



American Society of Human Genetics 68th Annual Meeting

POSTER ABSTRACTS

Abstract Topic	Abstract/Poster Board Numbers	
	Start #	End #
Genome Structure and Function	373	447
Prenatal, Perinatal, and Reproductive Genetics	448	540
Genetic Counseling, ELSI, Education, and Health Services Research	541	654
Cancer Genetics	655	975
Mendelian Phenotypes	976	1340
Bioinformatics and Computational Approaches	1341	1707
Omics Technologies	1708	1823
Epigenetics and Gene Regulation	1824	2015
Developmental Genetics and Gene Function	2016	2114
Complex Traits and Polygenic Disorders	2115	2664
Evolution and Population Genetics	2665	2786
Molecular and Cytogenetic Diagnostics	2787	3062
Cardiovascular Phenotypes	3063	3249
Statistical Genetics and Genetic Epidemiology	3250	3567

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

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

976T

Unravelling the genetic mechanisms of neonatal diabetes mellitus: An Egyptian experience. R. Elkaffas^{1,3}, H. Madani¹, B. AlKholi¹, N. Musa², Y. Shalaa², R.M.H. Elkaffas², M. Hassan², M. Hafez², S. Flanagan⁴, E. DeFranco⁴, K. Hussain^{2,5}. 1) Chemical pathology, Faculty of medicine-Cairo university, Cairo, Giza, Egypt; 2) Pediatric department, Cairo university; 3) UCL GOS Institute of Child Health, UCL, UK; 4) Institute of Biomedical and Clinical Science, University of Exeter Medical School, UK; 5) Sidra Medical & Research center, Qatar.

Background: Neonatal Diabetes Mellitus (NDM) is a rare monogenic form of Diabetes Mellitus presenting typically before the age of 6 months. NDM can be transient (TNDM) with 70% of these cases caused by a methylation defect in 6q24 or permanent (PNDM) with 40% caused by mutations in potassium channel subunits coding genes (*KCNJ11*, *ABCC8*). It may present in an isolated form or as a part of syndrome. Several studies have reported the genetic causes of NDM among different populations but to the best of our knowledge no previous studies had reported the genetic mechanisms of NDM among the Egyptian population. **Objective and hypotheses:** Our aim was to understand the genetic causes of NDM in Egyptian neonates presenting before the age of 6 months referred to a tertiary diabetes centre between the period 2013-2016. **Method:** In a cohort of 22 patients, anti-GAD antibodies and C-peptide testing were done in some cases to exclude type 1 Diabetes Mellitus. Sanger sequencing for *KCNJ11*, *ABCC8*, *INS* and *EIF2AK3* genes was done as a first tier genetic analysis for all cases. In absence of a causative mutation in these genes, we further undertook methylation analysis of 6q24 in cases presenting with TNDM, while in those presenting with PNDM targeted next generation sequencing (tNGS) for the coding regions and conserved splice sites of another 18 known genes to cause NDM were done. Targeted gene analysis was done in 4 syndromic cases with a distinctive phenotype suggesting a candidate gene. **Results:** Genetic causes were identified in 15/22 (68%) patients. These included 10 missense, 3 nonsense, 2 frameshift mutations and a complete loss of maternal methylation on chromosome 6q24. Mutations involved different genes; 4 in *ABCC8*, 3 in *GCK*, 3 in *EIF2AK3*, 2 in *KCNJ11*, 1 in *NEUROD1*, 1 in *ZFP57* and 1 in *SLC19A2*. Most of these patients (11/15) had previously reported mutations. In 4 patients, no mutation was identified in the first tier analysis but insufficient DNA was available for tNGS. In 3 patients no genetic cause could be identified.

977F

The Personalized Diabetes Medicine Program: Identification, diagnosis and treatment of monogenic diabetes. H. Zhang¹, J.W. Kleinberger¹, T.J. Mathias¹, Y. Guan², K.A. Maloney¹, E.A. Streeten¹, K. Blessings³, M.N. Snyder⁴, L. Bromberger⁵, J. Goehring⁶, A. Kimball⁷, C.M. Damcott⁸, C.O. Taylor⁹, M. Nicholson¹, D.C. Nwaba¹, K. Palmer¹, N. Ambulos¹, L.J.B. Jeng¹, A.R. Shuldiner¹, P. Levin⁹, D.J. Carey⁹, T.I. Pollin¹. 1) Division of Endocrinology, Diabetes, and Nutrition, University of Maryland School of Medicine, Baltimore, MD; 2) Rollins School of Public Health, Emory University, Atlanta, GA; 3) Geisinger Health System, Danville, PA; 4) MODEL Clinical Research, Research Division of Bay Endocrinology Associates, Baltimore, MD; 5) Harvey Institute for Human Genetics, Greater Baltimore Medical Center, Baltimore, MD; 6) Division of General Internal Medicine; Johns Hopkins University School of Medicine, Baltimore, MD; 7) University of Maryland Marlene and Stewart Greenebaum Cancer Center, University of Maryland School of Medicine, Baltimore, MD; 8) Regeneron Genetics Center, Tarrytown, NY; 9) Bay West Endocrinology Associates, Baltimore, MD.

Background Monogenic diabetes, including maturity onset diabetes of the young (MODY) comprises approximately 1-2% of all diabetes but is usually misdiagnosed as type 1 diabetes (T1D) or type 2 diabetes (T2D) and currently diagnosed correctly only ~5% of the time. A genomic diagnosis of monogenic diabetes informs proper treatment (sulfonylureas for *HNF1A/HNF4A* MODY, no treatment for *GCK* MODY vs. insulin for T1D, and for example, metformin for T2D) and familial risk assessment. The Personalized Diabetes Medicine Program (PDMP), part of the NIH IGNITE (Implementing GeNomics In PracTice) Network, was designed to implement, disseminate and evaluate a sustainable method for identifying, genomically diagnosing, and promoting individualized therapy for monogenic diabetes. **Methods** Patients were recruited into the program through a screening questionnaire completed in the waiting room or online or clinician- or self-referral. Patients suspected of monogenic diabetes underwent next generation sequencing for 40 genes, including 13 MODY genes as well as genes implicated in neonatal diabetes, lipodystrophy, syndromic diabetes, severe obesity and hyperinsulinemia. Automated filtering generated a list of nonsynonymous exonic or splicing variants with <5% minor allele frequency, which were then classified by the 2015 ACMG/AMP Guidelines. **Results** To date, 312 probands with suspected monogenic diabetes have enrolled and 269 have been sequenced. Among 362 high quality variants classified, 24 pathogenic or likely pathogenic variants were found in 26 patients, for a hit rate of 26/269 (9.7%). The variants were found in *GCK* (13), *HNF1A* (5), *LMNA* (2), and *INS* (2), *KCNJ11* (1) and *MC4R* (1). We also found 176 benign or likely benign variants and 162 variants of uncertain significance. Among the 14 patients with *GCK* variants, prior to sequencing, only two patients were suspected to have MODY, nine were diagnosed with T2D, two were diagnosed with diabetes with no type and one had pre-diabetes. Follow-up of these patients to assess effects of genomic diagnosis of monogenic diabetes on diagnosis, treatment, clinical outcomes, healthcare utilization/cost-effectiveness of screening, and cascade testing is in progress. Results will be used to inform a clinical service model to assure that all individuals and families with monogenic diabetes have access to diagnosis and customized treatment and more broadly access to personalized medicine across disease areas.

978W

High resolution ex vivo magnetic resonance imaging of the mucopolysaccharidosis I canine brain and associated white matter pathology. *P. Dickson*^{1,2}, *M. Egeland*², *S. Chen*³, *I. Nestrasi*⁴, *S.Q. Le*^{1,2}, *S. Kan*², *J. Jens*⁵, *E. Snella*⁵, *N.M. Ellinwood*⁶, *M. Guzman*⁶, *L.E. White*⁷, *J.D. Cooper*^{1,2}, *J. Provenzano*³. 1) Department of Pediatrics, Washington University School of Medicine, Saint Louis, MO; 2) Department of Pediatrics, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA; 3) Department of Radiology, Duke University, Durham, NC; 4) Department of Pediatrics, University of Minnesota, Minneapolis, MN; 5) Department of Animal Science, Iowa State University, Ames, IA; 6) Department of Pathology, Saint Louis University School of Medicine, Saint Louis, MO; 7) Department of Orthopedic Surgery, Duke University, Durham, NC.

Mucopolysaccharidosis I (MPS I) is a lysosomal storage disorder causing progressive physical and neurological disability. Neurological manifestations include progressive white matter hyperintensities on brain magnetic resonance imaging (MRI) that are thought to reflect dysmyelination. To explore white matter in MPS I brain, we performed high-resolution, ex vivo brain MRI in the canine model of MPS I and performed comparative neuropathology. MPS I and unaffected heterozygous control animals were sacrificed at seven weeks (n=3/group) or eighteen weeks (n=2 MPS and 4 controls). Following perfusion and fixation in formalin and gadolinium, excised brains were imaged on a 7T small-animal MR imaging system at Duke Center for In Vivo Microscopy as previously published (Middleton et al; *Neurorad J.* 2017; Middleton et al; *Neurorad J.* 2017). We placed 5 non-overlapping regions of interest 12px by 12px per slice on the left and right anterior and posterior internal capsule and anterior and posterior corticospinal tract in each contiguous slice containing the structures. Diffusion tensor imaging showed normal to slightly high fractional anisotropy of all white matter regions of MPS dogs at 7 weeks (p=NS). In the 18-week group, fractional anisotropy was lower in both MPS dogs compared to controls in the white matter regions. Luxol-fast blue staining of internal capsule showed reduced percent thresholding in 18-week MPS dogs versus controls. Additional work is underway to evaluate the composition of white matter. Diffusion tensor imaging may permit sensitive, early detection of abnormal myelination in MPS I brain.

979T

Screening strategy to classify mutations in a rare Mendelian disease: Glycogen storage disease type 1a. *K.L. Plona*¹, *E.M. Ponting*¹, *M.L. Drumm*^{1,2}.

1) Department of Genetics and Genome Sciences, Case Western Reserve University School of Medicine, Cleveland, OH; 2) Department of Pediatrics, Case Western Reserve University School of Medicine, Cleveland, OH.

The source of clinical heterogeneity among patients with the same monogenic disorder is largely speculative and may reflect genetic differences, stochastic events, or environmental sources. Understanding sources of clinical variation between patients with the same monogenic disorder can provide insight into treatment strategies. Our work uses glycogen storage disease 1a (GSD1a [MIM 232200]) as a Mendelian disease model and focuses on understanding how different mutations in *G6PC* [MIM 613742] contribute to differences in molecular phenotype. GSD1a is a rare metabolic disease characterized by an inability to break down glycogen to release glucose from cells during fasting, resulting in life-threatening hypoglycemia. Currently there are no curative therapies, and patients rely on symptomatic treatment and a regimented diet to maintain normoglycemia and sustain life. It is well established that a reduction in glucose-6-phosphatase enzyme function causes GSD1a, however it remains unclear how each of the 85+ reported pathogenic mutations in the *G6PC* gene, along with variants of unknown significance, alter the function of G6PC protein. Further molecular classification of the mutations can help inform which therapeutic strategies may be appropriate. We have developed an initial screening system in which *G6PC* cDNAs, fused in frame with a C-terminal fluorescent reporter, are transiently transfected into cultured cells. Our initial studies compare a panel of 11 mutations reported as pathogenic and 2 variants of unknown significance, assessing molecular outcomes for each. We have selected the panel to represent multiple mutation types, gene locations, protein domains, and patient ethnicities. Using high-throughput fluorescent imaging we have assessed these mutations for G6PC protein presence, relative quantity, and localization. As expected, one frameshift and two nonsense mutations abolish protein production. We also discovered that four of the nine tested missense mutations, including the most prevalent mutation (p.R83C), result in reduced protein levels. Functional studies are now underway to determine the relative amounts of catalytic activity each variant possesses, providing an opportunity to relate enzymatic function to clinical outcome and predict which preclinical tests would be most the appropriate next step for a given mutation. We hope to use this systematic approach as a model to enhance understanding of molecular pathogenesis in other rare diseases.

980F

Improved survival and amelioration of disease-related liver pathology in a mouse model of homocystinuria with a novel homocysteine degrading enzyme. C.L. Daige¹, M. Bonem², W. Lu², J. Wiggins¹, G. Agnello¹, J. Wooldridge¹, G. Georgiou², E. Stone², S. Rowlinson¹. 1) Aeglea BioTherapeutics, Austin, TX; 2) University of Texas, Austin, TX.

Background: Classical homocystinuria (HCU) is an inherited disorder of sulfur metabolism due to a genetic defect in cystathionine- β -synthase (CBS). The enzyme deficiency leads to homocysteine accumulation, which plays a key role in the progressive and serious disease-related complications, including skeletal abnormalities, neuropsychiatric impairment, and thromboembolism. Current disease management, which includes dietary protein and methionine restriction, is problematic with poor compliance. We investigated the potential of an alternative approach to disease management utilizing a novel homocysteine degrading enzyme derived from human cystathionine gamma lyase (CGL). **Methods:** *In vivo* efficacy of ACN00121 was tested in a murine model of homocystinuria (CBS^{-/-}) that presents with growth impairment and severe hepatopathy and results in 90% mortality during the first 2 weeks of neonatal life (PMID: 7878023). ACN00121 (25 mg/kg, i.p., twice per week) was administered to CBS^{-/-} mice (supplemented with betaine through weaning) starting on post-natal day 10 until day 50. The effect of treatment on lifespan, and liver histopathology was assessed. **Results:** Untreated CBS^{-/-} mice survived a median of 25 days and severe liver abnormalities, including marked hepatic steatosis, was observed. Treatment with ACN00121 markedly improved survival with no deaths at study termination on day 60. The reduction in mortality was accompanied by marked improvements in liver abnormalities, including resolution of the steatosis. **Discussion:** Enzymatic reduction of plasma homocysteine levels prevents neonatal death and corrects pathological liver manifestations in a homocystinuria mouse model. Given the limitations of current disease management approaches, ACN00121 warrants further investigation as a new potential treatment approach for this devastating disorder. **Disclosures:** C. Daige, G. Agnello, J. Wiggins, J. Wooldridge, and S. Rowlinson are employees of and have an equity interest in Aeglea BioTherapeutics, Inc. E. Stone and G. Georgiou have an equity interest in Aeglea BioTherapeutics, Inc. .

981W

20p11.2 deletions cause Hyperinsulinemic Hypoglycemia. T.W. Laver¹, W.N. Wakeling¹, R. Caswell¹, B. Bunce², J.A.L. Houghton³, M. Kershaw³, V. Saraff⁴, E.M. Honey⁴, N. Murphy⁵, I. Banerjee⁶, K. Hussain⁷, S. Ellard^{1,2}, S.E. Flanagan¹. 1) University of Exeter Medical School, Exeter, United Kingdom; 2) Department of Molecular Genetics, Royal Devon & Exeter NHS Foundation Trust, Exeter, United Kingdom; 3) Birmingham Children's Hospital, Birmingham, United Kingdom; 4) Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Pretoria, South Africa; 5) Department of Paediatric Endocrinology, Children's University Hospital, Dublin, Ireland; 6) Department of Paediatric Endocrinology, Royal Manchester Children's Hospital, United Kingdom; 7) Department of Paediatric Medicine Sidra Medical & Research Center OPC, Doha, Qatar.

Large deletions are an important cause of rare genetic disease which are often associated with syndromic disorders as a result of multiple genes being disrupted. Large deletions can provide important insights in gene discovery studies as they can act to narrow the potential list of causative genes. An example of a disease where new genetic aetiologies remain to be discovered is congenital hyperinsulinism, a disorder of unregulated insulin secretion causing severe hypoglycaemia, where screening of the known genes does not identify a mutation in 55% of patients. We screened 44 patients with permanent hyperinsulinism without a genetic diagnosis for large deletions (>100kb) using genome sequencing. We found one deletion within the cohort; this was a 3.8Mb *de novo* deletion on chromosome 20p11. We subsequently searched for 20p11 deletions using off-target reads in a cohort of 435 patients with permanent hyperinsulinism without a genetic diagnosis who had a sample sequenced by targeted next generation sequencing. We identified two further patients with an overlapping deletion on 20p11 (minimal deleted region 20:202000000-226000000). All three deletions were confirmed by digital droplet PCR (ddPCR). Two were shown to have arisen *de novo* in the proband while no parental samples were available for the third family. In all three patients ddPCR confirmed that the minimal deleted region encompasses the coding regions of seven genes and that the expression of an eighth gene, *FOXA2*, is potentially eliminated by disruption of the topologically associating regulatory domain. Three of the genes within the minimal deleted region (*NKX2-2*, *INSM1* and *FOXA2*) are known to have a role in the regulation of insulin secretion from the pancreatic beta cell. Functional studies will be required to determine which of the eight genes is causative of the hyperinsulinism in our patients. We have identified 20p11 deletions in three unrelated patients with hyperinsulinism confirming this as a novel genetic aetiology for this condition. None of our patients had been flagged for screening of deletions based on their phenotype, highlighting the utility of routinely searching for large deletions in gene discovery studies.

982T

Characterization of a murine model of glycogen storage disease type VI. Y. Lee^{1,3}, A. Estrella^{1,3}, D. Weinstein^{1,2,3}, T. Chengsupanimit³, L. Brown³. 1) GSD Program, Dept. of Pediatrics, University of Connecticut School of Medicine, Farmington, CT; 2) GSD Program, Connecticut Children's Medical Center, Hartford, CT; 3) GSD Program, University of Florida College of Medicine, Gainesville, FL.

Glycogen storage disease type VI (GSD-VI, OMIM232700) is an autosomal recessive disorder caused by deficiency of liver isoform of glycogen phosphorylase (PYGL). GSD-VI patients manifest hepatomegaly, growth retardation, ketotic hypoglycemia after an overnight fast and mild hypoglycemia after prolonged fasting. The disease symptoms are relatively mild in infancy and childhood, however, severe and recurrent hypoglycemia, severe hepatomegaly, and post-prandial lactic acidosis have been described before. The affected patients have increased risk of hepatic adenoma formation in late childhood and adulthood. Corn starch therapy may prevent hepatomegaly and hypoglycemia. And the secondary complications such as short stature, delayed puberty and osteoporosis will be improved with better metabolic control. To develop therapies and investigate the detailed mechanism of the disease, we generated and characterized in detail of the phenotypes *Pygl* deficient mouse (*Pygl*^{-/-}) model by inserting promoter driven cassette in between exon 2 and 3 of *Pygl* gene. In this study, as like affected human patients, GSD-VI mice showed mild hypoglycemia with normal growth patterns. In fasting glucose test and blood biochemistry, GSD-VI mice also showed mild hypoglycemia during fasting but with normal blood metabolites concentrations compared to wild type (*Pygl*^{+/+}) or heterozygous (*Pygl*^{+/-}) mice. However, histological analysis showed increased accumulation of glycogen in the liver of *Pygl*^{-/-} mice. With continued hepatomegaly, fibrosis and fibrotic lesions were found in aged GSD-VI mouse livers. In conclusion, the *Pygl*^{-/-} mice successfully mimic human disorder of GSD-VI. This animal model will help to understand detailed phenotypes the human disease and contribute future development of therapies.

983F

An 11-year-old with lysinuric protein intolerance presenting with xanthomas, failure to thrive, and hepatic adenomas. D.T. Segarra^{1,2,3}, R. Bonilla-Guerrero⁴, K. Spence², A. Sanchez-Valle^{1,2,3}. 1) University of South Florida Morsani College of Medicine, Tampa, FL; 2) Department of Pediatrics, Division of Genetics and Metabolism, USF Health, Tampa, FL; 3) Tampa General Hospital, Tampa, FL; 4) Quest Diagnostics, Inc., San Juan Capistrano, CA.

Background: Lysinuric protein intolerance (LPI) is an autosomal recessive metabolic disorder caused by a mutation in the solute carrier gene *SCL7A7*, which leads to renal reabsorption and intestinal absorption defect of the cationic amino acids lysine, arginine, and ornithine. In infancy, this disorder typically presents with vomiting and diarrhea. Other early clinical features include stupor and/or coma after protein rich meals, poor feeding, failure to thrive, hepatomegaly/splenomegaly, and muscular hypotonia. LPI is a rare inborn error of metabolism that should be included in the evaluation of children with severe failure to thrive and other systems' involvement such as hepatic adenomas. Typical failure to thrive work-up for inborn errors of metabolism would not necessarily identify these cases, as urine amino acids analysis is not part of the usual assessment. **Objectives:** To emphasize important clinical considerations and expanded testing for failure to thrive and liver involvement. **Case description:** We present an 11-year old male referred for genetics evaluation due to hepatic adenomas and a liver biopsy suggestive of a glycogen storage disease. On evaluation, he was found to have short stature, poor appetite, failure to thrive and significant leg xanthomas. His metabolic evaluation revealed a mildly elevated ammonia of 81 (ref <49umol/L) and an increased urinary excretion of lysine and arginine suggestive of LPI. DNA analysis of the *SLC7A7* gene confirmed a homozygous deletion of exons 6-11. **Conclusions:** Although LPI is a rare metabolic disorder, it should be considered in cases of failure to thrive with multiple organ involvement. Our patient had not been diagnosed even though he had been evaluated with plasma amino acids, urine organic acids and acylcarnitine profile. Urine amino acids analysis is essential in these cases.

984W

Would double heterozygotes of *SLC3A1* and *SLC7A9* pathogenic variants symptomatic for cystine stones? A perspective from population genetics and biostatistics. C. Wu¹, B. Eisner², N. Meeks³, A. Tsai⁴, G. Berry⁴.

1) Harvard Medical School Affiliated Hospitals, Boston, MA; 2) Massachusetts General Hospital, Boston, MA; 3) University of Colorado School of Medicine, Denver/Aurora, CO; 4) Boston Children's Hospital, Boston, MA.

INTRODUCTION AND OBJECTIVES: Cystine stones are caused by cystinuria, the most commonly-seen inborn error of metabolism. Gene mutations at *SLC3A1* and *SLC7A9* have been mapped as the causes of the disease. The inheritance model for cystine stones is autosomal recessive. However, would double heterozygotes of *SLC3A1* and *SLC7A9* (individuals who carry both *SLC3A1* and *SLC7A9* pathogenic variants in heterozygote state) present with cystine stones? Is there a synergistic heterozygosity effect? Since the prevalence of pathogenic variants is low, the probability of direct observation of double heterozygotes is even lower. We, therefore, designed this study with genomic and biostatistics approaches to answer this question. **METHODS:** We've parsed the 1000 Genomes Database Phase 3 (1KG) for identification of variants in the general population. To identify pathogenic mutations, we parsed the Human Gene Mutation Database (HGMD). We hypothesized that double heterozygotes do not present with cystine stones. With a neutral phenotype, no symptoms, no natural selection effect, we calculated the theoretical double heterozygote frequency based on Hardy-Weinberg Equilibrium (HWE) and Mendel's Law of Independent Assortment. One-sample proportion statistical test was performed to compare observed frequency to the theoretical frequency. **RESULTS:** In 1KG, the pathogenic variants have an allele frequency of 0.52% for *SLC3A1* (q1), and 0.24% for *SLC7A9* (q2). There were no homozygotes, compound heterozygotes, double homozygotes, or double heterozygotes. Based on HWE, the theoretical pathogenic *SLC3A1* carrier rate (2p1q1) is calculated as 1.03%, and 0.48% for *SLC7A9* (2p2q2). The two genes are located on different chromosomes and will obey Mendel's Second Law. Thus the theoretical double carrier frequency (2p1q1*2p2q2) would be 4.94×10^{-5} . By one-sample proportion statistical test, we compared the theoretical double heterozygote rate (2p1q1*2p2q2) to the observed double heterozygote rate (0). Z value is 0.352, and p-Value is 0.7251. The statistical null hypothesis that two proportions are equal is not rejected. **CONCLUSIONS:** We did not find any double heterozygotes for *SLC3A1* and *SLC7A9* mutations. However, this observation is not significantly different from the theoretical frequency under the hypothesis that double heterozygotes have a neutral phenotype with no cystine stone symptoms. Therefore, statistically, double heterozygotes would be asymptomatic and do not present with cystine stone.

985T

Clinical and functional characterization of Melanocortin 4 Receptor (*MC4R*) variants in African-American and Hispanic children with severe early onset obesity (SECO). M.C. De Rosa¹, A. Ches², S. McCormack², B. Weaver³, M. McDonald⁴, K. Liimatta¹, M. Rosenbaum¹, C. Doege¹, J.N. Hirschhorn¹, S. Grant², V.V. Thaker¹. 1) Columbia University Medical Center, New York, NY; 2) Children's Hospital of Philadelphia, Philadelphia, PA; 3) Boston University Medical School, Boston, MA; 4) Boston Children's Hospital, Boston, MA.

Introduction: *MC4R* mutations are the most common cause of monogenic obesity in children of European ancestry, with 2.5-6% prevalence. Data are limited on the prevalence of *MC4R* mutations in children with SECO in African-American and Hispanic ancestry/ethnicity. The objective of this study was to identify the prevalence of rare variants in *MC4R* in this population and perform their functional characterization. **Methods:** We assessed the prevalence of rare variants in *MC4R* in children with SECO, defined as body mass index > 120% of the 95th percentile for age, documented before age 6, and present for at least 6 months. Subjects were recruited from 3 tertiary care U.S. hospitals and identified through clinical encounters or by using a validated algorithm for identification of SECO from Electronic Health Records (PMID: 27452794). Longitudinal analysis of BMI trajectory was performed using Super Imposition by Translation and Rotation (SITAR) to generate size, tempo, and velocity estimates for individual subject (PMID: 20647267). DNA was obtained under institution approved consent from the Biobank, or in-person recruitment. *MC4R* variants were identified by whole exome or targeted Sanger analysis. For 3 novel variants, we performed functional assessment of cAMP response to its native ligand, α -MSH, and measured expression by surface biotinylation assay in HEK293 cells. **Results:** Our cohort comprised of 298 children, 86% were African-American and 20% were Hispanic by self-reported ancestry/ethnicity, and 51% were female. We identified 8 individuals with heterozygous rare variants (MAF < 1%, R7S, F202L (n=2), M215I, G252A, V253I, I269N, F284I) and 15 with common variants (V103I, I198=, Q156=, I251L). Three of these were novel (M215I, G252A, F284I). In an *in vitro* system of *MC4R* overexpression in HEK293 cells, the initiation of cAMP response to α -MSH required 10-fold higher concentration in M215I & G252A, and 1000-fold higher in F284I. The peak cAMP response achieved in WT protein was not seen in the tested variants with increasing concentrations of α -MSH. All of the identified variants were deemed pathogenic based on our functional studies and/or prior literature reports. **Conclusions:** In our cohort of children with severe and persistent early onset obesity from underrepresented minorities, 2.7% of the children had rare, pathogenic heterozygous variants in *MC4R*. We established the pathogenic nature of 3 novel variants in *MC4R* in our cohort by *in vitro* functional studies.

986F

Variation in the *Apolipoprotein L1 (APOL1)* gene—a search for pleiotropy beyond renal disease. A. Ihegword¹, D. Carranza-Leon¹, W.Q. Wei², L. Bastarache³, J.C. Denny², Q. Feng¹, C.M. Stein^{1,3}. 1) Division of Clinical Pharmacology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN; 2) Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN; 3) Department of Pharmacology, Vanderbilt University, Nashville, TN.

Objective: 1) To assess for pleiotropic clinical effects of 2 *APOL1* variants. 2) To determine if 1 *APOL1* variant imparts increased renal risk. **Background:** Risk of end stage renal disease (ESRD) is increased 3-5 fold in African Americans (AAs). Two common *APOL1* variants, G1 and G2, contribute to this disproportionate risk; ~15% of AAs carry 2 *APOL1* variants and have a marked increased risk of ESRD (OR ~7). *APOL1* protein is widely expressed and has roles in trypanolysis, autophagy and other immune mechanisms, apoptosis, anti-viral activity, and lipid biology; thus, there is substantial potential for pleiotropy. In some studies, in addition to renal risk, cardiovascular risk was increased with 2 *APOL1* variants; however, the pleiotropic clinical consequences of *APOL1* variants are undefined. Also, the risk of renal disease with 1 *APOL1* variant is unclear. **Methods:** We performed a phenome wide association scan (PheWAS) to identify pleiotropic effects of G1 and G2 *APOL1* alleles in 4,960 AA patients in BioVU, a DNA biobank linked to de-identified EHR. The G1 allele (rs73885319/rs60910145) was available on standard platforms. The G2 allele was imputed using rs12106505 [in high LD ($r^2 > 95\%$) with rs71785313]. Analyses were adjusted for age, body mass index, and sex. PheWAS significance was set at 2.75×10^{-5} . **Results:** Of 4,960 patients, 1455 (29.3%) were heterozygous and 286 (5.8%) homozygous for G1 (MAF 0.23), and for G2, 805 (16.2%) were heterozygous and 87 (1.8%) homozygous (MAF 0.14). Two *APOL1* risk alleles were present in 643 patients (13%) and 2057 (41.4%) carried neither variant. Comparing individuals with 2 risk alleles to those with 0 or 1, the top 30 consecutive PheWAS hits were all related to renal disease (e.g., dialysis (OR 3.89, $P=4.91 \times 10^{-21}$), ESRD (OR 3.33, $P=1.73 \times 10^{-16}$), kidney transplantation (OR 3.69, $P=3.37 \times 10^{-16}$), nephritis (OR 3.02, $P=2.45 \times 10^{-12}$), and chronic renal failure (OR 2.20, $P=1.49 \times 10^{-11}$). There was no strong evidence of clinical pleiotropy; the strongest non-renal association was acute upper respiratory infection (OR 0.64, $P=5.92 \times 10^{-4}$). Comparing G1/G0 or G2/G0 to G0/G0 individuals, no phenotypes were significant for either allele. **Conclusion:** There was a strong association between the presence of 2 *APOL1* risk alleles and renal disease, but no strong evidence for pleiotropic clinical risk. The presence of 1 *APOL1* risk allele was not associated with renal disease.

987W

Confirmation of *ACER3*-related recessive neurodegeneration, and preliminary evidence for feasibility of biochemistry-based *ACER3* variant classification. P. Bauer¹, C. Cozma¹, A.M. Bertoli-Avella¹, C. Beetz¹, A. Shukla^{2,3}, N. Kamath^{3,4}, K.M. Girisha^{2,3}, M. Alfadhel⁵, K.K. Kandaswamy¹, O. Paknia¹, O. Brandau¹, A. Rolfs^{1,6}. 1) Centogene AG, Rostock, Germany; 2) Department of Medical Genetics, Kasturba Medical College, Manipal, India; 3) Manipal Academy of Higher Education, Manipal, India; 4) Department of Paediatrics, Kasturba Medical College, Mangalore, India; 5) Genetics Division, Department of Genetics, King Abdulaziz Medical City, Riyadh, Saudi Arabia; 6) Albrecht-Kossel-Institute for Neuroregeneration, Medical University Rostock, Rostock, Germany.

Ceramidases cleave ceramides into fatty acids and sphingosine; they are thereby important enzymatic players in lipid metabolism. *ACER3* encodes alkaline ceramidase 3, i.e. one of the five human ceramidases. Knockout of the murine homologue *Acer3* results in dysregulation of sphingosine, sphingosine-1-phosphate and various ceramides in the brain, and is associated with adult onset degeneration of Purkinje cells. In humans, the homozygous *ACER3* missense variant c.98A>G (p.Glu33Gly) was recently suggested to underlie progressive leukodystrophy in two patients from a single family. We investigated two unrelated children which had initially developed normally, but started to loose motor milestones and became underweight at around 6 months of age. Routine biochemical workup revealed lactic acidosis, and brain MRIs were suggestive of delayed myelination and diffuse cerebral atrophy. The patient from family 1 is currently 17 months old, whereas the patient from family 2 deceased at age three years. Upon whole exome sequencing and confirmatory Sanger sequencing we identified homozygous *ACER3* variants in both index cases (c.53T>C (p.Leu18Pro) in family 1; c.399delC (p.Thr133Thrfs*6) in family 2). Consistent with findings on the *Acer3* knockout mouse, targeted blood-based metabolomics revealed elevated levels of several ceramides and dihydroceramides in the index of family 2. Interestingly, the heterozygous parents appeared to be intermediate between patient and controls in these analyses. We therefore turned to untargeted metabolomics, and compared 16 wild-type controls with 17 healthy heterozygous carriers of rare *ACER3* coding variants (absent from gnomAD or minor allele frequency of 0.0006) incl. the parents from family 2. Principal component analysis and unsupervised clustering of the metabolomics data completely separated both groups. Our study confirms *ACER3* as a novel recessive disease gene in humans, and extends the clinical and mutational spectra. Our biochemical analyses suggest that the pathogenicity of *ACER3* variants in heterozygous carriers can be predicted by metabolic profiling. An extension of this concept to other recessive metabolic and potentially even non-metabolic disorders has the potential to significantly impact on the interpretation of variants of unclear significance.

988T

Common genetic variation contributes to the variable expressivity of monogenic familial hypercholesterolemia. M.T. Oetjens, J.B. Leader, A.C. Sturm, L.K. Jones, C.L. Martin, D.H. Ledbetter on behalf of the DiscovEHR collaboration. Geisinger, Danville, Pennsylvania, USA.

Familial Hypercholesterolemia (FH) is a genetic condition of high cholesterol that if untreated can lead to aggressive coronary artery disease. While FH is amongst the most prevalent monogenic diseases (~1:250), it remains underdiagnosed despite routine screening of cholesterol. FH exhibits variable expressivity where individuals can range from low/normal to extremely high levels of low-density lipoprotein cholesterol (LDL-C) even amongst carriers of the same genomic variant. To understand the factors that contribute to variable expressivity in FH, we tested the hypothesis that common genetic variation contributes to LDL-C levels in addition to the presence of a monogenic FH variant. In our genotype-first approach, we analyzed ~90,000 DiscovEHR individuals exome-sequenced and genotyped on whole-genome arrays, screened for FH variants, and evaluated for LDL-C polygenic risk scores (PRS_{LDL-C}). We included all *LDLR* sequence variants identified as likely pathogenic or pathogenic (LP/P) loss-of-function variants as well as one known pathogenic exonic duplication. *LDLR* missense variants were also included if annotated in ClinVar as LP/P with a two-star review status. In total, we identified 260 individuals with pathogenic *LDLR* variants. Maximum EHR-documented LDL-C (maxLDL) was used as an approximation of the untreated state. The PRS_{LDL-C} was calculated from 207 variants that passed genome-wide significance in an external cohort and in linkage equilibrium. Compared to non-carrier first-degree relatives (FDRs) the *LDLR* carriers' LDL-C levels were on average 88.53 mg/dL higher (44 pairs; non-carrier: mean=143.00 mg/dL, SD=41.40 mg/dL; carrier mean=231.53 mg/dL, SD= 85.29 mg/dL). When we restricted our analysis to unrelated individuals, one SD of the PRS_{LDL-C} increased maxLDL by 22.61 mg/dL ($p = 0.0019$). We next applied the same approach to another monogenic FH gene, *APOB* ($n=127$) and did not find evidence of an association ($b=-2.00$ mg/dL; $p=ns$), suggesting a gene-specific effect of the PRS_{LDL-C} on FH. We confirm that rare pathogenic *LDLR* variants increase LDL-C by ~2 SD and demonstrate that common variation contributes to the variability in FH expression (~0.5 SD per PRS_{LDL-C} SD). These findings provide a potential explanation for asymptomatic monogenic FH and may provide a precision medicine approach to predicting clinical severity and CAD risk based on rare and common genetic variation.

989F

Marked hypertriglyceridemia as a biochemical hallmark of familial partial lipodystrophy caused by LMNA mutations. L. Rutkowska, K. Salacinska, I. Pinkier, D. Salachna, L. Jakubowski, A. Gach. Polish Mother's Memorial Hospital Research Institute, Department of Genetics, Lodz, Poland.

Familial partial lipodystrophy (FPL; MIM 150330) is a rare AD transmitted disorder mostly caused by mutation in *LMNA* gene, located on the long arm of chromosome 1. The *LMNA* gene (MIM 151660) encodes lamin A and lamin C proteins providing integrity of cell membranes. Typical clinical features in patients with FPL are excessive subcutaneous fat accumulation in the neck and face, reduced subcutaneous fat in the arms and legs and muscular definition with variable muscular hypertrophy and prominent superficial veins. The extent of adipose tissue loss usually determines the severity of the associated metabolic complications such as hyperglycemia, hyperinsulinemia, diabetes mellitus, increased serum triglycerides, hepatic steatosis, pancreatitis, high blood pressure, and premature atherosclerosis with an increased risk of coronary heart disease. What's more, after puberty some women may develop polycystic ovary syndrome (PCOS). A group of 60 patients with familial atherogenic dyslipidemia were tested using custom NGS panel. In two patients R482Q alteration in *LMNA* gene was found. We have analysed genotype-phenotype correlation in affected family members in 2 generations. The two probands were referred for genetic counselling due to history of infertility and atypical adipose tissue distribution. Medical history of 36-year old woman and 30-year old woman, showed severe hypertriglyceridemia, PCOS and insulin resistance with diabetes mellitus and impaired glucose tolerance, respectively. Detection of a heterozygous missense mutation at codon 482 c.1444C>T in *LMNA* gene, confirmed the diagnosis of FPLD2. In the Family 1 study revealed presence of FPLD2 in 8 family members in two generations. In Family 2 *LMNA* mutation was confirmed in proband's mother. Interestingly, most of them did not demonstrate phenotype as evident as the probands. The only common feature, shared by all affected family members was significant hypertriglyceridemia ranging 400-580 mg/dl. Spectrum of other clinical problems was wide, but we observed strong variation in the feature severity and distribution. To conclude, it is great diagnostic challenge to identify FPL patients, in terms of clinical diagnosis and qualification for molecular diagnostic, based on abnormal fat distribution and other physical features, especially in men. It seems to be that only common symptom is severe hypertriglyceridemia, therefore it is reasonable to diagnose the patient, with very high triglycerides level, for FPL.

990W

AAV8-mediated hLDLR gene transfer in patients with homozygous familial hypercholesterolemia: Interim analysis of the safety profile in the first 6 subjects. *M. Cuchel¹, R. Carr¹, A. Bajaj¹, C. Brent¹, P.B. Duell¹, J.C. Tardif¹, M. Linton¹, R. Fiorentino², S. Yoo³, E.A. Meagher¹, D.J. Rader¹.* 1) Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 2) Oregon Health and Science University, Portland, OR; 3) Montreal Heart Institute, Montreal, Quebec; 4) Vanderbilt University, Nashville, TN; 5) Regenxbio, Rockville, MD.

AAV-based vectors are emerging as an effective tool for gene transfer targeting the liver. Bi-allelic mutations in the LDL receptor (*LDLR*) gene are the most common cause of Homozygous Familial Hypercholesterolemia (HoFH), which is characterized by markedly elevated cholesterol levels and early atherosclerotic cardiovascular disease. Reconstituting functional *LDLR* in the liver may be a valuable treatment strategy. We are conducting a first-in-human trial of AAV8-mediated *hLDLR* gene transfer in HoFH subjects with documented *LDLR* mutations. We enrolled 6 subjects, 3 subjects in each of 2 dosing cohorts that received a single IV administration at the doses of 2.5×10^{12} GC/kg and 7.5×10^{12} GC/kg respectively. Weekly blood sampling was performed during the first 12 weeks, followed by further safety assessment up to 52 weeks post-vector administration. One subject who received the lower dose experienced a mild transitory activation of the innate immune system accompanied by hypotension and elevation in transaminases approximately 22 hours post dosing that resolved within a day. All three subjects who received the higher dose experienced an elevation in transaminases 4-6 weeks post-dosing. The peak ALTs were 165, 331, and 1496 IU/L in the 3 subjects (normal range: 6-41 IU/L). All 3 subjects were asymptomatic and responded rapidly to the initiation of prednisone at 60 mg/day followed by a slow taper. A sustained transgene expression has been observed in AAV-mediated gene transfer trials in hemophilia, and could be preserved by the administration of steroids after mild to moderate increases in transaminases observed in some subjects that were attributed to a T-cell response. Similarly, a T-cell response is likely to be the cause of the transaminases elevations observed in the 3 subjects that received the higher dose of the AAV8-*hLDLR* vector. However, the earlier timeframe and the degree of the elevation were unexpected, suggesting that HoFH phenotype or the *LDLR* transgene may alter the immunological response following AAV vector administration. Our ability to assess short-term efficacy on LDL-C levels is confounded by the hepatocyte injury and the steroid therapy. These results highlight the need to further our understanding of the host response of HoFH patients to the administration of AAV-based vectors and provide insights for a more appropriate design to evaluate the safety and efficacy of AAV8-*hLDLR* gene therapy in HoFH patients.

991T

Genetic and dietary manipulation of stearate levels rescues a model of retinal degeneration and ER stress. *R.A.S. Palu, C.Y. Chow.* Human Genetics, University of Utah School of Medicine, Salt Lake City, UT.

An important goal of the Precision Medicine Initiative is to address the phenotypic heterogeneity that impedes diagnosis and treatment in both Mendelian and complex genetic diseases. Cryptic genetic variation is a key contributor to this heterogeneity, but the underlying genetic architecture and modifiers are largely unknown. Understanding the roles of genetic modifiers in disease processes will enable the development of individualized therapeutic approaches. One process that commonly contributes to phenotypic heterogeneity is the ER stress response. In a previous study, we utilized the *Drosophila* Genetic Reference Panel, a collection of ~200 genetically diverse and fully sequenced fly lines, to study the effects of natural genetic variation on a model of ER stress-associated retinal degeneration. Overexpression of a mutant, misfolded rhodopsin protein (*Rh1^{G69D}*) induces ER stress in the developing eye, ultimately resulting in apoptosis and a small, degenerate adult eye that is incredibly variable among DGRP strains. Using genome-wide association methods, we identified 84 conserved candidate modifiers, many of which are associated with ER stress and apoptosis. One of these, *ELOVL6*, is an ER-associated fatty acid elongase that extends palmitate to stearate. Both *ELOVL6* activity and high stearate concentrations have been linked to ER stress, insulin sensitivity, and obesity in mammals. We demonstrate that loss of *ELOVL6* activity rescues cell death and retinal degeneration by reducing activation of the IRE1 and PERK branches of the ER stress response and JNK pro-apoptotic signaling in the *Rh1^{G69D}* model. Loss of *ELOVL6* in other tissues and cell types reduces ER stress and cell death in a similar manner, suggesting that *ELOVL6* is a general modifier of ER stress. Intriguingly, we find that administering dietary stearate during development is sufficient to increase degeneration and ER stress signaling, bypassing the need for *ELOVL6*. This suggests that *ELOVL6* regulates the ER stress response by controlling stearate concentration. These results raise the possibility that reducing *ELOVL6* function or stearate levels genetically, with small molecules, or through the diet could be a useful therapy that reduces ER stress signaling. Our findings suggest that *ELOVL6* is a general modifier of the ER stress response and a possible therapeutic target not only in retinal degeneration, but in many ER stress-associated diseases.

992F

The human transcriptional profile of Berardinelli-Seip congenital lipodystrophy. L.C. Ferreira^{1,2,3}, J.G. Lima^{1,4}, F.P. Freire-Neto^{1,2,3}, C.O. Mendes-Aguiar^{1,3}, S.M.B. Jeronimo^{1,2,3}. 1) Federal University of Rio Grande do Norte, Natal, Brazil; 2) Department of Biochemistry; 3) Institute of Tropical Medicine of Rio Grande do Norte; 4) Department of Clinical Medicine.

Mutations in *AGPAT2* (1-acylglycerol-3-phosphate O-acyltransferase 2) and *BSCL2* (codes for seipin protein) are the main cause of Berardinelli-Seip syndrome, a rare autosomal recessive disorder. *AGPAT2* catalyzes the reaction that produces phosphatidic acid, while seipin is a structural protein involved with lipid droplet biogenesis. The syndrome comprises complex metabolic disarrangement characterized by insulin resistance and diabetes. Of note, only *BSCL2*-affected individuals develop cognitive and intellectual disabilities. The mechanism linking genetic mutations to the clinical symptoms is still unsolved. This study aimed to compare the transcriptional profile in peripheral blood mononuclear cells (PBMC) from 23 individuals, grouped as: wild type control (WT, n=3); heterozygous for *BSCL2* mutation (*BSCL2*het, n=6); homozygous for *BSCL2* mutation (*BSCL2*mut, n=11) and homozygous for *AGPAT2* mutation (*AGPAT2*mut, n=3). All comparisons were performed using WT as the reference group. It was detected 28 differentially expressed genes (DEG) in *BSCL2*mut, with *NUAK2* as the top1. This finding is consistent with the notable hypertrophy in these patients, since *NUAK2* is a key regulator of muscle mass. Surprisingly, the unaffected *BSCL2*het group had 68 DEG. No gene achieved the genome wide significance threshold for comparison WT vs *AGPAT2*mut. Gene set enrichment analysis revealed key pathways for energy metabolism (Glycolysis, Krebs' Cycle and Oxidative Phosphorylation) and neurodegenerative diseases (Alzheimer's, Huntington and Parkinson's diseases) as enriched in *BSCL2*mut while *BSCL2*het individuals presented a remarkable inflammatory signature including the up-regulation of genes *IL1R2*, *CXCL1*, *IL8RB* and *IL8* ($P_{adj} < 0.05$). In addition, potential candidate genes for cardiovascular (*OLR1*, $\log_2FC = 2.55$, $P_{adj} = 0.00048$) and neurodegenerative (*TOR4A*, $\log_2FC = 1.75$, $P_{adj} = 0.033$; *KIF5C*, $\log_2FC = -1.28$, $P_{adj} = 0.045$; and *ACTG1*, $\log_2FC = 0.88$, $P_{adj} = 0.048$) diseases were differentially expressed in *BSCL2*het and *BSCL2*mut, respectively. A better understanding about the molecular mechanism of Berardinelli syndrome may shed light into our current knowledge concerning diseases such as atherosclerosis and Alzheimer.

993W

Distinguishing between monogenic and polygenic inheritance of extreme phenotype: Quantifying the value of family history in cases of extreme LDL-C. O. Soylemez^{1,2,3}, Z. Ahmad^{4,5}, S. Sunyaev^{1,2,3}. 1) Brigham and Women's Hospital, Boston, MA; 2) Harvard Medical School, Boston, MA; 3) Broad Institute, Cambridge, MA; 4) Division of Nutrition and Metabolic Diseases, Center for Human Nutrition, UT Southwestern Medical Center, Dallas, TX; 5) Department of Internal Medicine, UT Southwestern Medical Center, Dallas, TX.

Extreme values of quantitative traits or rare severe phenotypic presentations can be caused by individual mutations of very large effects segregating in Mendelian fashion. Alternatively, such phenotypes may reflect extremes of polygenic or oligogenic genetic inheritance. Beyond importance for basic genetics, this distinction is practically important for the design of genetic studies that can follow either Mendelian or complex trait paradigm. Currently, sequencing studies of rare phenotypes attribute only a fraction of cases to an obvious monogenic cause. Frequently, individuals with extreme phenotype have family members with milder presentation. Thus, it is expected that family history is informative about genetic architecture of severe phenotypes. We investigated this hypothesis in a series of population genetics computer simulations and the analysis of cases of severe familial hypercholesterolemia (FH). Recent work on FH in patients without known causative mutations in *LDLR*, *APOB* and *PCSK9* suggests that a large fraction of such cases may be attributed to polygenic inheritance whereby a large number of small-effect mutations collectively result in extreme LDL-C levels that are comparable to levels in patients with monogenic FH. It remains unclear to what extent the family history is informative to distinguish monogenic and polygenic FH patients with extreme LDL-C levels. We developed a population genetics framework to simulate families of individuals with extreme LDL-C levels and calculated the fraction of monogenic FH cases conditional on population genetics parameters (effect size distribution of FH mutations, demographic history), varying penetrance, and pedigree structure. Our simulation results show that the fraction of monogenic FH in patients with extreme LDL-C levels increases when the mutation's effect on LDL-C level and its selective effect are tightly coupled and decreases when the mutational target size is highly polygenic including hundreds of genes. Furthermore, we show that the family history is informative only for sufficiently large pedigrees. We applied our framework to real pedigrees that are clinically ascertained to have autosomal dominant FH with no causative mutations, and corroborated our simulation results. These findings have significant implications for study design of rare and unexplained genetic conditions as well as genetic testing strategies for new gene discovery, prevention and treatment of cardiovascular diseases.

994T

10-year experience with identification of cerebrotendinous xanthomatosis in newborns in Northern Israel; newborn screening followed by treatment from birth onward provides a “functional cure”.

T.C. Falik-Zaccai^{1,2}, A.E. DeBarber³, M. Yusim², S.B. Haroush¹, N. Chasnyk¹, N. Samra⁴, L. Even⁵, B. Higazi², N. Kfir^{1,2}, D. Savitzki⁶, A. Fedida^{1,2}, V.F. Sheffer⁷, H. Mandel⁸, E.S. Shinwell⁸, L. Kalfon¹. 1) Institute of Human Genetics Galilee Medical Center Naharia, 22100, Israel; 2) The Azrieli Faculty of Medicine, Bar Ilan University, Safed, Israel; 3) Physiology & Pharmacology Department, Oregon Health & Science University (OHSU), Portland, OR; 4) Unit of Medical Genetics, Ziv Medical Center, Safed, Israel; 5) Department of Pediatrics and pediatric endocrinology, The Galilee Medical Center, Naharia, Israel; 6) Pediatric Neurology unit, The Galilee Medical Center, Naharia, Israel; 7) Department of Neonatology, Galilee Medical Center, Naharia, Israel; 8) Department of Neonatology, Ziv Medical Center, Tzfat, Israel.

Cerebrotendinous Xanthomatosis (CTX), an autosomal recessive disorder of bile acid synthesis manifested by progressive multisystem damage, is rare worldwide but highly prevalent among several populations in Israel. Disease progression may be fully prevented by treatment with bile acids but when diagnosis is delayed neurodegeneration is not readily reversed by treatment. We describe our experience in Northern Israel performing population screening for CTX in one Druze village and performing biochemical screening for CTX in a high-risk newborn population. We genetically screened the child-bearing age population in an isolated Druze village for carriers of a pathogenic c.355delC founder genetic variant in the *CYP27A1* gene, ascertaining couples at risk and providing the option of genetic counseling and prenatal diagnosis. We then tested newborns and their siblings for this variant, providing 2-12 year follow-up and treatment for affected newborns and siblings identified. We recently expanded our screening program to perform screening for CTX in a high-risk newborn population. Newborn dried bloodspots obtained were subject to biochemical testing for CTX. Ten percent of samples consecutively collected were analyzed to identify CTX-causing founder genetic variants common among Druze and Moroccan Jewish populations. For individuals residing in the isolated Druze village a carrier rate for the c.355delC genetic variant of 1:7 was determined, leading to an estimated disease prevalence of 1:400. For Druze newborns from different villages born at local hospitals and screened for CTX a carrier rate of 1:17 was determined for the c.355delC genetic variant leading to an estimated disease prevalence of 1:1,200 in this newborn population. Five affected neonates born in the isolated Druze village were identified. Four were treated with chenodeoxycholic acid from age 1 month onward. One newborn was treated with cholic acid. Clinical outcomes were significantly improved for patients treated from infancy, with normal growth and no neurological symptoms, compared to six patients diagnosed and treated between ages 2-14. Our results validate prenatal diagnosis and newborn screening for CTX, and define pre-symptomatic treatment with bile acids as a functional cure preventing its common symptoms in childhood. The newborn screening pilot study data supports the feasibility of screening newborn dried bloodspots for CTX, setting the stage for large-scale prospective pilot studies.

995F

Homozygous mutation in *USP43*, a deubiquitinase gene, is linked to a novel interferon-mediated autoinflammatory disease. *I. Aksentijevich, H. Wang, D. Beck, D.L. Kastner.* National Human Genome Research Institute, Bethesda, MD.

Autoinflammatory diseases are a distinct group of rheumatologic diseases that are driven by abnormal activation of the innate immune system leading to recurrent episodes of systemic inflammation and a broad spectrum of other comorbidities. Autoinflammatory diseases are caused by dysregulation in critical immune regulatory complexes such as inflammasomes, cytokine receptors or inhibitors, enzymes, and proteasome complex. The excessive secretion of inflammatory cytokines can lead to chronic and debilitating phenotypes and may be life-threatening. These disorders are particularly amenable to treatment with targeted biologic agents such as cytokine inhibitors. Here, we report a patient from a founder population who presented with early-onset recurrent fevers, rash, subcutaneous nodules, lipodystrophy, and arthritis suggestive of CANDLE (Chronic Atypical Neutrophilic Dermatositis with Lipodystrophy and Elevated Temperature) syndrome. Based on the phenotypic similarities with CANDLE, we suspected that the patient might have an interferon-mediated disease. We found an upregulated type I interferon gene expression signature in peripheral blood, high basal levels of phospho-STAT1 in IFN α -stimulated PBMCs, and increased production of many inflammatory cytokines in serum and supernatants of stimulated primary cells. Similar to CANDLE, proteasome activities, especially the chymotrypsin activity, were significantly reduced in patient's EBV-B cells. After failing to identify mutations in genes associated with CANDLE, WES was performed and identified a novel predicted deleterious homozygous candidate variant within a poorly characterized deubiquitinase, *USP43* (p.G837K), present only in the proband. Deubiquitinases catalyze removal of ubiquitin chains from target proteins, many of which are marked for proteasome-mediated protein degradation. Analysis of stimulated patient's EBV-B cells and fibroblasts, or 293T cells depleted of *USP43*, led to increased levels of intracellular ubiquitinated (Ub) proteins, and this cellular phenotype was rescued by transfection of wild-type *USP43*. TNF-stimulated fibroblasts displayed higher levels of K48 ubiquitin chains, which suggests a dysregulation in proteasome function. The patient has responded well to treatment with JAK inhibitors that are known to ameliorate interferon-mediated diseases. Thus, mutations in *USP43* lead to a novel interferonopathy and *USP43* may function as a novel regulator of proteasome assembly.

996W

Clinical and pathophysiological characterization of a novel form of ectodermal dysplasia with anhidrosis and immunodeficiency (EDA-ID). M. Cuk^{1,2}, J. Lian³, S. Kahlfuss³, M. Maus³, L. Kozhaya⁴, M. Vaeth³, F. Rieux-Laucat^{5,6}, C. Picard^{6,7}, M.J. Benson⁸, A. Jakovcovic¹, K. Bilic¹, I. Martinac¹, P. Stathopoulos⁸, I. Kacsokovics⁹, T. Vraetz¹⁰, C. Speckmann^{10,11}, S. Ehli^{10,11}, T. Issekutz¹², D. Unutmaz⁴, B. Patel¹³, M. Ouimet¹⁴, U. Kaufmann³, J. Yang³, R. Horvath¹⁵, H.T. Hornig-Do¹⁵, Z. Chrzanowska-Lightowlers¹⁵, K.J. Moore¹⁴, A.M. Cuervo¹³, A.R. Concepcion³, S. Feske³. 1) University Hospital Centre Zagreb, Zagreb, Croatia; 2) University of Zagreb School of Medicine, Zagreb, Croatia; 3) New York University School of Medicine, New York, USA; 4) The Jackson Laboratory for Genomic Medicine, Framington, USA; 5) INSERM UMR Laboratory of the Immunogenetics of Pediatric Autoimmune Diseases, Paris, France; 6) INSERM UMR Imagine Institute, Paris Descartes-Sorbonne, Paris Cite University, Paris, France; 7) The Study Center for Primary Immunodeficiencies, Necker-Enfants Malades Hospital, Necker Medical School, Paris, France; 8) Schulich School of Medicine and Dentistry, Western University London, London, United Kingdom; 9) ImmunoGenes, Budapest, Hungary; 10) Center for Pediatrics, University of Freiburg, Freiburg, Germany; 11) Center for Chronic Immunodeficiency, Medical Centre, University of Freiburg, Freiburg, Germany; 12) Dalhousie University, Halifax, New Scotia, Canada; 13) Department of Developmental and Molecular Biology, Albert Einstein College of Medicine, Bronx, NY 10461, USA; 14) Department of Medicine, New York University School of Medicine, New York, NY 10016, USA; 15) Wellcome Trust Centre for Mitochondrial Research, Institute of Neuroscience, Newcastle University, Newcastle upon Tyne NE2 4HH, UK.

We present 4 patients with EDA-ID, autoimmunity, myopathy, mitochondrial dysfunction and disturbed lipid metabolism caused by novel *null* mutations in *ORA1* (p.V181SfsX8, p.L194P, p.G98R) and abolished Store-Operated Ca²⁺ Entry (SOCE). Patients had dry and exfoliate skin, thin and fragile hair, enamel defects, attacks of facial flushing, tachycardia, tachypnea and hypertension, anhidrosis with heat intolerance due to Ca²⁺-activated chloride channel TME-M16A dysfunction suggesting that SOCE is important for eccrine sweat glands function. Patients were prone to infections, especially with viruses (CMV, EBV, RSV, rotavirus). SOCE defects were associated with T cell dysfunction and reduced numbers of invariant NK T cells and regulatory FOXP3+ Treg cells, and altered composition of $\gamma\delta$ T cell and NK cell subsets. Patients had lymphadenopathy, hepatosplenomegaly, autoimmune-mediated pancytopenia and antiphospholipid syndrome, loss of naive CD45RA+ T cells and concomitant expansion of CD45RO+ or HLA-DR+ activated T cells. Patients also had reduced numbers of CD25+FOXP3+ Treg cells suggesting that SOCE is required for immunological tolerance. At the cellular level, we observed dysmorphic and dysfunctional mitochondria in cells of SOCE-deficient patients that were characterized by reduced mitochondrial volume and abnormal cristae structure, reduced expression of mitochondrial proteins, increased basal mitochondrial membrane potential but decreased electron and proton transport and superoxide production. Furthermore, we found reduced free fatty acid oxidation and decreased expression of neutral lipases, resulting in accumulation of lipid droplets in cells of SOCE deficient patients. Collectively these findings demonstrate severe mitochondrial dysfunction in the absence of SOCE, which likely contributes to the myopathy, EDA and immunodeficiency in our patients. The characterization of the pathophysiological mechanisms underlying EDA-ID contributes to the early recognition of these patients and potential new treatment options such as mitochondrial cocktail and regulated diet providing more favorable outcomes.

997T

A severe case of Aicardi-Goutières Syndrome-7 with novel pathogenic variant in IFIH1. T. Kaname¹, S. Amari², M. Iso¹, K. Hata³, K. Okamura⁴, N. Kobayashi¹, K. Yanagi¹, A. Ishiguro⁵, Y. Matsubara⁶. 1) Genome Medicine, National Center for Child Health and Development (NCCHD), Tokyo, Tokyo, Japan; 2) Neonatology, NCCHD, Tokyo, Tokyo, Japan; 3) Maternal-Fetal Biology, NCCHD, Tokyo, Tokyo, Japan; 4) Systems Biomedicine, NCCHD, Tokyo, Tokyo, Japan; 5) Hematology, NCCHD, Tokyo, Tokyo, Japan; 6) Research Institute, NCCHD, Tokyo, Tokyo, Japan.

Aicardi-Goutières syndrome (AGS) is a rare inflammatory disorder characterized by early onset encephalopathy and neurologic impairment. Till now, up to the present seven different subtypes of causative genes are known. Type 7 (AGS 7) presents with delayed psychomotor development, axial hypotonia, spasticity and brain imaging abnormalities, and is milder than other subtypes. Here we report a very severe case of AGS 7 who had fetal dysfunction and died of 3 months after birth with a combination of prolonged anemia and thrombocytopenia after birth. The patient was pointed out fetal heart expansion and significant pericardial fluid at 29 weeks of pregnancy, was diagnosed as fetal dysfunction together with Non-Stress Test, and was delivered by emergency caesarean section. After birth he showed symptoms of severe inflammation. There was no obvious calcified lesion in the brain by ultra sound but Lenticulostriate vasculopathy was observed. Neither Toxoplasma, parvovirus, rubella virus, cytomegalovirus, or herpes simplex virus was detected in the blood by PCR. Platelet transfusion almost every day since birth, nearly every day from the birth two days after birth required red blood cell transfusion due to severe inflammation. The effect of dexamethasone and vasodilator administration to him was not enough and he died of three month after birth. Whole exome analysis with trio was performed using the SureSelect Human All Exon V6 kit for capturing and a HiSeq2500 for sequencing with 101-bp paired-end reads. Then, a novel variant of the IFIH1 gene, NM_022168:c. A2439T:p.E813D, was found in the patient. The variant was validated by Sanger sequencing and was *de novo*. The variant is at the Helicase domain and is predicted as disease causing by PolyPhen-2, SIFT, and mutation tasting software. We concluded that AGS 7 might have a phenotypic heterogeneity from mild case to severe case.

998F

Beyond whole exome sequencing--molecular diagnosis for patients of primary immunodeficiency diseases who failed to have causal mutations identified after whole exome sequencing approach. J. Yang, W. Yang, P.P.W. Lee, Y.L. Lau. University of Hong Kong, Hong Kong, China.

Eighty-eight primary immunodeficiency diseases (PID) patients who have gone through whole exome sequencing (WES) but failed to have causal mutations identified are reexamined through systematic reanalysis of WES data, resequencing on the most updated PID genes with whole gene coverage, whole genome sequencing and RNA-seq, as well as PID-specific algorithms on data analysis. WES data are reevaluated to reflect the newest understanding on PID diagnosis and modified strategies. For those still without definitive diagnosis, a PID gene panel are designed and captured in their entirety, followed by deep sequencing to increase sensitivity. A data analysis algorithm is also being developed, which considers sequence specificities of each PID gene. For those still being negative on causal mutations, whole genome sequencing and/or RNA-seq are applied. Functional annotations and analysis algorithms are developed to help identify novel PID genes, making use of information such as immune related functions, protein interaction with known PID genes, phenotypes in animal models, expression and pathway analysis, and regulation patterns based on ENCODE data. These approaches are currently evaluated in terms of potential increase in detection sensitivity for PID patients who failed WES.

999W

A novel CYCS mutation in the α -helix of the CYCS C-terminal domain causes non-syndromic thrombocytopenia. Y. Uchiyama^{1,2}, K. Yanagisawa³, S. Kunishima⁴, M. Shiina⁵, Y. Ogawa³, M. Nakashima^{1,6}, S. Miyatake¹, S. Mitsuhashi¹, A. Takata¹, N. Miyake¹, K. Ogata⁵, H. Handa³, N. Matsumoto¹, T. Mizuguchi¹. 1) Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Kanagawa, Japan; 2) Department of Oncology, Yokohama City University Graduate School of Medicine, Yokohama, Japan; 3) Department of Hematology, Gunma University Graduate School of Medicine, Gunma, Japan; 4) Department of Medical Technology, Gifu University of Medical Science, Seki, Japan; 5) Department of Biochemistry, Yokohama City University Graduate School of Medicine, Yokohama, Japan; 6) Department of Biochemistry, Hamamatsu University School of Medicine, Hamamatsu, Japan.

We report a patient with thrombocytopenia from a Japanese family with hemophilia A spanning four generations. Various etiologies of thrombocytopenia, including genetic, immunological, and hematopoietic abnormalities, determine the prognosis for this disease. In this study, we identified a novel heterozygous mutation in a gene encoding cytochrome c, somatic (CYCS, MIM123970) using whole exome sequencing. This variant (c.301_303del:p.Lys101del) is located in the α -helix of the cytochrome c (CYCS) C-terminal domain. *In silico* structural analysis suggested that this mutation results in protein folding instability. CYCS is one of the key factors regulating the intrinsic apoptotic pathway and the mitochondrial respiratory chain. Using the yeast model system, we clearly demonstrated that this one amino acid deletion (in-frame) resulted in significantly reduced cytochrome c protein expression and functional defects in the mitochondrial respiratory chain, indicating that the loss of function of cytochrome c underlies thrombocytopenia. The clinical features of known CYCS variants have been reported to be confined to mild or asymptomatic thrombocytopenia, as was observed for the patient in our study. This study clearly demonstrates that thrombocytopenia can result from CYCS loss-of-function variants.

1000T

Determination of causal variants in inherited thrombocytopenias. M.

Pesova¹, K. Stano Kozubik^{1,2}, L. Radova¹, K. Pal¹, K. Reblova¹, J. Trizuljak^{1,2}, H. Urbankova³, P. Smejkal⁴, J. Gumulec⁵, V. Fiamoli⁶, S. Pospisilova^{1,2}, M. Doubek^{1,2}. 1) Central European Institute of Technology (CEITEC), Masaryk University, Brno, Czech Republic; 2) Department of Internal Medicine, Hematology and Oncology, University Hospital and Faculty of Medicine, Brno, Czech Republic; 3) Department of Hematologic Oncology, University Hospital and Palacký University, Olomouc, Czech Republic; 4) Department of Clinical Hematology, University Hospital, Brno, Czech Republic; 5) Department of Hematologic Oncology, University Hospital, Ostrava, Czech Republic; 6) Department of Pediatric Hematology, University Hospital, Brno, Czech Republic.

Inherited thrombocytopenias (IT) are a heterogeneous group of rare disorders. To date, forty genes have been described as being associated with IT. Pathogenic variants in these genes usually lead to the disruption of megakaryopoietic and thrombopoietic processes and present as the thrombocytopenia phenotype. Patients are occasionally misdiagnosed with the more common idiopathic thrombocytopenic purpura and unsuccessfully treated with steroid therapy and splenectomy. Additionally, IT patients are often identified incidentally during blood examination in adulthood, due to the absence of spontaneous or life-threatening bleeding that is a usual characteristic of IT. In some patients, accurate diagnosis of IT can only be established based on the results of molecular genetic testing. In our patient cohort, we identified four families with at least two thrombocytopenia patients. Proband from two of the families presented with macrothrombocytopenia and probands from the other two families presented with normal platelet size and volume; one of them had hematological disease in their medical history. Germline DNA analysis was performed on all available samples and somatic DNA analysis was done for the oncological patient. Sequencing libraries were prepared according to the NimblegenSeqCap EZ Exome v3 protocol and sequencing was performed on NextSeq500. Within each family, the obtained variants were compared between the individuals with thrombocytopenia phenotype and their disease-free relatives. Using Exome sequencing, we identified a unique variant segregating with thrombocytopenia phenotype for each of the four families: *GP1BA*:NM_000173 exon2:c.176T>G:p.L59R and exon2:c.98G>A:p.C33Y, *CYCS*:NM_018947 exon2:c.59C>T:p.T20I and *RUNX1*:NM_001754 exon8:c.866delG:p.G289fs. *In silico* analysis revealed a structural defect of amino acid changes found in the protein structures, indicating that all of the variants are likely pathogenic. We identified a causal variant for each of the families analysed. Accurate diagnosis of IT would allow clinicians to conduct further examinations if the identified variant poses additional risk to the carrier, i.e., a higher risk of oncological disorders. Thus, patients could receive appropriate treatment and avoid the unnecessary side effects from long-term interventions such as steroids and splenectomy. The research was done according to the Declaration of Helsinki. Supported by Ministry of Health of the Czech Republic, grant No 16-29447A.

1001F

Genetics and pathogenesis of familial multiple sclerosis. C. Vilari-

no-Guell¹, M. Encarnacion¹, C. Bernales¹, I. Yee¹, M. Criscuoli¹, A. Traboulsee², Z. Wang³, B. Herculano³, W. Song³, A. Zimprich⁴, E. Reinthaler⁵, A. Deutschländer⁶, A.D. Sadovnick^{1,2}. 1) University of British Columbia, Vancouver, BC, Canada; 2) Division of Neurology, Faculty of Medicine, University of British Columbia, Vancouver, Canada; 3) Townsend Family Laboratories, Department of Psychiatry, University of British Columbia, Vancouver, Canada; 4) Department of Neurology, Medical University of Vienna, 1090 Austria; 5) Department of Neurology, Department of Clinical Genomics, and Department of Neuroscience, Mayo Clinic Florida, USA.

Background: Multiple sclerosis (MS) is an inflammatory autoimmune disease of the central nervous system characterized by myelin loss and neuronal dysfunction. Although the majority of patients do not present a family history of MS, the prevalence of familial aggregation has been estimated at 12.6% globally. Mutations identified in families are likely to account for the highest attributable risk towards disease, and provide a better understanding of the biological processes underlying the genesis of MS. This is critical for the development of therapeutic options addressing the cause of disease, in addition to managing the clinical symptoms, as currently available treatments primarily target inflammatory components and have a modest impact on disease progression. **Methods:** We implemented whole-exome sequencing (WES) in 132 MS patients from 34 families for the identification of disease-causing mutations. Candidate variants were assessed for co-segregation with MS within families, and genotyped in a large multi-ethnic cohort that includes over 13,000 samples from 2,054 MS families, to confirm pathogenicity. **Results:** WES analysis of MS families identified 12 mutations co-segregating with disease in genes of the fibrinolysis and complement pathways (*PLAU*, *MASP1*, *C2*), inflammasome assembly (*NLRP12*), Wnt signaling (*UBR2*, *CTNNA3*, *NFATC2*, *RNF213*), nuclear receptor complexes (*NCOA3*), and cation channels and exchangers (*KCNQ4*, *SLC24A6*, *SLC8B1*). Biological characterization of these mutations suggest a disruption of interconnected innate immunity and pro-inflammatory pathways as the initial events in the pathophysiology of familial MS. **Conclusions:** Our data demonstrates the existence of Mendelian forms of MS, which can be attributed to a single mutation of major effect that is largely responsible for the onset of MS and its transmission across generations. The genes identified in this study play critical roles in cellular cation homeostasis, and the regulation of transcription and activation of inflammatory mediators; suggesting a disruption of the innate immune system as the biological mechanism of disease. Identified mutations highlight targets for therapeutic development, and provide the foundations for the generation of cellular and animal models based on human genetic etiology in which to assess their efficacy. In addition, screening for these mutations in MS patients and healthy family members can be used for disease confirmation, risk prediction and prognosis.

1002W

Expedited whole exome sequencing diagnoses activated PI3K-delta syndrome-1 in an infant presenting with chronic diarrhea. *J.P. Schacht¹, C. Umandap¹, A. Carey², A. Iglesias².* 1) Department of Pediatrics, Division of Clinical Genetics, Columbia University Medical Center, New York City, NY; 2) Department of Pediatrics, Division of Pediatric Critical Care Medicine, Columbia University Medical Center, New York City, NY.

Introduction: Chronic diarrhea presents a diagnostic challenge due to the robust differential diagnosis. Diagnosis of rare etiologies frequently involves time consuming and invasive approaches including endoscopy and biopsy, which may be nondiagnostic, or difficult to implement in acutely ill patients. Expedited whole exome sequencing (WES) may be a more feasible diagnostic tool. We present a case of an infant presenting with chronic diarrhea diagnosed with activated PI3K-delta syndrome-1 via expedited WES. **Case Description:** A 36 day old ex full term male presented with a one day history of fever, and worsening of chronic diarrhea since birth. Initial exam was notable for a nondysmorphic infant with respiratory distress, sunken fontanel, and height, weight, and head circumference all less than the 5th centile. Initial labs showed hyperchloremic, hypernatremic, metabolic acidosis with an elevated anion gap. Infectious evaluation with respiratory, GI, and CSF PCR was negative. Differential diagnoses considered included infection, formula intolerance, inborn errors of metabolism, cystic fibrosis, and other genetic diseases. Metabolic studies showed ketosis but were otherwise normal. Hospital course was complicated by dural venous sinus thrombosis. Based on all of the above, an expedited WES was ordered as part of an ongoing study. **Methods/Results:** Expedited trio WES, with turnaround time of 16 days, was completed at a commercial lab using current standardized NGS technology, and demonstrated a de novo missense variant, p.Leu855Arg (CTG>CGG):c.2564 T>G, in exon 20 of the *PIK3CD* gene, classified as likely pathogenic following ACMG guidelines for variant interpretation, as it has not been reported previously, causes non-conservative amino acid substitution, and in-silico analyses support a deleterious effect. *PIK3CD* pathogenic variants cause activated PI3K-delta syndrome-1 (APDS1). **Conclusions:** Expedited WES results were consistent with APDS1, which is associated with chronic diarrhea in a minority of patients. As no other etiology for diarrhea was found, we believe these results explain the phenotype. This diagnosis was reached much faster than the typical turnaround time for routine WES. These results quickly clarified the etiology of the diarrhea and demonstrated the need for immunologic evaluation. This case proves that expedited WES can aid the diagnosis and management of acutely ill patients.

1003T

Gene expression analysis of endothelial cells from sickle cell anemia patients and its relation with stroke. *M.B. Melo¹, M.T. Ito¹, S.M. Silva-Costa¹, L.C. Baptista¹, G.Q. Carvalho-Siqueira¹, D.M. Albuquerque², F. Cendes³, S. Ospina-Prieto³, C. Lanaro³, V.M. Rios¹, M.C. Ozello², F.F. Costa².* 1) CBMEG, University of Campinas, Campinas, São Paulo, Brazil; 2) Hematology and Hemotherapy Center, University of Campinas, Campinas, São Paulo, Brazil; 3) Neuroimaging Laboratory, University of Campinas, Campinas, São Paulo, Brazil.

Sickle-cell anemia (SCA) results from the homozygosity of a single mutation at position 6 of the β -globin locus. The clinical aspects of this disease are very heterogeneous and hence different patients may present significantly different clinical evolution. Almost all organs can be affected, with emphasis on the Central Nervous System (CNS), where transient ischemic events, infarcts and cerebral haemorrhage are observed, affecting approximately 25% of the patients with SCA. Although sickle cell anemia is characterized by falcization of red blood cells, the vascular endothelium plays an important role in the pathophysiology of this hemoglobinopathy. Differences in the expression of molecules produced by endothelial cells may be associated with clinical heterogeneity among individuals affected by vascular diseases. In this study, we investigated the differential expression of genes involved in the endothelial cell biology in patients with SCA with and without stroke. Endothelial cells from four patients with SCA with stroke and six patients with SCA without stroke (confirmed by magnetic resonance imaging) were evaluated. The approach included the isolation and culture of BOECs (blood outgrowth endothelial cells) and expression analysis of 84 genes by PCR Array (Human Endothelial Cell Biology RT² PCR Array). The analysis of the array showed that 18 genes were upregulated and 11 were downregulated, from which 10 have been chosen for validation (5 upregulated - *FLT1*, *IL6*, *MMP1*, *PGF* and *PROCR* and 5 downregulated - *CCL2*, *ICAM1*, *KDR*, *SELPLG* and *TNFSF10*). *PGF*, *MMP1*, *ICAM1* and *TNFSF10* genes have been validated by qRT PCR. Among these genes, *MMP1* is noteworthy once its expression was 204.64 higher in patients with stroke than in patients without stroke ($p = 0.000428$). In stroke, *MMPs* may be involved in damage of the blood-brain barrier, degradation of components of the extracellular matrix (ECM), facilitation of cellular migration through the ECM, cleavage of cytokine precursors from the cell surface and exacerbation of post ischemic edema. These results may help to understand the mechanisms involved in the development of stroke in patients with the disease, in the preventive diagnosis and also contribute to the employment of useful approaches for their treatment.

1004F

Single cell transcriptomes of sickle cell disease PBMC identifies markers of disease severity in distinct cell populations. S. Lessard, H. Ling, D. Levasseur. Bioverativ, Waltham, MA.

Each year, around 300,000 individuals are born with sickle cell disease (SCD), a recessive genetic disorder caused by a single base pair mutation in the β -globin gene. SCD patients display altered circulating immune cell populations and reticulocyte counts, which can complicate transcriptome studies of peripheral mononuclear cells (PBMCs). Here, we sequence the transcriptome of PBMCs from healthy and SCD donors using a single-cell approach to account for this heterogeneity. We purified PBMCs from 2 SCD and 4 healthy donors. Single cell library preparation using the 10X Genomics 3' gene expression assay (v2) and high throughput sequencing was performed at Johns Hopkins University genetics resource core facility. We processed reads using Cell Ranger (v2.1.0) and used Seurat (v2.2.1) to aggregate all 6 datasets and to perform differential gene expression analyses. In total, we sequenced 24,524 cells with a mean of 85,034 reads/cell (1.8×10^5 total reads) achieving on average ~92% sequencing saturation. We identified a mean of 573 genes per cell. We found that SCD samples displayed a greater T cell fraction (53%), along with fewer monocyte counts (20%) compared to PBMCs from healthy volunteers (32% and 42% respectively). One SCD sample, "SCDB", exhibited particularly altered PBMC fractions, with few monocytes (6%) and greater T cell (65%) and B cell fractions (22%). SCD cell clusters showed a global down-regulation of immunity-related genes. Monocytes displayed increased expression of the ferritin heavy chain gene *FTH1*, a marker of iron balance, and *GPX1*, a marker of oxidative stress. Cytotoxic T-cells and natural killer cells showed increased expression of *S100B*. *S100B* serum levels are increased following stroke or brain hypoxia. Finally, we found that monocytes from SCDB showed strong up-regulation of *IL8*, which has been proposed as a marker of sickle cell crisis. We sequenced 6 PBMC transcriptomes from healthy and SCD patients at the single cell level. PBMCs from SCD samples showed altered PBMC populations compared to healthy donors. Monocytes showed the largest changes in gene expression in SCD samples, expressing markers of iron overload and SCD complications. In conclusion, single cell transcriptomics can help account for different cell composition and identify cell-specific gene expression changes that could possibly be markers of disease.

1005W

A new mechanism and vulnerability for fibrosis. M.A. Seman-Senderos¹, D.H. Kim¹, J.D. Beckett¹, V. Nagpal¹, R.A. Gould^{1,2}, T.J. Creamer³, Y. Chen¹, D. Bedja⁴, J.T. Butcher⁵, W. Mitzner⁶, R. Rouff⁷, S. Hata⁸, D.S. Warren⁹, H.C. Dietz^{1,7,8,9}. 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; 2) Meinig School of Biomedical Engineering, Cornell University, Ithaca, New York; 3) Department of Surgery, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; 4) Department of Cardiology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; 5) Department of Environmental Health Sciences, Johns Hopkins University Bloomberg School of Public Health, Baltimore, Maryland, USA; 6) Department of Advanced Science for Biomolecules, Tokyo Metropolitan Institute of Medical Science, Tokyo, Japan; 7) Howard Hughes Medical Institute, Chevy Chase, Maryland, USA; 8) Department of Pediatrics, Division of Pediatric Cardiology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; 9) Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA.

Fibrosis is a common pathologic outcome of chronic disease resulting in the replacement of normal tissue parenchyma with a collagen-rich extracellular matrix. The mechanisms by which transforming growth factor β (TGF β) – the dominant profibrotic cytokine – drives fibrosis are varied but include obligate mesenchymal transition (MT) of progenitor cells to invasive and synthetic myofibroblasts. In a candidate culture-based screen we identified dimeric calpain activity as an essential effector of TGF β -induced MT using a variety of progenitors. Furthermore, calpain activity was also required for maintenance of myofibroblast differentiation and performance despite ongoing exposure to TGF β . siRNA-based gene silencing demonstrated the specific requirement for induction of expression of calpain 9 (CAPN9), an isoform previously thought to be restricted to the gastrointestinal tract. Mice lacking functional CAPN9 due to biallelic targeting of *Capn9* were viable and fertile, but showed overt protection from bleomycin-induced lung fibrosis, carbon tetrachloride-induced liver fibrosis, and angiotensin II-induced cardiac fibrosis and dysfunction. In order to determine if CAPN9 inhibition had the potential to modify a chronic and naturally-occurring (as opposed to acute and chemically-provoked) fibrotic predisposition, we assessed for therapeutic modification of arrhythmogenic cardiomyopathy (ARC) – caused by homozygosity for targeted *Dsg2* alleles – on the *Capn9*-null background. ARC mice, as with affected patients, show early onset and highly penetrant myocardial fibrosis in association with progressive heart failure. CAPN9 deficient ARC mice showed dramatic protection from heart fibrosis, as assessed by histomorphometric and biochemical analyses, in association with complete preservation of myocardial function, as assessed by echocardiographic measurement of ventricular size, ejection fraction and shortening fraction. Database analyses revealed that a human SNP that disrupts the obligate splice acceptor for exon 4 in CAPN9, leading to out-of-frame exon skipping and predicted loss-of-function, is common in Southeast Asia, with the frequency of homozygosity (~1%) matching the prediction of Hardy-Weinberg equilibrium; together with the highly spatially-restricted pattern of CAPN9 expression under physiologic circumstances and the heartiness of the murine knockout, these data provide a strong signature for tolerance of therapeutic strategies aimed at CAPN9 antagonism.

1006T

Type IA isolated growth hormone deficiency due to *GH1* gene complete homozygous deletion. M. Manotas^{1,2}, E. Castaño^{1,2}, R. Quero^{1,2}, A. Paredes^{1,2}, C. Cespedes^{1,2}, F. Suarez^{1,2}. 1) Pontificia Universidad Javeriana, Bogotá D.C., Colombia; 2) Hospital Universitario San Ignacio.

Isolated GH deficiency type IA (IGHDIA [MIM 262400]) is an infrequent cause of severe congenital growth hormone deficiency. Underlying genetics are predominantly *GH1* gene deletions, although nonsense or frameshift mutations have also been reported. *GH1* gene is located on 17q22–24, within a cluster of 65 kb including five homologous genes that predispose to deletions after unequal recombination and crossing over at meiosis. The deletion of the *GH1* gene of 6.7 kb is involved in 76% to 83% of the reported cases. The appearance of neutralizing anti-GH antibodies (anti-GH Abs) upon recombinant human GH (rhGH) treatment is a characteristic feature of IGHDIA. We present the case of A 16-year-old male who was referred to our genetic unit after several years of lapsed follow-up. The patient was diagnosed in early childhood with severe growth retardation (–8.89 S.D. at 9 months years of age). Birth length at 35 weeks of gestation was already poor with and height of 43.6 cm and weight of 1524 gm. In infancy, he had a delay in the motor development. Hormonal analysis revealed undetectable IGF-1 and serum GH concentration was below the detection limit even after administration of clonidine. Other pituitary hormonal axis functions were normal, and the pituitary MRI did reveal the gland in a normal position, however relatively small. rhGH supplementation was started at the 18-month years, but no significant increase in growth was achieved, compatible with the development of anti-GH Abs. Directed del/dup analysis revealed a complete homozygous *GH1* gene deletion. Actually, his height is 83 cm (–8.75 S.D.), he has a disproportionately short stature with particular phenotypic features such as immature facies, slightly down-slanting palpebral fissures, depressed bridge of nose, absence of secondary sexual characteristics and mildly impaired cognition. He also suffers from asthenia and osteomuscular pain. The patient described herein is, to our knowledge, the first case of IGHDIA due to *GH1* gene complete homozygous deletion described in the literature. GH complete absence and the consequent severe IGF1 deficiency may underlie the patient's present puberty delay, osteomuscular symptoms, impaired cognition and youthful face. It should be noted, that although evidence about growth hormone and aging is currently controversial, the juvenile face, noncorresponding with his chronological age, can be a clinical signal of delayed aging associated with GH deficiency.

1007F

OFD1 and autophagy: Implication for renal cystic disease. M. Manuela^{1,2}, U. Formisano¹, S. Brillante¹, D. Iaconis¹, A.S. Maione¹, E. Damiano¹, R. Tammaro¹, C. Settembre^{1,2}, B. Franco^{1,2}. 1) Telethon Institute of genetics and Medicine, Pozzuoli, Naples, Naples, Italy; 2) Medical Genetics Unit, Department of Translational Medicine, University of Naples "Federico II", Naples, Italy.

The OFD1 protein is codified by the gene mutated in Oral-facial-digital type I syndrome (OFD type I), an X-linked ciliopathy characterized by abnormalities of face, oral cavity and digits, involvement of the central nervous system and renal cystic disease. OFD1 is a centrosomal/basal body protein necessary for primary cilia formation. Interestingly, recent data established a link between autophagy, cilioproteins and renal cystic disease (CK). In particular it has been shown that autophagy activation rescues renal cystogenesis in autosomal dominant polycystic kidney disease. Mass spectrometry analysis revealed, among putative OFD1 interactors, autophagy related proteins and we thus sought to determine whether OFD1 might have a role in this process. We demonstrated that OFD1-inactivated human renal cells (KO-OFD1) show increased autophagic flux and autophagosome biogenesis, by analysis and quantification of a) LC3 and WIPI2 puncta; b) autophagosome-lysosome fusion rates; d) rate of delivery of autophagosomes to lysosomes using an RFP-GFP tandem tagged LC3 protein. The specificity of our findings was supported by the observation that OFD1 overexpression rescues the autophagic phenotypes observed *in vitro* and results in decreased autophagosome biogenesis in renal control cells. Using a variety of approaches we also determined that the increased autophagic flux is not due to decreased mTORC1 activity. These data suggest that OFD1 acts as inhibitor of the autophagy. In addition, we demonstrated *in vivo* enhanced autophagic flux both at precystic and cystic stages in different *Ofd1*-inactivated models. Finally, to test the role of autophagy modulation on CK *in vivo* we achieved conditional inactivation in the kidney of both *Ofd1* and *Atg7*, an essential gene for autophagy. Histological analysis showed a significant reduction in the number and size of cysts in *cre^{Ksp};Ofd1^{+/+};Atg7^{+/+}* mutants compared to *cre^{Ksp};Ofd1^{+/+};Atg7^{+/+}* mice, suggesting that the increased autophagy might be strictly associated to renal cystogenesis in OFD type I syndrome. These data suggests that alterations of autophagy may be a common pathogenic mechanism in CK. The dissection of the molecular mechanisms underlying the initial phases of renal cyst formation in OFD type I could allow elucidating the role of autophagy in CK and could disclose new therapeutic avenues for renal cystic disease. .

1008W

Investigating *PRSS1* copy number variants showing MLPA peak height ratio sensitivity. M.J. Deshotel¹, G. Pont-Kingdon¹, E. Lyon^{1,2}. 1) Molecular Genetics, ARUP Laboratories, Salt Lake City, UT; 2) Department of Pathology, University of Utah, Salt Lake City, UT.

Hereditary chronic pancreatitis (HCP [MIM 167800]) is caused by pathogenic variants in four or more genes. One gene, *PRSS1*, codes for cationic trypsinogen, and pathogenic variants in this gene cause premature activation of trypsinogen within the pancreas, leading to pancreatic calcifications and exocrine insufficiency. While Sanger sequencing identifies the causal mutation in about 80% of HCP patients, about 6% of cases are due to increased copy number of *PRSS1*. Our laboratory is evaluating the multiplex ligation-dependent probe amplification (MLPA) kit from MRC-Holland to detect *PRSS1* and *SPINK1* copy number variants in HCP patients. We first tested whether 9 common polymorphisms and pathogenic SNPs located under MLPA probe binding sites would obstruct detection; we found no evidence of interference. However, in 3 out of 21 wild-type patient samples tested during assay development, we observed lower than average peak height ratios around 0.8 in the 6 *PRSS1* probes. *PRSS1* deletions are not pathogenic, but to avoid potential misinterpretation of results in patient samples, we explored possible explanations for the reduced ratios. We considered pseudogene interference, differences in DNA elution buffers, and lower input concentrations for MLPA. Probe competition with the pseudogene *PRSS3P2/TRY6* was a possible cause for the lower ratios because it shares 90.1% identity with *PRSS1* and deletions have been described in many different ethnicities. To test this hypothesis, we designed and tested qPCR primers specific to *PRSS1* and *TRY6*. The qPCR results did not reveal differences in *PRSS1* or *TRY6* copy number, but did reflect the MLPA data, with average $2^{-\Delta\Delta C_T}$ gene dosage ratios of 0.72 for both *PRSS1* and *TRY6* measured in the samples with low *PRSS1* peak height ratios. To further exclude pseudogene interference, we tested Coriell DNA samples with the rs1323002 SNP, which was previously shown to be in 100% linkage disequilibrium with *TRY6* deletion. When tested with MLPA, these samples did not show a dose-dependent *TRY6* influence on the *PRSS1* probes. We also performed ethanol precipitations, which increased the MLPA ratios from an average of 0.81 to 0.86, closer to expected values. Additionally, running MLPA on the aberrant samples with low input concentrations of DNA (10 ng/μL) did not resolve the low *PRSS1* ratios. Our results highlight the importance of carefully choosing wild-type reference samples in clinical laboratories.

1009T

***FOG2* and *GATA4* haploinsufficiency is a novel cause for primary ovarian insufficiency.** D. Baetens¹, A. Bachelot², N. De Vriese³, F. De Zegher⁴, N. Reynaert⁵, R. Colombo⁶, Y. Van Bever⁷, J. Dulon², P. De Sutter⁶, P. Touraine², M. Cools¹, E. De Baere¹. 1) Center for Medical Genetics Ghent, Ghent University and Ghent University Hospital, Ghent, Belgium; 2) UPMC Université Pierre et Marie Curie, Université Paris, Paris, France; 3) Pediatric Endocrinology, University Hospital Gasthuisberg, University of Leuven, Leuven, Belgium; 4) Center for the Study of Rare Hereditary Diseases, Niguarda Ca' Granda Metropolitan Hospital, Milan, Italy; 5) Department for Clinical Genetics, Erasmus Medical Center Rotterdam, Rotterdam, The Netherlands; 6) Department of Obstetrics and Gynaecology, Ghent University and Ghent University Hospital, Ghent, Belgium; 7) Department of Pediatric Endocrinology, Ghent University and Ghent University Hospital, Ghent, Belgium.

Background: Primary ovarian insufficiency (POI) is a major cause of female infertility. Although many POI genes have been identified, the genetic cause remains unknown in the majority of cases. The starting point of this study was a 3.2 Mb deletion encompassing *Friend of GATA 2 (FOG2)* in a patient with a 46,XY disorder of sex development (DSD). This deletion was inherited from his mother, displaying low levels of serum anti-Müllerian hormone (AMH), suggestive of a decreased ovarian reserve. Based on this human case and supported by the role of the *Fog2/Gata4* complex in murine ovarian development, we hypothesized that haploinsufficiency of the *FOG2* and the *GATA binding protein 4 (GATA4)* genes could contribute to the pathogenesis of POI. **Methods:** Targeted next-generation sequencing of *FOG2* and *GATA4* was conducted in a Belgian-French POI cohort (n=388). Functional validation of the identified variants consisted of luciferase assays, co-immunoprecipitation of mutated *FOG2* and *GATA4* and *in silico* structural analysis. **Results:** A heterozygous deletion was found in a child with isolated 46,XY DSD and his mother with presumed POI. Subsequent targeted resequencing of *FOG2* and *GATA4* in a Belgian-French POI cohort revealed monoallelic likely pathogenic variants in *FOG2* (n=5) and *GATA4* (n=4). An altered transcriptional activation of the mutated *FOG2/GATA4* complex was shown for several of the identified variants, substantiated by co-immunoprecipitation and *in silico* structural analyses. **Conclusion:** Our study put forward *FOG2* and *GATA4* as novel POI genes, supporting a role for both genes in human ovarian maintenance. We expand the phenotypic spectrum of *FOG2* and *GATA4* haploinsufficiency, thus far only associated with cardiac anomalies, congenital diaphragmatic hernia and 46,XY DSD.

1010F

Germline genetic causes of pediatric Cushing's disease: Findings in a large single-center cohort. L.C. Hernández Ramírez¹, F.R. Faucz¹, M.B. Lodish¹, N. Pankratz², P. Chittiboina³, J. Lane⁴, D.M. Kay⁴, C. Tatsi⁴, G. Trivelpin¹, F. Chasseloup¹, J.L. Mills⁵, C.A. Stratakis¹. 1) Section on Endocrinology and Genetics, Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), National Institutes of Health (NIH). 10 Center Drive, CRC, Room 1E-3216, Bethesda, MD 20892-1862, USA; 2) Department of Laboratory Medicine and Pathology, University of Minnesota Medical School. Room 1-156 Moos Tower, 515 Delaware Street SE. Minneapolis, MN 55455, USA; 3) Surgical Neurology Branch, National Institute of Neurological Disorders and Stroke (NINDS), National Institutes of Health (NIH). 10 Center Drive, CRC, Room 3D20, Bethesda, MD 20892-1414, USA; 4) Newborn Screening Program, Wadsworth Center, New York State Department of Health, P.O. Box 509, Albany, NY 12201-0509, USA; 5) Epidemiology Branch, Division of Intramural Population Health Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), National Institutes of Health (NIH). 6710B Building, Room 3117. Bethesda, MD 20892, USA.

Introduction: Although most patients present sporadically, Cushing's disease (CD) may occur in the setting of various syndromes of familial isolated pituitary adenoma (FIPA) or multiple endocrine and non-endocrine neoplasia. Such cases are found in the literature as individual case reports or small case series; therefore, the frequency of inherited forms of CD and their genetic causes remain largely unknown. Given that familial pituitary adenomas usually develop at a young age, pediatric patients represent an excellent population for estimating the prevalence of known genetic defects leading to CD and for gene discovery. **Methods:** We studied 268 pediatric CD patients (10.3±3.3 years at disease onset and 12.4±3.5 years at diagnosis) referred to our Center between 1997-2018, including 111 males (41.4%) and 157 females (58.6%). Seventy-one patients were screened for germline mutations in pituitary adenoma-associated genes (*AIP*, *CDKN1B*, *CDKN2C*, *MEN1* and *PRKAR1A*) by Sanger sequencing as part of a previous study. Ninety-eight patients (including 29 from the previous study) underwent whole-exome sequencing and, based on the results, one or more genes were analyzed in up to 153 additional patients by Sanger (42 from the previous study). Relevant data were obtained from the clinical records or directly from the patients. **Results:** Seven patients (2.6%) had familial disease: five cases presented as multiple endocrine neoplasia type 1 (four of them with confirmed *MEN1* mutations) and two cases presented as FIPA (unknown genetic cause). Among the 261 sporadic patients (97.4%), genetic analyses identified nine simplex cases (3.4%) with mutations in the *AIP* (n=1), *CABLES1* (n=2), *CDKN1B* (n=2), *MEN1* (n=1), *PRKAR1A* (n=1), *TSC2* (n=1), or *USP8* (n=1) genes. Altogether, cases with potentially inherited genetic causes (familial and simplex) accounted for 6% (16/268) of all patients. There were no statistically significant differences in age at disease onset, age at diagnosis or tumor diameter between patients with potentially inherited genetic defects and the rest of the cohort, although there was a trend toward larger tumors (median 6.5mm [interquartile range: 4.3-10.8] vs 4 mm [4-6.7] $P=0.0589$) in the former group. **Conclusions:** Known germline genetic defects cause a small but not negligible fraction of the CD cases presenting in childhood. Identifying such cases should lead to a more precise genetic testing and counselling, as well as to developing more targeted therapeutic strategies.

1011W

Preliminary results from whole exome sequencing in 100-sequential patients referred to Victorian KidGen Renal Genetics clinics reveals unexpected diagnoses. M. Wallis^{1,13}, J. Whittam^{2,13}, G. Valente^{1,15}, A. Leaver^{1,15}, A. Talbot^{3,13}, K. Nicholls^{3,13}, K. West^{4,15}, E. Allen^{4,15}, A. Trainer^{4,13}, L. Johnstone^{10,11}, J. Ryan^{5,12}, K. Jayasinghe^{5,12}, M. Hunter^{6,11}, Y. Praver^{6,15}, E. Krzesinski^{6,11}, S. White^{7,8,14}, Z. Stark^{7,8,14}, E. Wilkins^{7,15}, L. Wardrop⁸, C. Quinlan^{8,9,14}, *The KidGen Collaborative*. 1) Clinical Genetics Department, Austin Health, Heidelberg, Melbourne, Victoria, Australia; 2) Department of Nephrology, Austin Health, Heidelberg, Melbourne, Victoria, Australia; 3) Department of Nephrology, Royal Melbourne Hospital, Parkville, Melbourne, Victoria, Australia; 4) Adult Genetics Service, Royal Melbourne Hospital, Parkville, Melbourne, Victoria, Australia; 5) Department of Nephrology, Monash Medical Centre, Clayton, Melbourne, Victoria, Australia; 6) Monash Genetics, Monash Medical Centre, Clayton, Melbourne, Victoria, Australia; 7) Victorian Clinical Genetics Service, Parkville, Melbourne, Victoria, Australia; 8) Murdoch Children's Research Institute, Parkville, Melbourne, Victoria, Australia; 9) Department of Nephrology, Royal Children's Hospital, Parkville, Melbourne, Victoria, Australia; 10) Department of Nephrology, Monash Children's Hospital, Clayton, Melbourne, Victoria, Australia; 11) Department of Paediatrics, Monash University, Melbourne, Victoria, Australia; 12) Department of Medicine, Monash University, Melbourne, Victoria, Australia; 13) Department of Medicine, University of Melbourne, Victoria, Australia; 14) Department of Paediatrics, University of Melbourne, Melbourne, Australia; 15) Melbourne Genomics Health Alliance, Melbourne, Australia.

Background: Genetic kidney disease accounts for up to 70% of children and 20% of adults with end stage kidney disease. The emergence of genomic technologies has the potential to lead to better delineation of these conditions, the identification of new renal disease genes and subsequent targets for therapy. We describe the patient characteristics of 100-sequential referrals for individuals with genetic kidney disease across four Renal Genetics clinics in Victoria, Australia, and outline early results from exome sequencing in an initial consignment of 19 patients from this cohort. **Methods:** Patients were seen according to the KidGen clinic model and recruited through the Australian Genomics and Melbourne Genomics Renal Genetics Flagships. Demographic features, diagnosis at referral, and reason for referral were recorded. Preliminary results from whole exome sequencing (WES) were available in the first 19 patients. **Results:** The majority of patients (87) were referred by Nephrology, approximately half (51) were female, and 53 reported at least one affected first-degree relative. The reasons for referral included genetic testing (86), diagnosis (41), management (5), and genetic counselling (58). Common diagnoses at referral were Alports/hereditary nephritis (17), nephrotic syndrome (17), cystic kidney disease (7), ADTKD (7), ARPKD (6), Nephronophthisis (4) and ADPKD (2). Of the first 19 cases where WES results were available, a molecular diagnosis was suspected in 15 patients, leading to a revised diagnosis in seven. **Conclusions:** Ongoing analysis of this cohort will allow comparison of WES to current diagnostic methods, health economic analysis, and the identification of clinical characteristics or conditions that are more predictive of a positive molecular diagnosis. Early results suggest that a significant proportion may receive a revised diagnosis after WES, further highlighting the utility of identifying a specific molecular aetiology.

1012T

Chronic intestinal pseudo-obstruction: Successful genetic diagnosis through whole exome sequencing. *F. Djukiadmodjo¹, P. James^{1,2}, Z. Stark², M. Walsh^{1,2}.* 1) Genomic Medicine, The Royal Melbourne Hospital, Melbourne, Victoria, Australia; 2) Victorian Clinical Genetics Services, Murdoch Children's Research Institute, Melbourne, Victoria, Australia.

Introduction Chronic intestinal pseudo-obstruction (CIPO) is characterised by symptoms and signs of intestinal obstruction, but without evidence of a physical occlusion. It represents the severe end of a spectrum of gut motility disorders resulting in significant mortality (10-25%) and morbidity, with up to 80% of childhood-onset CIPO requiring total parenteral nutrition (TPN). Most cases are sporadic, with an estimated incidence of 1 in 40,000-100,000, but an increasing number of inherited causes are being described. **Aim** To describe two unrelated probands with causative *ACTG2* variants identified on whole exome sequencing (WES). **Methods** Family 1 presented with a four-generational history of CIPO. 9 of 28 family members were affected suggesting an autosomal dominant inheritance with high penetrance. The proband presented at 2 years of age and had steadily increasing severity of her condition, requiring multiple bowel resections and long-term TPN. Extensive genetic tests, including molecular karyotype and an extended panel for mitochondrial disorders, were negative. In contrast, the second proband is a neonate born at 36 weeks gestation who was found to have an atonic bladder at birth requiring ongoing catheterisation, following an antenatal diagnosis of an abdominal cyst. She did have initial bilious vomiting and required short-term TPN. **Results** WES performed in the proband of Family 1 identified a heterozygous novel missense variant (c.739C>A) at a highly conserved amino acid in the *ACTG2* gene (p.Gln247Lys). Segregation testing of other family members enabled us to confirm pathogenicity. The second patient was found to have a likely pathogenic heterozygous missense variant in *ACTG2* (c.613G>A, p.(Ala205Thr)) and she has been closely monitored for signs of intestinal obstruction. *ACTG2* encodes gamma (γ)-2 actin, a sarcomeric component of enteric smooth muscle involved in peristalsis. Inherited *ACTG2* gene mutations are associated with dominant visceral myopathy and de novo mutations have been found to be responsible for megacystis-microcolon-intestinal hypoperistalsis syndrome. **Discussion** CIPO is a rare but serious condition. Through WES, we identified *ACTG2* as the cause of CIPO in two families with contrasting presentations. This molecular diagnosis would enable early diagnosis and better management of CIPO that could avoid the need for bowel resection and the associated significant morbidity. It also enables reproductive testing options for affected family members.

1013F

Pancreatic cancer in a hereditary pancreatitis kindred is associated with a mutation in a pancreatic cancer susceptibility gene. *D.C. Whitcomb^{1,2}, C.A. Shelton^{1,2}, W. Zhan¹, P.J. Greer¹, D. Yadav¹.* 1) Division of Gastroenterology, Hepatology & Nutrition, University of Pittsburgh & UPMC, Pittsburgh, PA; 2) Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA.

Hereditary pancreatitis (HP) is an autosomal dominant syndrome of recurrent acute and chronic pancreatitis typically caused by mutations in *PRSS1*. HP is associated with an increased risk for pancreatic cancer (7.2%-40% by age 70). However, incidence of pancreatic cancer can vary dramatically between families and does not correlate well with age of onset or severity. HP probands and their family members were prospectively recruited into the HP study. Minimum three-generation pedigrees were constructed, subjects completed questionnaires, and medical records were obtained. DNA was tested for core pathogenic *PRSS1* mutations and patients were followed up to 20 years. Cause of death information was obtained from the family and the National Death Index (NDI). We identified clustering of HP-pancreatic cancer cases in HP kindreds, which suggests the presence of additional modifying factors. Whole exome sequencing (WES) was performed on available DNA from 6 individuals (of 4 *PRSS1* R122H kindreds) that developed pancreatic cancer (3 affected *PRSS1* carriers, 2 unaffected *PRSS1* carriers, and 1 non-*PRSS1*). Variants were filtered for known pancreatic cancer germline risk genes and evaluated for pathogenicity. Three study participants that developed pancreatic cancer were identified in one *PRSS1*-HP kindred, and all three were found to harbor a known cancer-susceptibility variant (rs664143) in the *ATM* gene – a critical component of DNA double-strand break repair. This kindred did not have known pathogenic variants in other familial pancreatic cancer-associated genes analyzed (*BRCA1*, *BRCA2*, *PALB2*, *SPINK1/2*, *GGT1*, *CTRC*, *CTRC*, *CFTR*, *TP53*, *SMAD3*, *CHECK2*, *CDKN2A*, *STK11*, *PTEN*, *BARD1*, *FANCA*, *FANCC*, *FANCG*, *FANCM*, *MLH1*, *MSH2/6*, *PMS2*, *EPCAM*, *POLN*, *POLQ*, *NBN*, *PALLD*, *APC*). The fact that *ATM* is a known oncogene adds biological plausibility to the argument that this variant contributed to development of pancreatic cancer in these cases. Ten ≥ 4 generation pedigrees were further investigated for incidence of cancer. Families exhibited high variability in pancreatic cancer age of diagnosis and incidence, ranging from (0% -28.6% of HP individuals $\geq$ age 50). These data demonstrate that in addition to smoking, risk for cancer in patients with HP may be linked to germline mutations in pancreatic cancer risk genes. Family history for pancreatic cancer should be considered when counseling HP patients for risk of pancreatic cancer.

1014W

Identifying causative genetic variants of pheochromocytoma and paraganglioma by next-generation sequencing (NGS) in Taiwan. Y.T. Chou¹, C.S. Chen², P.L. Chen^{1,2,3,4}. 1) Graduate Institute of Molecular Medicine, National Taiwan University, Taipei City, Taiwan; 2) Department of Medical Genetics, National Taiwan University Hospital, Taipei, Taiwan; 3) Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan; 4) Graduate Institute of Medical Genomics and Proteomics, National Taiwan University, Taipei, Taiwan.

Pheochromocytoma (PCC) and paraganglioma (PGL) are neuroendocrine tumors arising from adrenal and extra-adrenal chromaffin cells respectively. They mostly present benign, yet show high morbidity and mortality due to the overproduction of catecholamine, which leads to hypertension, arrhythmia and even ischemia stroke. About one thirds of PCC/PGL are caused by germline genetic variants; therefore, here we established a NGS panel to detect possible disease-causing variants. After literature review, we aimed at the top 12 PCC/PGL causative genes (*RET*, *VHL*, *NF1*, *FH*, *SDHA*, *SDHAF2*, *SDHB*, *SDHC*, *SDHD*, *MAX*, *TMEM127*, *EGLN1*(*PHD2*)), 10 somatic mutation genes (*ARNT*(*HIF1β*), *ATRX*, *BRAF*, *CSDE1*, *EPAS1*(*HIF1α*), *FGFR1*, *HRAS*, *IDH1*, *SETD2*, *TP53*), 3 fusion genes (*BRAF*, *MAML3*, *NGFR*) and other 7 genes (*CDKN2A*, *H3F3A*, *IDH2*, *KMT2D*, *MDH2C*, *MERTK*, *MET*) which were reported relative to PCC/PGL. Paired-end reads were generated from Illumina MiSeq platform and analyzed with in-house pipeline (BWA-MEM, Picard SortSam, MarkDuplicates, GATK-BQSR and ANNOVAR). Variants were filtering according to the allele frequencies in gnomAD, ExAC and Taiwan Biobank, and the pathogenicity interpretation were facilitated by disease/gene database, scientific literature and the 2015 ACMG Guidelines. In this study, we sequenced 41 probands and yielded approximately 25% diagnosis rate with identifying 6 different disease-causing variants and several variants of unknown significant (VUS). Among the results, 5 unrelated probands carried a same causative variant in *SDHD* (c.3G>C, p.M1X, NM_003002), which implied a founder mutation in Taiwan. One variant in *TMEM127* and one variant in *VHL* show co-segregation pattern in their pedigrees but remain absent in literature so far, which necessitated further functional assay.

1015T

Genotype and phenotype correlation in Turkish Alport syndrome patients with different COL4A3, COL4A4 and COL4A5 mutations. P. Ata, T. Celik, T. Pak, A. Isik, H.C. Yakut, S. Yildirim, A. Turkyilmaz. MARMARA UNIVERSITY , SCHOOL OF MEDICINE, ISTANBUL, TURKEY.

Introduction and Purpose: The clinical effects of *COL4A3*, *COL4A4* and *COL4A5* genes mutations differ from patient to patient. Herein we aimed to search the correlation between the genotype and differing phenotypes at Alport Syndrome. **Method:** This is a cross-sectional observational investigation covering patients admitted between December 2009 and September 2017. We established a scoring system; "ALSCORE" according to auditory, visual finding, biopsy, urinary, genetic findings and mutations of *COL4A3*, *COL4A4*, *COL4A5*. Also According to ACMG guidelines mutations were classified into groups (Group A, B, C,D) according to their pathogenicity Variants and mutations were analyzed with up-to-date genetic databases. For clinical and genotype correlation analysis SPSS v11.0 was used. Statistical significance score was determined as $p < 0.05$. Chi-square, Spearman, Kruskal-Wallis, Mann-Whitney-U tests were used. **Findings:** In a total of 37 patients 40.5% were males and 59.5% were females. Mean age was 13.49±6.83 years, mean age at diagnosis was 10.55±8.2 years. Mean observation duration was 47.01±28.36 months. Diseased relative history was present at 62.2% of patients and consanguinity was present at 40.5%. Eighty eight percent of all variants were (n=16) located at *COL4A4*. Mean "ALSCORE" was 5.78±2.49 (max: 10, min: 2). ALSCORE levels of mutation carriers were significantly higher than those of variant carriers ($p=0.002$, $p<0.01$). Those who were only variant carriers also had lower ALSCORE than those who were carriers of both variant and mutation at the same time. ($p=0.05$; $p<0.05$). For Statistical Analysis NCSS (Number Cruncher Statistical System) 2007 (NCSS, LLC, Kaysville, Utah, USA) programme was used. **Results:** Among all our patients *COL4A3* gene mutation was more frequent. The majority of patients with mutations had consanguineous marriage. Having two mutations at the same time had a higher impact on occurrence of visual findings at early ages. Patients with mutations had higher ALSCORE clinical scoring than those carrying mutation and variation at the same time. Most pathogenic mutations were accompanied with proteinuria. Macrohematuria was observed at younger ages.

1016F

Systematic characterization of variants that affect RNA splicing to optimize assignment of precision therapies for cystic fibrosis. A. Joynt¹, T. Evans¹, E. Davis¹, M. Pellicore¹, M. Atalar¹, S. Han¹, A. McCague¹, Z. Lu¹, N. West¹, C. Merlo¹, K. Raraigh¹, P. Sosnay¹, C.U. Cotton², N. Sharma¹, G.R. Cutting¹. 1) Johns Hopkins University, Baltimore, MD; 2) Case Western Reserve University, Cleveland, OH.

Treatment of cystic fibrosis (CF) has been revolutionized by modulators that target CFTR protein. To maximize the number of individuals who can benefit, we determined the molecular consequences of 46 naturally occurring CF-causing variants (intronic=37 and missense=9) that were predicted to affect RNA splicing by CryptSplice and NNSplice or by their proximity to consensus splice sites. Expression mini-genes (EMGs) encompassing 20 of the 26 *CFTR* introns were used to experimentally determine variant effect upon RNA splicing and protein synthesis. CFTR function was evaluated by short circuit current measurement of CF bronchial epithelial cells that stably expressed variant EMGs. Five variants (intronic=3 and missense=2) had no detrimental effect on RNA splicing. The remaining 41 variants exhibited aberrant splicing resulting in either frameshifts (intronic=25 and missense=4) or in-frame deletions (intronic=13 and missense=5). Twenty variants generated more than one splice isoform. Importantly, 9/20 variants (intronic=8 and missense=1) generated reduced amounts of normal spliced transcript. Remarkably, 3 variants (intronic=2 and missense=1) used non-canonical GC as a splice donor site. Our observations using EMGs were corroborated by splicing patterns observed in primary nasal cells from CF individuals (n=5) harboring the same splicing variants. Immunoblotting revealed that the 41 splicing variants fell into three groups: Group A generated no protein due to introduction of a premature termination codon (PTC) (intronic=7), Group B had truncated protein due to an in-frame deletion or PTC at the 3' region of *CFTR* (intronic=19, missense=5), and Group C synthesized reduced amounts of full-length protein from normally spliced RNA isoforms (intronic=8, missense=1). Strikingly, splice variants at the 5' region of *CFTR* (intronic=6) produced truncated protein, likely due to use of an alternative translational start site. Ten variants that generated protein were selected to evaluate CFTR function and the effect of modulators. Four variants from group B synthesized stable truncated forms of CFTR that did not respond to modulators. However, all six variants selected from group C displayed residual CFTR function and responded to both classes of CFTR modulator. Variants that alter RNA splicing patterns can generate targetable protein thereby increasing the number of individuals who could receive CFTR modulator therapy.

1017W

Natural genetic modifiers of *NGLY1* deficiency suggest a role for protein homeostasis in determining inter-individual outcomes. D. Talsness, K. Owings, E. Coelho, K. Peralta, C.Y. Chow. Department of Human Genetics, University of Utah, Salt Lake City, UT.

Autosomal recessive loss-of-function mutations in *N-Glycanase 1 (NGLY1)* cause *NGLY1* deficiency, the only known disease of deglycosylation. *NGLY1* deficiency is a devastating, rare, neglected disease. Patients with *NGLY1* deficiency present with developmental delay, movement disorder, seizures, liver dysfunction, and alacrima. *NGLY1* is a conserved component of the endoplasmic reticulum associated degradation (ERAD) pathway. ERAD degrades misfolded proteins that accumulate in the lumen of the ER. *NGLY1* deglycosylates misfolded proteins as they are translocated from the ER lumen to the cytoplasm for degradation. Patients with *NGLY1* deficiency show a wide spectrum of severity. This is especially striking since all identified patients carry two null mutations. This phenotypic heterogeneity poses a great challenge to treating this disorder. To understand what may be underlying inter-individual differences in severity of *NGLY1* deficiency, we undertook a natural genetic variation screen. We developed a *Drosophila* model of *NGLY1* deficiency and crossed it into 200 strains from the *Drosophila* Genetic Reference Panel (DGRP), a collection of wild-derived *Drosophila* strains. We assessed the effect of natural genetic variation in modulating the phenotypic impact of loss of *NGLY1* activity by quantifying the number of mutant flies that survived to adulthood. Across the 200 strains, we found a very large phenotypic spectrum in survival in the absence of *NGLY1*. We found that some strains showed nearly 100% survival without *NGLY1* function and other strains were completely lethal without *NGLY1*. We performed an association analysis to identify natural polymorphisms that modify survival in flies lacking *NGLY1*. Modifier genes fall into diverse categories involved in protein homeostasis, including ERAD (*SEL1L*), ER Ca²⁺ homeostasis (*TMTC1/2*), ER proteostasis (*ERMP1*), ubiquitination (*MYCBP2*), and vesicle trafficking (*RAB26*). We will present functional evidence, including tissue-specific studies, demonstrating how each modifier alters lethality associated with *NGLY1* deficiency in our *Drosophila* model. The genes identified from this study have excellent biological support and may serve as candidate modifiers of *NGLY1* deficiency. This study represents an important step to understanding the pathogenesis underlying *NGLY1* deficiency and provides a general framework for incorporating phenotypic variability, when developing personalized therapies for rare diseases.

1018T

Genetic insight into Birt-Hogg-Dubé syndrome: A rare disease in India. A. Ray¹, S. Kundu², S. Rajan³, S. Roy Chowdhury⁴, A. Bera⁵, U. Desai⁶, P. Prabhu-desai⁷, M. Kumar Sarma⁸, A. Kumar Prakash⁹, P. Singhal³, A. De², S. Mitra¹⁰, J. Madanmohan Joshi⁶, A. Jaiswal⁶, B. Roy¹. 1) Human Genetics Unit, Indian Statistical Institute, Kolkata, West Bengal, India; 2) Department of Pulmonary Medicine, Chest, RG Kar Medical College & Hospital, Kolkata, India; 3) Chest Medicine, Bombay Hospital Institute of Medical Sciences, Mumbai, India; 4) Pulmonology, Apollo Gleneagles Hospital Limited, Kolkata, India; 5) Critical Care Medicine, Medica Superspeciality Hospital, Kolkata, India; 6) Department of Pulmonary Medicine, TNMC and BYL Nair Hospital, Mumbai, India; 7) Chest Medicine, Lilavati Hospital and Research Centre, Mumbai, India; 8) Alipurduar District Hospital, Alipurduar, India; 9) Department of Respiratory and Sleep Medicine, Medanta -the Medicity, Gurugram, India; 10) Department of Chest Medicine, Calcutta National Medical College, Kolkata, India.

Background Birt-Hogg-Dubé (BHDS) is a rare genetic condition manifesting in either or all of the three phenotypes - primary spontaneous pneumothorax (PSP), cutaneous fibrofolliculomas and kidney neoplasms. Familial PSP is mostly associated with inherited mutations in *Folliculin* (*FLCN*). Other similar pulmonary phenotypic disorders –LAMS, Homocystinuria, Marfan syndrome, alpha-1 antitrypsin deficiency, Ehlers-Danlos syndrome can sometime make a confounding diagnosis of BHDS making it important to distinguish BHDS from other PSP related phenotypes in the context of inherited mutations.

Most of the BHDS studies have been reported from Western countries with considerably fewer reports from Asia, including one study from India. Here, we are reporting mutations in *FLCN* and other PSP related genes in 86 individuals from 13 Indian families, having at least one BHDS patient in each family.

Methods Bidirectional Sanger DNA sequencing was performed to detect inherited mutation at *FLCN* from genomic DNA. Germline variations at *MTHFR* and *SERPINA1*, generally observed in Homocystinuria and alpha-1 antitrypsin deficiency respectively, were also evaluated. Sequence alignment software and protein prediction tools were used to detect the effect of mutations on protein structure. **Results** Sequencing data revealed a heterozygous C-deletion in poly-cytosine (C_n) tract of exon 11(c.+1285) at *FLCN* in 19 individuals from 5 families, leading to a truncated protein. A family with a 41-yr-old female index patient (symptomatic), her mother and sister (both asymptomatic) were found to have a heterozygous 11-base deletion (c.+1150_1160del11) in exon 10 of *FLCN*. This mutation has not, yet, been reported in any of BHDS patients from all over the world. Protein modelling predicted a stop codon (c.+1164) resulting in a truncated *FLCN* protein. No other pathogenic inherited mutations in *FLCN* were found in remaining 64 individuals from 7 families. Two deleterious SNPs - *rs1801133* and *rs762920973* were also found in *MTHFR* and *SERPINA1* respectively. Therefore, search for other structural mutations in *FLCN* and other PSP related genes are being carried out in these samples. **Conclusion** Apart from reported exon 11 *FLCN* mutation (c.+1285), a novel 11-nucleotide deletion mutation(c.+1150_1160del11) in *FLCN* was also detected for the first time. Both the mutations are predicted to produce truncated *FLCN* protein. The truncated *FLCN* may play a role in deregulating metabolic pathways to generate the BHDS phenotype.

1019F

Autosomal dominant visceral myopathy in connection with a *de novo* MYH11 truncating mutation. A.C. Lionel¹, J. Silver², Y. Avitzur³, J. So^{1,2,4}. 1) Department of Medicine, University of Toronto, Toronto, Ontario, Canada; 2) The Fred A. Litwin Family Centre in Genetic Medicine, University Health Network and Mount Sinai Hospital, Toronto, Ontario, Canada; 3) Division of Gastroenterology, Hepatology and Nutrition, The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada; 4) Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada. The disruption of *MYH11*, the gene encoding smooth muscle myosin heavy chain, results in a range of diverse clinical phenotypes. Heterozygous mutations, with suggestive evidence for a dominant negative mechanism of action, have been linked to familial thoracic aortic aneurysms and dissections, as well as patent ductus arteriosus. In contrast, autosomal recessive loss-of-function *MYH11* mutations have been reported in visceral myopathy, a rare inherited form of functional obstruction in the gastrointestinal and urinary tracts, also known as megacystis-microcolon-intestinal hypoperistalsis syndrome (MMIHS) or Berdon syndrome. We present a family with the first reported instance of autosomal dominant visceral myopathy connected to *MYH11*. The proband is a 32-year-old male with a history of congenital visceral myopathy and intermittent atonic bladder during childhood, with no history of cardiac issues and normal echocardiogram. He was dependent on total parenteral nutrition until successful cluster transplantation of stomach, small bowel, liver, and pancreas in young adulthood. Genetic testing revealed a *de novo* heterozygous frameshift deletion (c.5770_5779del; p.Leu1924Serfs*84) in *MYH11*, resulting in premature truncation of the coiled-coil domain. The pathogenic variant was subsequently also found in the proband's infant son with symptoms of visceral myopathy. Our findings expand the phenotypic spectrum of *MYH11* mutations and increase understanding of the genetic architecture of visceral myopathy. This will aid in genetic diagnosis and counseling related to this disorder.

1020W

New digenic combinations with mutations of the genes associated with GnRH-dependent pubertal disorders contribute to variable phenotypes in idiopathic hypogonadotropic hypogonadism. A. Gach¹, I. Pinkier¹, U. Wysocka¹, K. Salacinska¹, D. Salachna¹, M. Szarras-Czapnik², L. Rutkowska¹, M. Rybak-Krzyszowska³, L. Jakubowski¹. 1) Department of Genetics, Polish Mother's Memorial Hospital Research Institute, Lodz, Poland; 2) Department of Endocrinology and Diabetology, Children's Memorial Health Institute, Warsaw, Poland; 3) Department of Obstetrics and Perinatology, Jagiellonian University Medical College, Cracow, Poland.

Introduction Idiopathic hypogonadotropic hypogonadism (IHH) results from a dysfunction of the hypothalamic-pituitary-gonadal axis (HPG), which is essential for development and function of the reproductive system. IHH may be associated with anosmia/hyposmia (Kallmann syndrome, KS [MIM 308700]) or normal sense of smell. For many years IHH has been considered a monogenic disorder, with a single gene mutation conditioning up to 30% of cases. Numerous studies have proven that hypogonadotropic hypogonadism is not simply a monogenic Mendelian disease, but more than one gene may be involved in its pathogenesis in a single patient. **Objective** The aim was to explore digenic mutations spectrum in IHH/KS patients. **Material and methods** The study was carried out in the group of 32 patients with hypogonadotropic hypogonadism (HH) with or without anosmia using next generation sequencing (NGS) panel of 51 genes associated with HH. was performed on the MiniSeq platform with TrueSeq Custom Amplicon protocol for library preparation. Variant Studio Software was used for variant filtering and reporting. All disease-relevant and biologically significant variants were validated by Sanger sequencing. **Results** Of 32 IHH subjects five unrelated patients have been found to carry digenic mutations. Subject No 1 is a woman who presented to genetic counseling with primary amenorrhea, discrete features of dysmorphism and normosmia. She was identified with mutations in *CCDC141* (MIM 616031) and *POLR3B* (MIM 614366) genes. Subject No 2 is a woman with KS and anosmia, who was found to carry mutations in *SRA1* (MIM 603819) and *SEMA7A* (MIM 607961) genes. Subject No 3 is a man with micropenis, cryptorchidism and anosmia. Mutations in *CHD7* (MIM 608892) and *PROKR2* (MIM 607123) genes were reported. Subject No 4 and no 5 are men referred to genetic specialists due to clinical diagnosis of IHH. They were found to carry *SPRY4* (MIM 4607984)/*SEMA3A* (MIM 603961) and *STON1* (MIM 605357)/*SEMA3A* mutations respectively. Digenic combinations identified in this study have not been reported elsewhere. **Conclusion** The identification of 5 patients with IHH each harboring a unique combination of two rare variants in genes involved in GnRH neural migration, GnRH secretion or affecting pituitary gland function demonstrates the complexity of the genetic architecture of the disorder. In the era of personalized medicine understanding of full phenotypical spectrum of genotypes associated with IHH is increasingly important.

1021T

Novel gene defect involved in the surfactant metabolism associated with canine lethal neonatal pulmonary alveolar proteinosis. K.J. Dillard^{1,2,3,4}, M. Ochs⁵, J. Niskanen^{1,2,3}, M. Anttila⁴, M.K. Hytönen^{1,2,3}, H. Lohi^{1,2,3}. 1) Department of Veterinary Biosciences, University of Helsinki, Helsinki, Finland; 2) Molecular Neurology, Research Programs Unit, University of Helsinki, Helsinki, Finland; 3) The Folkhälsan Research Center, Helsinki, Finland; 4) Finnish Food Safety Authority - Evira, Pathology, Helsinki, Finland; 5) Institute of Functional and Applied Anatomy, Hannover Medical School, German Center for Lung Research (DZL), Hannover, Germany.

Pulmonary surfactant is produced by alveolar epithelial type II cells (PCII) and is a mixture of lipids and proteins with a property to reduce surface tension in the alveolus necessary for normal gas exchange in the lung. Proper lung surfactant metabolism is crucial for life and genetic disorders affecting this homeostasis are rare in human or any other species. Dog has emerged as an important natural model for human inherited diseases with clinicopathological and genetic similarities. We have studied here an extended family of dogs with neonatal mortality due to dyspnea from birth that generally leads to death in a few hours to days. Pathological examination of the affected puppies revealed a severe congenital pulmonary alveolar proteinosis resembling the group of human disorders of surfactant metabolism. Pedigree established around the affected dogs suggested an autosomal recessive inheritance. To identify the genetic cause, a combined approach of a genome wide association study (5 cases and 24 controls) and exome sequencing (2 cases and 2 controls) was utilized. GWAS using canine high density arrays revealed a disease associated ~3.9 Mb homozygous haplotype in chromosome 34 ($p_{raw} = 7.161 \times 10^{-10}$, $p_{Bonferroni} = 6.689 \times 10^{-6}$). Variant filtering in the identified locus with the assumption of a recessive inheritance revealed only three coding changes. Further analysis of these three variants in additional cases and controls in the breed (n=399), indicated a full segregation and penetrance of one of the variants with the disease. This variant is located in a functional candidate gene abundantly expressed in the lung and involved in surfactant metabolism. Our ongoing experiments aim to confirm the causality of this variant and the novel association of the gene with the disorders of surfactant metabolism.

1022F

Characterization of candidate genes for short stature by establishing evolutionary conserved networks of human growth. C.T. Thiel¹, N.N. Hauer¹, L. Taher², B. Popp¹, C. Vogl¹, R. Ahmadian³, P.S. Dhandapany⁴, C. Büttner¹, S. Uebe¹, H. Sticht⁵, F. Ferrazzi¹, A.B. Ekici¹, A. de Luca⁶, E. Schoeller¹, S. Schuhmann¹, K.E. Heath¹, A. Hisado-Oliva⁷, P. Klinger⁸, S. Boppud⁹, J. Kelke¹⁰, A.M. Juung¹⁰, C. Zweier¹, E. Kunstmann¹¹, K. Kutsche¹², A. Rauch¹³, D. Wiczorek^{14,15}, T. Rohrer¹⁰, M. Zenker⁸, H.-G. Doerr¹⁶, A. Reis¹. 1) Institute of Human Genetics, Friedrich-Alexander-University of Erlangen-Nuremberg, Erlangen, Bavaria, Germany; 2) Bioinformatics, Department of Biology, Friedrich-Alexander-University of Erlangen-Nuremberg, Erlangen, Germany; 3) Institute of Biochemistry and Molecular Biology II, Medical Faculty, Heinrich-Heine University, Düsseldorf, Germany; 4) The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 5) Institute of Biochemistry, FAU Erlangen-Nürnberg, Erlangen, Germany; 6) Mendel Laboratory, Casa Solievo della Sofferenza Hospital, IRCCS, San Giovanni Rotondo, Rome, Italy; 7) Institute of Medical and Molecular Genetics (INGEMM) and Skeletal dysplasia Multidisciplinary Unit (UMDE), Hospital Universitario La Paz, Universidad Autónoma de Madrid, IdiPAZ, and CIBERER, Madrid, Spain; 8) Department of Orthopaedic Rheumatology, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany; 9) Institute of Human Genetics Otto von Guericke University Magdeburg, Magdeburg, Germany; 10) Division of Pediatric Endocrinology, Department of Pediatrics and Neonatology, Saarland University Hospital, Homburg/Saar, Germany; 11) Institute of Human Genetics, University of Würzburg, Würzburg, Germany; 12) Institute of Human Genetics, University Medical Center Hamburg-Eppendorf, Hamburg, Germany; 13) Institute of Medical Genetics, University of Zurich, Zurich, Switzerland; 14) Institute of Human Genetics, University of Essen, Essen, Germany; 15) Institute of Human Genetics, University Dueseldorf, Dueseldorf, Germany; 16) Department of Pediatrics and Adolescent Medicine, FAU Erlangen-Nuremberg, Erlangen, Germany.

The Height of an individual is a heritable and highly heterogeneous trait. So far, rare, low-frequency and common height-associated variants together explain about 27.4% of the heritability of height. Consistent with a model complex human trait, rare, large effect-size variants are important for growth development in individuals and are the underlying cause in monogenic forms of short stature. The missing causes in many patients represents an unmet problem for the evaluation of effective and disease-specific treatment. We established a clustered network of human growth and short stature genes to facilitate the characterization of novel candidate genes for short stature identified from WES in 254 unrelated families without known cause. Genes potentially involved in growth delay or growth regulation were selected from MedGen and OMIM. We performed enrichment analysis and clustering of terms from the Gene Ontology, UniProt, OMIM and KEGG in order to retrieve the biological framework of growth. The most significant cluster were related to Skeletal system development, Appendage development, Metabolic processes and Ciliopathy. Moreover, clusters were often related to central processes like chromatin binding or extracellular matrix formation. Based on the systematic characterization of WES variants in the 254 patients we classified 13 candidate genes as 1st level (highly confident) and 50 as 2nd level (confident) candidate genes. 6 of these 63 1st and 2nd level candidate genes (9.5 %) have previously been reported as candidate genes for short stature or syndromes including short stature (IFT81, ZBED4, AMMECR1, BRD4, FZD2, and LZTR1). Considering the enrichment cluster analysis 11 (85%) of the identified 1st level candidate genes (CPZ, EDEM3, FBRS, IFT81, KCND1, PLXNA3, RASA3, SLC7A8, UBR4, USP45, and ZFH3) are part of 21 defined clusters significantly enriched for growth (49 (98%) of the 2nd level candidate genes in 25 clusters). 2 (15%) of the identified 1st level candidate genes (LAMA5, MED24) and one of the 2nd level candidate genes (GREB1L) were not included in enriched clusters for growth indicating novel growth pathways. In conclusion, we build a map of enriched evolutionary functions and pathways for known growth retardation genes. We successfully applied this cluster to characterize 13 highly confident novel (50 confident) candidate genes for short stature.

1023W

Prevalence and treatment of OI-related hearing loss with bisphosphonates. E.M. Carter, C. Citron, S. Khan, K. Citron, C.L. Raggio. , Select a Country.

PURPOSE: Osteogenesis Imperfecta (OI) is a bone disorder that usually stems from dominant mutation in the type 1 collagen genes. The prevalence of hearing loss in OI varies from 50-92%. Bisphosphonates limit bone turnover and are used to treat OI. Recent research suggests bisphosphonates can halt the progression of OI-related hearing loss. **METHODS:** Retrospective review of 150 charts from individuals with a clinical diagnosis of OI identified 24 individuals (ages 2-66) with audiograms. Pure tone averages (PTA) of those with a single audiogram (N=17) were reviewed in the context of their age, OI type, and bisphosphonate treatment. PTA of individuals with more than one audiogram (N=7) were analyzed longitudinally with bisphosphonate treatment (hearing loss defined as ± 10 dB since previous audiogram). **RESULTS:** The majority of the population was female (N=21) with an average age of 30.08 ± 19.28 years. 24/150 individuals had an audiogram. 8/24 did not experience hearing loss, 2 of those took bisphosphonates. 16/24 individuals who had audiograms displayed hearing loss; 9 of those were treated with bisphosphonates. 8 individuals had OI Type I and hearing loss; 2 were treated with bisphosphonates, 6 never took bisphosphonates. 11/150 individuals were referred for ENT consultation but not given an audiogram. Left ear pure tone measurements at the 500 Hz level were the only significant difference between individuals treated with bisphosphonates and those who were not. Individuals with Types I, III, and IV OI significantly varied only in left ear pure tone measurements at 500 and 1000 Hz. There was no relationship between PTAs and age or years of bisphosphonate treatment. **CONCLUSION:** The comparisons made in this study suggest that bisphosphonate treatment in OI is equivocal in halting the progression of hearing loss. Because most individuals in this study did not have multiple audiograms, it is difficult to postulate whether bisphosphonate use impacts deterioration of hearing loss. These results also demonstrate the need for communication between specialists to optimize care for individuals with OI and raises questions about the lack of patient compliance in following physician recommendations. **SIGNIFICANCE:** Hearing loss may potentially be averted and/or the progression halted with bisphosphonate use. Emphasis needs to be placed on getting regular hearing tests so that proper treatment can be administered and its effects observed.

1024T

Isolated brachymetatarsia due to a mutation in *TBX4*. J. D'Alessandro, G. LaPorta. Podiatric Surgery, Lourdes Memorial Hospital, Binghamton, NY.

Introduction Brachymetatarsia is a disorder that causes premature closure of the physis associated with one or more metatarsal bones. The deformity can occur in isolation or with other comorbidities and has a predilection to the fourth metatarsal. Isolated Brachymetatarsia is rare and displays variable Mendelian patterns of inheritance. We performed whole exome sequencing (WES) in an affected Caucasian male presenting with isolated, idiopathic Brachymetatarsia and identified a pathogenic *de novo* variant in *TBX4*.

Materials and Methods -IRB approval and informed consent were obtained from Lourdes Memorial Hospital. -Genomic DNA from Proband and unaffected parents extracted from saliva. -Target capture on Agilent SureSelect V6.0. -Enriched library sequenced at Quick Biology. -Multiple filtering steps: Assuming a dominant inheritance pattern variants that were filtered out included: non-exonic, synonymous changes, or $MAF > 0.01\%$ in 1000 genomes, ESP (n=6500), ExAC (n~62000), or gnomAD (n~123,000). -Prioritization steps based on: Evolutionary conservation (PhyloP), prediction of pathogenicity (PolyPhen2/SIFT), Biological/Clinical relevance (Pathways / interaction networks). -Sanger sequencing was used for validation in each member of the trio with primer pairs designed to the relevant region of *TBX4* (Reference sequence NM_001321120.2). -Further exploration of this mutation and its clinical implications pending. **Main Findings** After WES in 3 samples and variant calling / filtering: -5 rare, nonsynonymous variants were retained, all with deleterious prediction by SIFT/PolyPhen. -the *TBX4* nonsense mutation chr17:59534035 NM_018488.3:c.184C>T p.(Gln62Ter) is the most relevant as *TBX4* is linked to Ischiopatellar dysplasia (OMIM: 601719.0002), implying a role in lower extremity development. -This *TBX4* mutation is not documented in any public database and is thus considered to be very rare; its nonsense nature could greatly compromise the proteins structure-function relationship. **Conclusion** -This is the first documented report of Brachymetatarsia caused by a *de novo* mutation of *TBX4*. -The discovery of the 184C>T p.(Gln62Ter) mutation in *TBX4* therefore has important implications for disease screening, prognosis, and genetic counseling. .

1025F

Childhood hypophosphatasia: Painful bone marrow edema mimicking chronic recurrent multifocal osteomyelitis improved after three months of asfotase alfa enzyme replacement therapy. G.S. Gottesman¹, D. Wenkert¹, W.H. McAlister², G. Khanna³, K. Mack¹, S. Mumm^{1,3}, M.P. Whyte^{1,3}. 1)

Center for Metabolic Bone Disease, Shriners Hospital for Children - St. Louis, Saint Louis, MO; 2) Mallinckrodt Institute of Radiology, Washington University School of Medicine, St. Louis, MO; 3) Division of Bone and Mineral Diseases, Washington University School of Medicine, St. Louis, MO.

Hypophosphatasia (HPP) is the inborn-error-of-metabolism that features hypophosphatasemia, tooth loss, and impaired skeletal mineralization with varying expressivity. In 2009, we reported an unrelated boy and girl with childhood HPP and recurrent periarticular soft tissue swelling and pain, primarily at the distal forearms and ankles. Radiographic, histopathologic, and MRI findings, including bone marrow edema, mimicked chronic recurrent multifocal osteomyelitis (CRMO). The pathogenesis remained unknown. Here, we report a teenage boy with childhood HPP and 7 episodes of recurrent wrist swelling and pain (5 episodes within the preceding 3 years). At age 9 years, he reported brief morning stiffness and intermittent leg pain treated with non-steroidal medications (NSAIDs). He had no fever, mucocutaneous or gastrointestinal symptoms. Each episode occurred with minor trauma causing rapid localized swelling and was associated with point tenderness. Conservative treatment with casting, for presumed Salter-Harris I growth plate fractures, improved the pain and swelling. He took NSAIDs for pain typically twice weekly. At 13 years of age, our review of the interval radiographs revealed slight worsening of the abnormalities typical of HPP at his wrists and knees. We then explored the most recent episode with an MRI of his forearms and hands, which showed metadiaphyseal signal increases consistent with CRMO. Additionally, functional testing confirmed new gross motor deficits consistent with childhood HPP. Mutational analysis of *ALPL* identified a novel heterozygous ~65 base pair deletion 5' of exon 4 that removed the splice site. With an evolving, worsening clinical status during his pubertal growth spurt, we started asfotase alfa enzyme replacement therapy (ERT). Three months later, pain had resolved (off NSAIDs), and follow-up MRI showed significant improvement of the metadiaphyseal bone edema (Figure). C-reactive protein and erythrocyte sedimentation rate, normal at baseline, remained normal. Procollagen type I amino-terminal peptide, elevated at baseline, fell 3.6-fold but was still elevated; collagen carboxy-terminal telopeptide also elevated at baseline, had normalized. Autoinflammatory mechanisms are considered an etiology for CRMO. Our patient's response to ERT alone suggests that CRMO-like lesions in HPP reflect a metabolic consequence of the alkaline phosphatase deficiency.

1026W

Temporal growth trends in achondroplasia: Results from a multicenter achondroplasia registry. S. Hashmi¹, D. Rodriguez-Buritica¹, E. Serna¹, A. Alade², J. Hoover-Fong², M. Bober³, M. Little³, R. Pauli⁴, J. Legare⁴, P. Modaff⁴, C. Smid⁴, J. Hecht⁴. 1) McGovern Medical School at UTHealth, Houston, TX; 2) Greenberg Center for Skeletal Dysplasias, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD; 3) Al DuPont Hospital for Children, Jefferson University, Wilmington, DE; 4) University of Wisconsin School of Medicine and Public Health, Madison, WI.

Objective: Achondroplasia is the most common form of dwarfism occurring in 1/25,000 newborns. The natural history has been defined in a number of small studies that included limited number of individuals with achondroplasia. Therefore, the information about medical issues is sparse or imprecise leaving uncertainties about specific health recommendations. Weight in achondroplasia is a vexing problem for those with achondroplasia and for the providers who care for them, height velocity in achondroplasia is poorly defined and head circumference changes need better definition. The objective of this study is to utilize the recently created achondroplasia natural history database to evaluate temporal trends in height, weight and head circumference and to produce new growth charts. **Methods:** Approval was obtained by all institutional IRBs. A large multicenter achondroplasia registry using Research Electronic Data Capture (REDCap) system was created that included longitudinal patient data from Johns Hopkins University, Al duPont Hospital for Children, McGovern Medical School UTHealth, and the University of Wisconsin. All available demographic and medical data from 1957-2017 was abstracted and entered by study coordinators at each site and checked for consistency. Anthropomorphic data was compiled to calculate growth curves and growth velocities. **Results:** 1,377 patients in the registry yielded 15,734 data points for weight, 13,427 for height and 9951 for head circumference. Linear growth continued through 19.3 years and 21.7 years, with final adult heights of 127.3 cm (4'2") and 120.2 cm (3'11"), for males and females, respectively. There was no evidence of a pubertal growth acceleration for either males or females. The average adult weight in the study population was 59.4 kg (SD 16.6) and 58.4 kg (SD 17.4) and the average adult head circumference was 59.4 cm (SD 2.5) and 58.1 cm (SD 3.4), for males and females, respectively. **Conclusions:** The new growth charts are based on the largest number of measurements from patients with achondroplasia across age groups in the US. They allow greater precision in clinical follow-up of patients. They may also serve as baseline controls for comparison to subjects treated with novel pharmacologic, medical and surgical interventions, and other unstudied components.

1027T

Application of noncontinuously binding (loop-out) oligonucleotide primers to avoid a polyT region in the untranslated region of *SHOX* gene exon 1. M.A. Jama¹, A. Milson¹, K. Sumner¹, G. Pont-Kingdon¹, Y. Ji^{1,2,3}. 1) 1ARUP Institute for Clinical and Experimental Pathology, Salt Lake City, Utah, Salt Lake City, UT. 500 Chipeta Way, Salt Lake City, UT 84108; 2) Department of Pathology, University of Utah, Salt Lake City, UT; 3) ARUP Laboratories, Salt Lake City.

Defects in the *SHOX* gene (short stature homeobox-containing gene), located in the pseudoautosomal region 1 (PAR1) of X and Y chromosomes, are associated with Léri-Weill, Langer syndrome, and idiopathic short stature. *SHOX* deficiency can be determined by MPLA analysis that detects ~80% pathogenic variants in patients, and DNA sequencing for sequence alteration in the remaining ~10-20% of patients. PCR primers were designed and M13 tails added to the 5'-end of the PCR primers to allow for touchdown PCR and Sanger sequencing using universal M13 sequencing primers. The noncoding region of exon 1, exons 2-5, and 6a were amplified. A highly polymorphic T (polyT) and a (GC)₅₋₆ regions were noted in the 5' untranslated region (UTR) of exon 1. A noncontinuous sequencing (loop-out) forward and reverse primers were therefore designed to obtain clean sequences for the detection of variants upstream and downstream of the polyT region. A forward sequencing primer was also designed upstream of the (GC)₅₋₆ region to achieve clean sequences only if the GC₅₋₆ region was polymorphic and heterozygous. Twelve DNA samples (10 wild-type and 2 Coriell samples with known *SHOX* mutations, NA20215 and NA20216) were amplified and sequenced with M13 primers and noncontinuous forward and reverse primers. DNA sequences with poor quality score around the polyT region were observed in all 12 samples with regular M13 sequencing primers, whereas clean DNA sequences were achieved from the same 12 PCR amplicons when the two loop-out primers were used. Only Sample #5 was noted to have the GC polymorphism and thus the upstream (GC)₅₋₆ forward primers were used to improve sequencing quality. Pathogenic variant and benign polymorphisms were previously reported in the 5'UTR of *SHOX* where both polyT and (GC)₅₋₆ tracks are technically challenging for optimal sequencing quality. Our study has demonstrated an easy solution for clinical laboratory by using a pair of noncontinuous loop-out primers. The (GC)₅₋₆ forward primers can be used only when a heterozygous (GC)₅₋₆ polymorphism is noted. This improved Sanger sequencing strategy can now be applied in the clinic lab to ensure the detection of sequence variants in up to 20% of patients who carry *SHOX* deficiency upon a negative *SHOX* MLPA result.

1028F

Genotypic and phenotypic characterization of Chinese patients with osteogenesis imperfecta. L. Li¹, B. Mao¹, S. Li¹, J. Xiao¹, H. Wang¹, J. Zhang¹, X. Ren², Y. Wang³, Y. Wu¹, C. Lu¹, Y. You¹, F. Zhao¹, X. Geng¹, C. Jiang¹, Y. Ye¹, T. Yang¹, X. Zhao¹, X. Zhang¹. 1) Medical Genetics Dept, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences, Beijing, China; 2) Pediatric Orthopaedics Dept, the People's Hospital of Wuqing District, Tianjin; 3) Pediatric Orthopaedics Dept, Shandong Provincial Hospital Affiliated to Shandong University, Jinan.

Osteogenesis imperfecta (OI) is a clinically and genetically heterogeneous group of rare inherited connective tissue disorders, characterized by multiple fractures, deformity of the long bones, short stature and a variety of extraskelletal features. It is estimated that heterozygous mutations in genes encoding either the $\alpha 1$ (*COL1A1*) or $\alpha 2$ (*COL1A2*) chains of type I collagen cause 85 – 90% of OI cases with autosomal dominant (AD) inheritance. With the improvement of molecular diagnoses, at least 15 genes involved in bone mineralization (*IFITM5* and *SERPINF1*), collagen post-translational modification (*CRTAP*, *P3H1* and *PP1B*), processing and crosslinking (*SERPINH1*, *FKBP10*, *PLOD2* and *BMP1*), and osteoblast differentiation and function (*SP7*, *TMEM38B*, *WNT1*, *CREB3L1*, *SPARC* and *MBTPS2*) have been identified in OI patients. In the present study, we recruited the largest OI cohort in Asian populations, and performed clinical investigation and genetic analyses (whole exome sequencing and targeted sequencing) in 668 Chinese OI patients from 378 families. Finally, we identified 274 mutations (230 in type I collagen encoding genes and 44 in non-collagen genes), including 98 novel mutations, in 340 probands with a detection rate of 90%. Compared with 47 loss-of-function mutations detected in *COL1A1*, either a nonsense or a frameshift mutation was not found in *COL1A2* ($P < 0.0001$). The major cause of autosomal recessive OI was biallelic mutations in *WNT1* (56%, 20/36). It is noteworthy that three genomic rearrangements, including one gross deletion and one gross duplication in *COL1A1* as well as one gross deletion in *FKBP10*, were detected in this study. Of ten individuals with N-terminal glycine substitutions in the $\alpha 1$ (I) chain, none exhibited hearing loss, suggesting a potential genotype–phenotype correlation. The findings in this study expanded the mutation spectrum and identified potential novel correlations between genotype and phenotype in Chinese OI patients.

1029W

Vertebral segmentation defect and hemivertebra due to rare autosomal biallelic *FBN2* mutations in a Chinese congenital scoliosis cohort. M. Lin^{1,2}, Y. Chen^{1,2}, W. Chen^{1,2}, Y. Niu³, X. Li³, Z. Wu^{3,4,5}, N. Wu^{1,3,4,5}, J. Lin^{1,2,4}, Y. Ye^{1,4,5}. 1) Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, B, Beijing, Beijing, China; 2) Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100730, China; 3) Beijing Key Laboratory for Genetic Research of Skeletal Deformity, Beijing 100730, China; 4) Department of Central Laboratory, Peking Union Medical College Hospital, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing 100730, China; 5) Medical Research Center of Orthopedics, Chinese Academy of Medical Sciences, Beijing 100730, China.

Background Congenital scoliosis (CS) is a common disorder with an incidence of around 0.5–1 in 1000 live births. The genetic basis of sporadic CS is still unclear. *FBN2* gene which lies in 5q23.3 region has been suggested as a candidate gene for CS. Diseases associated with *FBN2* include Arthrogryposis, Distal, Type 9 (DA9; MIM: # 121050) and Macular Degeneration, Early-Onset (EOMD; MIM: # 616118). *Fbn2* mouse models indicated skeleton phenotypes involving congenital kyphoscoliosis and vertebral segmentation defect in MGI. Previous report indicated that partial fusion of C2–C3 vertebral bodies could be characteristics of DA9. Moreover, congenital kyphoscoliosis was shown in about 45% in DA9 according to *FBN2* implicated in OMIM. **Methods** We consecutively enrolled 500 individuals with sporadic CS via Deciphering Disorders Involving Scoliosis and Comorbidities (DISCO) study (<http://www.discostudy.org/>) in Peking Union Medical College Hospital (PUMCH) from October 2010 to November 2017, 110 of which have familial samples. WES was performed on the patients and available family members. Variant calling and annotation was executed by an in-house developed analytic pipeline (PUMch Pipeline, PUMP), followed by data interpretation and reporting according to ACMG guidelines. **Results** We identified rare autosomal biallelic *FBN2* missense mutations (c.1180C>T(p.Arg394Cys) and c.6073G>A(p.Gly2025Ser)) in one sporadic CS case of XH483 by WES. XH483 is characterized by vertebral segmentation defect of T10–T11 and T11 hemivertebra. According to ACMG guidelines, this variant of Arg394Cys should be classified as VUS because the only support of potential pathogenicity comes from its complete absence in our in-house controls, MAF in ExAC (0.000016) and in silico prediction (PolyPhen (0.999), SIFT (0.16), and CADD (26.3)). Arg394 is in the TGFBP #01 domain and highly conserved. Moreover, the variant of Gly2025Ser was recurrent in 3 unrelated CS patients and two unaffected controls. We consider this variant was a risk allele due to the low MAF in ExAC (0.000074) and in silico prediction (PolyPhen (1), SIFT (0.19), and CADD (35)). Gly2025 is in the cb EGF-like #31 domain and highly conserved. Thus, the biallelic *FBN2* mutations were putatively deleterious. **Conclusions** Our findings indicated that biallelic *FBN2* missense mutations might contribute to CS, which also raises the possibility that some CS cases may be due to a primary pathogenetic mechanism involving the recessive *FBN2* variants.

1030T

Bruck syndrome variant lacking congenital contractures due to novel compound heterozygous *PLOD2* mutations. S. Mumm^{1,2}, G.S. Gottesman², P.M. Campeau³, A. Nenninger², M. Huskey⁴, V.N. Bijanki², D.J. Veis^{1,2}, A. Barnes⁴, J.C. Marini⁴, D. Wenkert², W.H. McAlister², M.P. Whyte^{1,2}. 1) Division of Bone and Mineral Diseases, Department of Internal Medicine, Washington University School of Medicine, St. Louis, MO; 2) Center for Metabolic Bone Disease and Molecular Research, Shriners Hospital for Children; 3) Department of Pediatrics, University of Montreal; 4) National Institute of Child Health and Human Development, NIH; 5) Mallinckrodt Institute of Radiology, Washington University School of Medicine at St. Louis Children's Hospital.

A 10-year-old girl was referred for an unidentified skeletal dysplasia featuring osteopenia with poorly-healing fragility fractures, short stature, Wormian bones in her skull, cleft soft palate and uvula, congenital fusion of cervical vertebrae, bell-shaped thorax, progressive scoliosis, restrictive and reactive pulmonary disease, blue sclera, and other dysmorphic features. Previously, chromosome 22q11 deletion analysis for velocardiofacial syndrome and *COL1A1* and *COL1A2* sequencing were negative. Iliac crest biopsy revealed low trabecular connectivity, abundant osteoid, and active bone formation featuring widely-spaced tetracycline labels consistent with osteogenesis imperfecta (OI). Alendronate treatment improved her bone density. Copy number microarray excluded a contiguous gene syndrome. Sanger sequencing of *LRP5*, *PP1B*, *FKBP10*, and *IFITM5* was negative. Instead, exome sequencing revealed two novel missense variants in *PLOD2* (exon 8, c.797G>T, p.Gly266Val; exon 12, c.1280A>G, p.Asn427Ser) which encodes procollagen-lysine, 2-oxoglutarate 5-dioxygenase 2 (lysyl hydroxylase 2, LH2). In ExAC, absence (Asn427Ser) and low frequency (Gly266Val, 0.0000419) implicated both as mutations. The father carried the exon 8 mutation whereas the mother and sister carried the exon 12 defect, and all were well. Autosomal recessive (AR) *PLOD2* mutations cause Bruck syndrome 2 (OMIM #609220) reported in at least 20 cases/15 families as a rare OI featuring congenital contractures of large joints, typically with pterygia, a feature not seen in our patient. Defective LH2 causes under-hydroxylated type I collagen telopeptides compromising crosslinking. Most *PLOD2* missense mutations causing Bruck syndrome 2 are located in exons 17-19 (of 20 total) encoding the C-terminal domain having LH activity, but others are truncating defects (frameshift, splice site mutations) found throughout the gene. AR *PLOD2* mutations causing mild or moderate OI without congenital contractures have been reported in only three reports. Our patient represents the fourth such patient. Her *PLOD2* missense mutations are in the central domain of unknown function, and one (Gly266Val) is the most N-terminal missense defect to date. Hence, our patient's novel *PLOD2* mutations underlie a distinctive skeletal phenotype. Atypical Bruck syndrome should be considered as a possibility in *COL1A1* and *COL1A2*-negative OI cases.

1031F

Loss of function variants in *KDM4C* in Ollier disease: A methylation-dependent mechanism for tumor development. S. Robbins^{1,2}, C. Halde-
man-Engler³, A. Dowiak⁴, E. Cirulli⁵, G. Semenza^{6,7}, D. Gorkin⁴, D. Valle^{1,6},
N.L.M. Sobreira^{1,6}. 1) McKusick-Nathans Institute of Genetic Medicine, Johns
Hopkins School of Medicine, Baltimore, MD., USA; 2) Predoctoral Training
Program in Human Genetics, Johns Hopkins School of Medicine, Baltimore,
MD., USA; 3) Fullerton Genetics Center, Mission Health, Asheville, NC., USA;
4) USCD Center for Epigenomics, University of California San Diego, San Di-
ego, CA., USA; 5) Duke University School of Medicine, Durham, NC., USA; 6)
Department of Pediatrics, Johns Hopkins School of Medicine, Baltimore, MD.,
USA; 7) Institute for Cell Engineering, Johns Hopkins School of Medicine,
Baltimore, MD, USA.

Ollier disease (OD) is a rare disorder characterized by the development of multiple enchondromas which can cause skeletal deformity and fractures. Maffucci syndrome (MS) is a related disorder characterized by enchondromas and vascular anomalies. Patients with OD or MS are also at increased risk for cancer, mainly chondrosarcomas and gliomas. Amary et al., (2011) and Pansuriya et al., (2011) identified specific somatic mutations in *IDH1* or *IDH2* in 80% of enchondromas, vascular anomalies, and chondrosarcomas isolated from patients with OD and MS. *IDH1* and *IDH2* catalyze the oxidative decarboxylation of isocitrate to α -ketoglutarate (α -KG). Specific gain of function variants found in patients with OD and MS convert α -KG to 2-hydroxyglutarate, hypothesized to be an oncogenic metabolite. We present a 16-year-old boy with hypotonia, mild learning disability, dysphagia, vocal cord dysfunction, short stature, and multiple enchondromas in his hands and feet first diagnosed at 3 years of age. We performed Sanger sequencing on DNA isolated from an enchondroma and found no variants in *IDH1* or *IDH2*. However, whole exome sequencing of his leukocyte DNA revealed compound heterozygous missense variants in *KDM4C*, [c.11A>G, p.Tyr4Cys];[c.2320G>A, p.Ala774Thr], each inherited from one of his parents. *KDM4C*-Tyr4Cys (gnomAD MAF = 1.1e-5) disrupts a conserved tyrosine and is highly deleterious by GERP and SIFT; *KDM4C*-Ala774Thr (gnomAD MAF = 1.4e-5) is deleterious by SIFT and Poly-Phen2. *KDM4C* encodes JMJD2C, an α -KG dependent histone demethylase that acts as a co-activator of HIF1A, and like HIF1A, JMJD2C levels increase in response to hypoxia (Luo et al., 2012). By immunoblot analysis of extracts of fibroblasts cultured under normoxia from our patient, we found reduced levels of JMJD2C (0.62 fold of control). After 48 hours of culture under hypoxic conditions, JMJD2C levels in control fibroblasts increased 2.4-fold, while levels in our patient fibroblasts increased only 1.2-fold, or 0.32-fold of control. Currently, we are performing ChIP-seq to determine the effect of these loss of function *KDM4C* variants on the H3K9me3 and H3K36me3 demethylase activity of JMJD2C. To further explore the effects of these *KDM4C* variants on the expression of downstream genes, we are performing qPCR and RNA-seq. Our results implicate *KDM4C* in the development of enchondromas and other features in our patient and uncover a possible mechanism for the association of *IDH1/IDH2* with cancer.

1032W

An unknown problem in a well-known disorder: OI fractures show delayed healing and increased possibility of re-fracture in murine models.

J. Zieba, E. Munivez, M.M. Jiang, B.H. Lee. Baylor College of Medicine, Houston, TX.

Osteogenesis Imperfecta (OI) is the most common genetic bone dysplasia that is phenotypically and genetically complex. It is characterized by bone deformities and fractures caused by low bone mass and impaired bone quality. Roughly 85-90% of cases are dominantly inherited and result from mutations in genes encoding type I collagen (*COL1A1* and *COL1A2*), the major protein of the bone matrix. 10-15% of OI cases are recessively inherited and the majority result from mutations in members of the prolyl-3-hydroxylation complex including Cartilage Associated Protein (CRTAP). OI patients are at an increased risk of fracture throughout their lifetimes and anecdotal evidence suggests successful fracture recovery. However, non-union has been reported in 24% of fractures and 52% of osteotomies and many stabilization techniques result in additional surgery due to re-fracture. Additionally, re-fractures typically go unreported making the frequency of re-fractures in OI patients unknown. Thus, there is an unmet need to better understand the mechanisms by which OI affects fracture healing. It is our hypothesis that OI fractures undergo suboptimal healing and that this process results in ultimately weaker bone leading to the increased possibility of re-fracture. Using an open tibial fracture model, we show a decrease in callus size in both *Col1a1*^{CG100/+} and *Crtap*^{-/-} OI mouse models post-fracture indicating delayed healing and decreased callus cell proliferation. Additionally, fracture calluses in both models exhibited a significant decrease in polar moment of inertia (pMOI) indicating a decrease in resistance to torsional stress supporting a potential functional deficit in newly healed bone. This data provides valuable insight into the effect of the ECM on fracture healing, a poorly understood mechanism. Most importantly, we performed biomechanical testing via three-point-bending of fully healed *Crtap*^{-/-} tibia to determine the mechanical strength of the fracture site. In wild type bone, the healed fracture site resulted in stronger bone when compared to the unfractured tibia. However, *Crtap*^{-/-} healed fractured tibia are mechanically weaker than the contralateral unfractured bone. This implies the possibility that OI fractures do not heal properly and may be a prime location for re-fracture. These data may support aggressive prevention of primary fractures as well as a need for therapies during fracture healing to decrease incidence of refracture and deformity in OI patients.

1033T

Whole exome sequencing identifies known and likely pathogenic variants in a Turkish tooth agenesis cohort.

R. Du¹, N. Dinckan^{2,3}, X. Song¹, L.E. Petty⁴, Z. Coban-Akdemir¹, I.S. Paine¹, Y. Bayram¹, S.N. Jhangiani⁵, E.H. Baugh⁶, A.P. Erdem⁷, H. Kayserili⁸, H. Doddapaneni⁹, J. Hu⁵, D.M. Muzny⁵, Y. Guven⁷, O. Aktoren⁷, N. Wu¹, E. Boerwinkle^{4,5}, R.A. Gibbs^{1,5}, J.E. Below¹, Z.O. Uyguner², A. Letra^{3,9,10}, J.R. Lupski^{1,5,11,12}, J.E. Posey¹, Baylor-Hopkins Center for Mendelian Genomics. 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA; 2) Department of Medical Genetics, Istanbul Medical Faculty, Istanbul University, Istanbul 34093, Turkey; 3) Center for Craniofacial Research, University of Texas Health Science Center at Houston School of Dentistry, Houston, TX 77054, USA; 4) Human Genetics Center, University of Texas Health Science Center at Houston School of Public Health, Houston, TX 77030, USA; 5) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA; 6) Department of Biology, New York University, New York, NY 10003, USA; 7) Department of Pedodontics, Faculty of Dentistry, Istanbul University, Capa, Istanbul 34390, Turkey; 8) Department of Medical Genetics, Koc University, School of Medicine, Istanbul 34010, Turkey; 9) Department of Diagnostic and Biomedical Sciences, University of Texas Health Science Center at Houston School of Dentistry, Houston, TX 77054, USA; 10) Pediatric Research Center, University of Texas Health Science Center at Houston McGovern Medical School, Houston, TX 77030, USA; 11) Department of Pediatrics, Baylor College of Medicine, Houston, TX 77030, USA; 12) Texas Children's Hospital, Houston, TX 77030, USA.

Tooth agenesis (TA), described as the failure of development of one or more permanent teeth, is a common craniofacial abnormality observed in different world populations. The genetic etiology of TA is heterogeneous; more than a dozen genes have been associated with nonsyndromic TA, while more than 80 genes for syndromic forms have been described. In this study, we applied whole exome sequencing to identify potential pathogenic variants for TA in 15 multiplex Turkish families affected with TA. Parental consanguinity was reported in 13 pedigrees. Array-based genotyping and linkage analysis were also employed to localize regions segregating with TA. Sanger sequencing and segregation analysis were performed for all variants of interest. Of these 15 families, six were characterized by syndromic TA, and potentially pathogenic variants identified in six genes. In particular, variants in *KREMEN1* (c.146C>G:p.T49R and c.773_778del:p.F258_P259del) were identified in two families, respectively; variants in *BCOR* (c.1651G>A:p.D551N) and *WNT10A* (c.682T>A:p.F228I) were identified in one family that had a dual molecular diagnosis; and variants in *TSPEAR* (c.[1726C>T; 1728delC]: p.V576Lfs*38), *ANTXR1* (c.1312C>T:p.R438C), and *LAMB3* (c.547C>T:p.R183C) are each identified in one family. Nine families were characterized by nonsyndromic TA, and putative deleterious variants were identified in six genes. Variants in *WNT10A* (c.697G>T:p.E233*, c.682T>A:p.F228I, and c.433G>A:p.V145M) were identified in six families, one of whom had a dual molecular diagnosis including a variant in *LAMA3* (c.1097G>A:p.R366H); variants in *LRP6* (c.3607+3_6del) and *TSPEAR* (c.1877T>C:p.F626S) were identified each in one family; and variants in *DKK1* (c.548-4G>T), *COL17A1* (c.3277+3G>C), and *LAMA3* (c.2798G>T:p.G933V) segregated with TA in one family, suggesting potential oligogenic inheritance. In conclusion, we identified variants in four WNT signaling pathway genes (*WNT10A*, *LRP6*, *KREMEN1*, *DKK1*), consistent with the role of this pathway in tooth development. The genotype-phenotype relationships described in the present study further represent an expansion of the clinical spectrum associated with *ANTXR1*, *BCOR*, *COL17A1*, *LAMA3*, *LAMB3*, and *TSPEAR*. Our data also support potential multilocus genomic variation, or mutational burden, in three families, underscoring the underlying genetic complexity in this common complex trait.

1034F

Patient with skeletal dysplasia, importance of medical criteria in *RUNX2* spectrum. *J. Rojas*^{1,2}, *A. Paredes*¹. 1) Instituto de Genética Humana, Pontificia Universidad Javeriana, Bogotá D.C., Colombia; 2) Clínicos programas de atención integral S.A.S.

RUNX2 (CBFA1) is a master regulatory gene in bone formation. Mutations in this gene have been described in patients with cleidocranial dysplasia (CCD [MIM 119600]) and metaphyseal dysplasia-maxillary hypoplasia-brachydacty syndrome (MDMHB [MIM 156510]), which is a rare autosomal dominant genetic disease, characterized by the triad of hypoplastic or aplastic clavicles, persistence of wide-open fontanels and sutures and multiple dental abnormalities. In general, skeletal abnormalities include short stature, scoliosis, genu valgum, pes planus, a wide pubic symphysis, dysplastic scapulae, and coxa vara. Nevertheless, the range of clinical manifestations can be very broad, where mild cases as isolated dental anomalies, may be underdiagnosed. We present a male patient of 8-years-old with short stature, disproportionate, dental anomalies, frontal bossing, midface hypoplasia, chest with pectus excavatum, scoliosis, genu valgum. The patient was previously studied by another genetics, whom consider a skeletal dysplasia and request a skeletal dysplasia panel of 59 genes, which report a variant of uncertain clinical significance in the *DYM* gene, however the phenotype of the patient does not correspond to that described the literature for diseases associated with mutations in this gene. We extend the molecular study for skeletal dysplasia, which reports the variant c.884C>T (p.Pro295Leu), classified as VUS in the *RUNX2* gene. The *RUNX2* gene (6p21) is a member of the RUNX family of transcription factors, which encodes a nuclear protein essential for osteoblastic differentiation and skeletal morphogenesis. The variant found in the patient, was neither in ExAC nor 1000 genomes, the deletion in this position results in CCD (HGMD CD992745). Currently some genotype-phenotype correlations have been established for dental manifestations, it is essential to identify characteristics that allow limiting the etiological suspicion, to perform the molecular diagnostic confirmation supported by the patient's phenotype. As we know more about the variability of expression of a disease, a whole clinical spectrum is recognized, as is the case of CCD Spectrum Disorder, in which a patient may have isolated dental anomalies, milder cases, skeletal compromise, such as aplasia of the clavicles and anomalies in the closure of cranial sutures, in typical cases. This is where the experience and semiology of the clinician impacts on the direction of molecular studies and the correct diagnosis. .

1035W

Bruck syndrome, a clinical case presentation and literature review. *L. Mora*^{1,2,3}, *A. Paredes*^{1,2}. 1) Pontificia Universidad Javeriana, Bogotá D.C., Colombia; 2) Instituto de Genética Humana; 3) Meintegral S.A.S.

Introduction: Bruck's syndrome (BRKS1 [MIM 259450]) was first described by Bruck et al. in 1897. It is a monogenic disease with autosomal recessive inheritance mechanism, characterized by the association of osteogenesis imperfecta and congenital joint contractures. There are two genes associated with this disease, which allow its classification. Its clinical manifestations include short stature, osteoporosis, wormian bones, scoliosis secondary to alterations in the vertebrae, bone fragility and progressive joint contractures associated with multiple pterygium. Case report: We present the case of a 26-month-old female patient, daughter of non-consanguineous parents who cursed with fractures in-uterus, later associated with joint contractures. Taking into account the physical examination findings, we decided to enlarge the skeletal dysplasia panel, where the homozygous pathogenic variant c.612C>G, p.(Tyr204*) in the *FKBP10* gene was reported. Discussion: Bruck syndrome is classified into 2 types, according to the gene involved. Type 1, the mutation is found in the *FKBP10* gene (17q215), in type 2 in the *POLD2* gene (3q24), which codes for telopeptide lysyl hydroxylase. Mutations in these genes cause deficiency of the bone-specific lysyl telopeptide hydroxylase enzyme, which generates an aberrant cross-linking of type I collagen. Also, pathogenic variants in *FKBP10* are associated with osteogenesis imperfecta type XI and Bruck syndrome, which can be differentiated from each other by means of the physical examination, like in this case. Due to unknown prevalence of Bruck syndrome and low knowledge we bring this case report. Conclusion: Although skeletal dysplasia is a large and heterogeneous group of diseases, it is essential to identify characteristics that limit diagnostic suspicion; as in the case presented, where the clinical concept of the doctor regarding the recognition of the association between pathological fractures and congenital joint contractures associated with pterygium, was essential to define the molecular test and diagnostic confirmation of Bruck syndrome.

1036T

A novel mutation in *NBAS* causes SOPH syndrome. G.F. Godínez-Zamora^{1,2}, P. Baeza-Capetillo¹, A. Villaseñor-Domínguez¹, A. Reyes-de-la-Rosa¹, C. Armijos-Torres¹, C. García-Delgado¹, V. Morán-Barroso¹, J. Aguirre-Hernández¹. 1) Laboratory of Genomics, Genetics and Bioinformatics, Hospital Infantil de México. México city; 2) Universidad Nacional Autónoma de México,.

Background: SOPH syndrome (OMIM 614800) is a disease characterized by short stature, optic nerve atrophy and Pelger-Huet anomaly. This disease was initially described in Siberian Yakut people, where most of the cases have been found. A small number of patients have been found in Europe, and none in other regions across the world. This disease is characterised by global developmental delay, bone chronological age delay, compression of the medullar channel and metabolic alterations. The syndrome shows phenotypic variability, with some patients, but not all, having liver failure. The gene underlying this syndrome is *NBAS*, which codes for a protein that is part of the syntaxin 18 sub complex that plays a role in retrograde transport from the endoplasmic reticulum to the Golgi apparatus. Alterations in *NBAS* have been shown to alter protein recycling and lead to defects in protein glycosylation. Here we describe a girl presenting short stature, bilateral optic atrophy and other multisystem features; global developmental delay, bone chronological age delay, compression of the medullar channel C7 to T5 and facial features suggestive of SOPH syndrome. **Materials and methods:** Whole exome sequencing was performed in the proband and her parents. After read alignment and variant calling, variants were filtered by depth, consequence, segregation, and frequency in public databases, as well as in a local database. Remaining variants were prioritised with the software Exomiser. **Results:** Whole exome sequencing data of the patient and her parents showed that the proband was homozygous for a novel transversion in the acceptor splice site immediately before exon 15 of *NBAS* (ENST00000281513.5:c.1342-1G>C hg19 2:15614449 C>G). Both parents were found to be carriers of this variant. **Conclusions:** SOPH is a very rare disease, previously known only from Siberia and a very small number of European patients. Here, a novel splice site mutation is reported in a very young female patient, from a small population in southern Mexico, presenting phenotypic features of SOPH.

1037F

Molecular and clinical characterization of Turkish patients with Steel syndrome. J.M. Fatih¹, G. Yesil², D. Pehlivan^{1,3}, Y. Bayram⁴, Z.C. Akdemir¹, Y. Sahin⁵, E. Karaca¹, S.N. Jhangiani⁶, D.M. Muzny⁶, R.A. Gibbs^{1,6}, J.E. Posey¹, J.R. Lupski^{1,7,8}, Baylor-Hopkins Center for Mendelian Genomics. 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Department of Medical Genetics, Bezmi Alem Vakif University Faculty of Medicine, Istanbul, Turkey; 3) Section of Neurology, Department of Pediatrics, Baylor College of Medicine, Houston, Texas; 4) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York; 5) Department of Medical Genetics, Necip Fazil City Hospital, Kahramanmaraş, Turkey; 6) Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas; 7) Department of Pediatrics, Baylor College of Medicine, Houston, Texas; 8) Texas Children's Hospital, Houston, Texas.

Steel syndrome (STLS) is an autosomal recessive disorder associated with osteochondrodysplasia, a rare form of skeletal dysplasia described in the Puerto Rican population with clinical manifestations that include dislocated hips and radial heads, short stature, scoliosis, carpal coalition, characteristic facies, cervical spine anomalies, and pes cavus. Clinical findings are limited to the skeletal system. We previously identified (Gonzaga-Jauregui, *et al.* (2015) *Eur J Hum Genet* 23: 342-346) a homozygous *COL27A1* variant (c.2089G>C; p.Gly697Arg) and provided evidence for a Puerto Rican founder allele. Since then, two distinct homozygous pathogenic variants have been reported in families from India and the UAE. We studied three unrelated Turkish families with reported parental consanguinity. We applied whole-exome sequencing (WES) in two patients from different families with a clinical diagnosis of an uncertain skeletal dysplasia. Due to a clinical diagnosis of STLS in the third proband, they were sent directly for Sanger sequencing of *COL27A1*. Clinical evaluation revealed distinctive craniofacial features in addition to rhizomelic shortening of the upper limbs, bowing of the femurs, pes calcanovalgus, and joint laxity. WES revealed a homozygous *COL27A1* missense variant (c.2683G>A; p.Gly895Arg) in patient BAB5133 and a homozygous *COL27A1* frameshift deletion (c.4976_4980delGAGGA; p.Gly1660Aspfs*3) in patient BAB8900. The third patient, BAB10793, who was clinically diagnosed with Steel syndrome, underwent Sanger sequencing and was found to have a homozygous 11bp *COL27A1* deletion involving the intron1/exon2 boundary causing c.63-5_68delTCTAGGGGGTT; p.22_23delGlyPhe. For the families with available data, the *COL27A1* gene mapped within large segments of absence of heterozygosity (AOH), of 9.2Mb for BAB5133 and 3.68Mb for BAB8900, suggesting identity by descent and consistent with the reported parental consanguinity. Our findings demonstrated that STLS is a clinically recognizable disorder with distinctive facial dysmorphic features and we further delineate the skeletal findings of this syndrome. Additionally, we expanded the mutational spectrum of *COL27A1* variants and further confirm the gene-disease association in STLS.

1038W

TBX6 missense mutations in congenital scoliosis: Expanding the**mutational spectrum of a complex trait.** W. Chen^{1,2,3}, J. Lin^{1,2}, X. Li^{4,7}, J.Liu^{5,6}, X. Song³, R. Du³, G. Qiu^{1,4,6}, J. Posey², F. Zhang^{12,13,14}, P. Liu^{3,8}, Z. Wu^{4,5,7},J. Lupski^{3,9,10,11}, N. Wu^{1,3,4,6}, *Deciphering Disorders Involving Scoliosis and**COMorbidities (DISCO) study.* 1) Department of Orthopaedic Surgery, Peking Union Medical College Hospital, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing 100730, China; 2) Graduate School of Peking Union Medical College, Beijing 100005, China; 3) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA; 4) Beijing Key Laboratory for Genetic Research of Skeletal Deformity, Beijing 100730, China; 5) Breast Surgical Oncology, Cancer Hospital of Chinese Academy of Medical Sciences, Beijing 100021, China; 6) Medical Research Center of Orthopaedics, Chinese Academy of Medical Sciences, Beijing 100730, China; 7) Department of Central Laboratory, Peking Union Medical College Hospital, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing 100730, China; 8) Baylor Genetics Laboratory, Houston, TX 77021, USA; 9) Department of Pediatrics, Baylor College of Medicine, Houston, TX 77030, USA; 10) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA; 11) Texas Children's Hospital, Houston, TX 77030, USA; 12) Obstetrics and Gynecology Hospital, State Key Laboratory of Genetic Engineering at School of Life Sciences, Institute of Reproduction and Development, Fudan University, Shanghai 200433, China; 13) Key Laboratory of Reproduction Regulation of NHFPC, Collaborative Innovation Center of Genetics and Development, Fudan University, Shanghai 200433, China; 14) Shanghai Key Laboratory of Female Reproductive Endocrine Related Diseases, Shanghai 200090, China.

Congenital scoliosis (CS) is a rare and severe malformation with a prevalence of ~0.5-1‰. Genetic underpinnings have been implicated in CS pathogenesis. We previously reported that compound inheritance of a rare null mutation and a hypomorphic common variant haplotype of *TBX6* could explain ~11% of CS cases. Here, we elucidate the role of missense mutations of *TBX6* in CS pathogenesis and further investigate the genetic underpinnings, disease biology and etiology of CS. We enrolled patients clinically diagnosed with CS in Peking Union Medical College Hospital between October 2010 and November 2017 as part of the Deciphering Disorders Involving Scoliosis and COMorbidities (DISCO) study. The CS diagnosis was confirmed with radiological imaging and patients with additional syndromic features were excluded. We performed whole exome sequencing for the entire cohort (N = 499) and identified 10 missense mutations of *TBX6* in 16 patients (16/499, 3.2%). Among them, four patients present as CS with hemi/wedge-vertebrae, three with segmentation defects and ten with hemi-/wedge-vertebrae combined with segmentation defects in thoracic or lumbar vertebrae. Five mutations were not present in ExAC, 1000 Genomes, ESP6500 databases, and our in-house database. CADD score showed three mutations (c.424G>T, c.440A>T and c.929C>T) were damaging/likely damaging (CADD = 21, 16.01 and 20.6, respectively), and two (c.1184G>A and c.1300C>T) as tolerated (CADD = 8.97, 11.06, respectively). The remaining five mutations were rarely present with frequency < 0.001 in ExAC, all absent in 1000 Genomes, < 0.005 in ESP6500, and 1, 7, 3, 0, 0 in 998 subjects respectively in our in-house database. CADD showed four mutations (c.356G>A, c.434C>T, c.499C>T and c.700G>C) as damaging/likely damaging (CADD = 34, 16.65, 20.9 and 19.27), and one (c.1112C>T) as tolerated (CADD = 11.36). We further investigated the transcriptional activity of *TBX6* mutations using a dual luciferase reporter assay. Interestingly, we identified two mutations (c.356G>A and c.424G>T) with significantly decreased luciferase activity compared to the wild-type. In addition, we identified the hypomorphic haplotype 'T-C-A' (the co-occurrence of three common SNPs, rs2289292-rs3809624-rs3809627) in *trans* in 9 out of the 16 subjects. Our findings indicate that the likely gene disrupting missense mutations of *TBX6* could contribute to the etiology of CS, thus expanding the mutational spectrum of *TBX6* associated congenital scoliosis (TACS).

1039T

Combining large-scale in vitro functional testing with protein stability simulations to develop new insights into hypophosphatasia genotype/phenotype correlation. G. del Angel¹, E. Mornet², J. Reynders¹.

1) Alexion Pharmaceuticals Inc., Lexington, MA; 2) Laboratoire de génétique constitutionnelle prénatale et postnatale. Centre Hospitalier de Versailles. Le Chesnay, France.

Hypophosphatasia (HPP) is a rare, systemic, inherited metabolic disorder that can involve bone, muscle, kidney and other organs. The disease is characterized by low alkaline phosphatase (ALP) enzymatic activity and is typically caused by mutations in the tissue non-specific alkaline phosphatase gene (*ALPL*, OMIM:171760). It has been previously shown that clinical presentation severity correlates with mutant protein *in vitro* enzymatic activity and with the strength of dominant-negative effects (Zurutuza 1999; Lia-Baldini 2001). Our study had two aims: first, to expand our understanding of *ALPL* genotype/phenotype relation by characterizing *in vitro* a larger portion of population variants, and to use this *in vitro* data to develop methods to predict variant enzymatic activity level. We first tested 62 *ALPL* missense mutations for residual enzymatic activity and dominant negative effect using site-directed mutagenesis and transient co-transfections of WT/mutant plasmids into MDCK cells. Twenty-nine of these mutations had been previously characterized and were used to validate the assay. The remaining 33 missense variants had not been previously characterized *in vitro*, and 24 of them would be marked as Variants of Unknown Significance (VUS) according to ACMG guidelines. Of these 33, 15 showed low residual enzymatic activity and 2 showed dominant negative effects, thereby increasing their likelihood of pathogenicity. Once combined with previously reported results, this dataset of 128 *in vitro*-tested variants, which encompasses all *ALPL* missense variants with gnomAD carrier frequencies above 1/10,000 (Lek 2016), was used to develop and validate a new algorithm that predicts low enzymatic activity for *ALPL* missense variants. We first measure effects of amino-acid changes on ALP dimer thermodynamic equilibrium by using free energy perturbation methods (Steinbrecher 2017), and we predict low enzymatic activity based on changes in protein structure stability, affinity and metal cross-link presence, yielding prediction accuracy, as measured by the area under the receiver operating curve (AUC), of 0.78. We then use in-silico scores like Polyphen-2 (Adzhubei 2010) as additional inputs to the model to improve performance with an AUC metric of 0.87, compared with an AUC of 0.83 for Polyphen-2 alone. This model can be used to score novel *ALPL* variants, and may help get new insights into HPP genotype/phenotype correlation.

1040F

Multilocus genomic variants may explain the clinical heterogeneity of patients with KBG syndrome. C. Zhang¹, J.F. Mazzeu^{2,3}, S.V. Braz², B.M. Ferreira², A.C. Acevedo-Poppe², Z.C. Akdemir¹, S.N. Jhangiani¹, J.E. Posey¹, D.M. Muzny⁴, R.A. Gibbs⁵, C.M.B. Carvalho¹, J.R. Lupski^{1,4,5,6}. 1) Department of Molecular and Human Genetics, BCM, Houston, TX; 2) University of Brasilia, Brasilia, Brazil; 3) Robinow Syndrome Foundation, Anoka, MN; 4) Human Genome Sequencing Center, BCM, Houston, TX; 5) Department of Pediatrics, BCM, Houston, TX; 6) Texas Children's Hospital, Houston, TX.

KBG syndrome (MIM:148050) is an autosomal dominant disorder characterized by macrodontia of the upper central incisors, distinctive craniofacial findings, skeletal anomalies, and neurodevelopmental involvement including developmental delay and intellectual disability (DD/ID). Diagnosis can be challenging due to extensive phenotypic variability and overlap with other neurodevelopmental conditions. In addition, many of the features, including intellectual disability, can be mild, and none are pathognomonic for diagnosis. Heterozygous variants in *ANKRD11* leading to haploinsufficiency are the cause of KBG syndrome with more than 60 mutations identified. We hypothesize that some of the clinical heterogeneity observed in patients with a diagnosis of KBG syndrome can be partially explained by multiple genetic variants in addition to mutations in *ANKRD11*. To test this hypothesis, we studied two patients with KBG-like phenotypes by SNP array and whole exome sequencing (WES). Both of these patients have *de novo* indel variants affecting exon 9 of 13 exons in *ANKRD11*. In addition, we also identified a hemizygous likely pathogenic variant in *MED12* in the first patient. This rare variant is a conserved missense at position c.3571G>A; p.Asp1191Asn. Missense variants in *MED12* have been implicated in three distinct X-linked intellectual disability syndromes with distinctive clinical features, phenotypes that largely overlap with KBG syndrome. Thus, some distinct phenotypes such as speech problems and upslanted palpebral fissures are more likely caused by the variant in *MED12*. Wide fontanel, short stature, anteverted nares, long philtrum and cryptorchidism, are more likely caused by the truncated variant in *ANKRD11*. Other overlapping phenotypes like high forehead and intellectual disability may be associated with either of these two variants. For the second patient, we identified a 18Mb AOH region at position 14q21.1q23.1 (g.41435763-59516209). Some distinct phenotypes, such as high forehead, upturned nose, anteverted nares, long philtrum, and 5th finger clinodactyly are more likely to be caused by truncated variant in *ANKRD11*. Other phenotypes, like development delay and intellectual disability may be associated with either the variant in *ANKRD11* or maternal/paternal AOH on chromosome 14. In summary, our data provide evidence that clinical phenotypic variability of KBG syndrome may result from multilocus genomic variation in some patients.

1041W

A missense mutation of COL2A1 gene in an achondrogenesis type II neonate (Case report). C.H. Lee^{1,2}, Y.N. Yang^{1,2,3}, P.J. Ko^{2,4}, M.L. Yeh^{2,4}, L.M. Chen^{1,2}, Y.T. Su^{1,2}, N.C. Lee^{5,6}. 1) Department of Pediatrics, E-DA Hospital, Kaohsiung, Taiwan; 2) School of Medicine, I-Shou University, Kaohsiung, Taiwan; 3) Graduate Institute of Medicine, College of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan; 4) Division of Pediatric Surgery, Department of Surgery, E-DA Hospital, Kaohsiung, Taiwan; 5) Department of Medical Genetics, National Taiwan University Hospital, Taipei, Taiwan; 6) Department of Pediatrics, National Taiwan University Hospital, Taipei, Taiwan.

Introduction: Achondrogenesis type II (OMIM #200610) was a rare lethal skeletal disorder which affected cartilage and bone development. The major clinical characteristics were extremely short trunk and limbs, a narrow chest with short ribs, and underdeveloped lung. Achondrogenesis type II was an autosomal dominant inheritance and caused by *de novo* mutation in the collagen type II-1 *COL2A1* gene. This gene made a protein that formed type II collagen which was essential for the normal development of bones and other connective tissues. **Case presentation:** This male neonate was born at 40 weeks and 2 days of gestation age via vaginal delivery and his parents were without consanguinity or know medical problems. After birth, he was admitted to neonatal intensive care unit due to severe tachypnea, dyspnea, and subcostal retraction. The physical examination showed body length 38cm (<3 percentile), body weight 2755gm (3-15% percentile), head circumference 36cm (85-97 percentile), arm span 36cm, pseudo low-set ears, mid-facial hypoplasia, anteverted nostrils, cleft palate, micrognathia, short neck, narrow chest cavity with bulging of anterior chest, prominent abdomen, bilateral simian crease, clinodactyly of bilateral 5th fingers, and bilateral clubfoot. Radiographs revealed short ribs, short and broad long bones, and a lack of ossification in the spine and pelvis. **Method and result:** Whole exome sequencing was performed for skeletal dysplasia survey. The heterozygote missense mutation c.1529G>A (p.Gly510Asp) was identified in exon 24 of *COL2A1* gene which was also confirmed by Sanger sequencing. According to Leiden Open Variation Database (LOVD-COL2A1 2015 update) and clinical features, this case was accord with achondrogenesis type II. **Conclusions:** To our knowledge, this case was the third time this mutation was reported in achondrogenesis type II. Through whole exome sequencing, the patient could gain a rapid, accurate, and cost-efficient diagnosis. In clinical, it was much important for appropriate medical care, identifying potential treatment, prognosis of the disorder, genetic counseling, and access to support groups.

1042T

Chondrodysplasia Punctata (CDPX1) presenting with respiratory distress. G.E. Tiller¹, D.D. Tieu², E.H. Rhee³. 1) Genetics, Kaiser Permanente, Los Angeles, CA; 2) Head and Neck Surgery, Kaiser Permanente, Los Angeles, CA; 3) Pediatrics, Kaiser Permanente, Los Angeles, CA.

X-linked recessive chondrodysplasia punctata (type 1; CDPX1) is a rare skeletal dysplasia presenting with mild dysmorphic features, epiphyseal +/- soft tissue stippling, and brachytelephalangy (shortened distal phalanges). We describe a term male infant with CDPX1 who presented with mild dysmorphic features and respiratory distress. He exhibited mild hypertelorism, epicanthal folds, small nose with flat nasal bridge, retrognathia, and malar hypoplasia. The patient developed stridor and respiratory distress in the first 12 hours of life. He was transferred to our NICU and underwent supraglottoplasty for shortened aryepiglottic folds; no tracheal stenosis was evident during direct laryngoscopy. Due to persistent stridor and increased work of breathing, a tracheostomy was performed at one month of life, which relieved his symptoms. Imaging studies revealed C3-C4 fusion, brachytelephalangy, as well as stippling of proximal femurs, ankles, and posterior thoracolumbar spinal elements. No cervical spinal stenosis nor CNS anomalies were noted. Microarray revealed no deletions or duplications, but sequencing of the ARSE gene revealed a Gly141Glu substitution. Glycine 141 is highly evolutionarily conserved, and predictive software suggests that the amino acid substitution is deleterious. This case serves to illustrate additional airway and cervical spinal anomalies that can be seen in CDPX1, as well as identifying an additional novel missense variant in ARSE.

1043F

Identification of a diagnostic fusion transcript in a patient with multiple exostoses clinically negative for EXT variants. G.R. Oliver¹, P. Blackburn¹, F. Pinto e Vairo¹, M.S. Ellingson², D. Babovic-Duksanovic¹, E.W. Klee¹. 1) Center for Individualized Medicine, Mayo Clinic, Rochester, MN; 2) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN.

Hereditary Multiple Exostoses (HME) are cartilaginous outgrowths inherited in an Autosomal Dominant fashion. Mutations in *EXT1* and *EXT2* are usually causative with 70-75% of cases due to protein truncating point mutations and 10% caused by deletions involving single or multiple exons. Some have suggested the remaining cases may be due to mutation at novel loci, while others have highlighted the role of mosaic mutations in *EXT1* and *EXT2*. We describe the case of a nine-year-old male born to healthy parents of mixed European descent. At age 4 months, the patient was referred to Clinical Genomics for failure to thrive. At 10 months, he was diagnosed with global developmental delay, short stature, dysmorphisms, and with multiple exostoses that have worsened at 6 years of age. Focal epilepsy was diagnosed at 2 years. The patient had a normal karyotype, and array genomic comparative hybridization (aCGH) revealed a de novo interstitial deletion of chromosome 16p13.3 spanning approximately 219 kilobases and encompassing 11 known genes unrelated to patient's phenotype. Multiplex ligation-dependent probe amplification (MLPA) and direct sequencing of all coding exons of *EXT1* was negative for mutations. Whole Exome Sequencing revealed a heterozygous, maternally inherited splice mutation in *DCX* (c.809-1G>A). This mutation was deemed diagnostic of the patient's neurological phenotype but exostoses remained unexplained. RNA sequencing revealed a candidate intrachromosomal *SAMD12-EXT1* fusion transcript. Subsequent reinspection of clinical aCGH identified a region of low probe intensity potentially indicative of a sub-calling threshold chromosomal loss between *SAMD12* exon 1 and *EXT1* exon 1. Re-analysis with a second array of higher probe density reinforced these findings and the event was categorized as a likely mosaic deletion resulting in gene fusion and *EXT1* loss of function. The event was deemed diagnostic of the multiple exostoses. This case corroborates previously published work describing the need to consider mosaic loss in cases of *EXT* mutation-negative HME patients. Furthermore the findings highlight the value of extended investigation in the resolution of undiagnosed cases and exemplify an emerging reality in genetic analysis whereby traditional approaches may benefit from novel analyses and hypothesis-building to maximize diagnostic yield.

1044W

KIAA1217: Evidence for a novel candidate gene associated with cervical segmentation defects. N. AlDhaheer¹, N. Wu², S. Zhao², Z. Wu², R. Blank³, C. Raggio⁴, M. Halanski⁵, K. Noonan⁵, B. Nemeth⁵, S. Sund⁵, N. Sobreira¹, P. Giampietro⁶. 1) McKusick Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, United States; 2) Peking Union Medical College Hospital, China; 3) Medical College of Wisconsin, Milwaukee, WI; 4) Hospital for Special Surgery, NY; 5) University of Wisconsin-Madison, WI; 6) Drexel University College of Medicine, Philadelphia, PA.

Background: Cervical segmentation defects (CSD) are caused by alterations in somitogenesis, which may occur in association with other renal, cardiac, or spinal cord birth defects (30-60% of cases), and are associated with disabling pain from myelopathy and disc disease. The genetic etiology of most CSD disorders remains unknown and its identification may aid in the development of novel treatment and prevention strategies. **Methods:** We performed whole exome sequencing (WES) on a cohort of eight unrelated probands with CSD and a constellation of additional anomalies affecting their cardiac, CNS, and musculoskeletal systems. Using PhenoDB analysis tool, heterozygous and homozygous, rare (MAF<1%) and functional (missense, nonsense, splicing and indels) variants in each proband were selected. These genes and variants were evaluated for their ClinVar, HGMD, OMIM, GWAS, mouse model phenotypes and other annotations to identify the best candidate variants in each proband. Genes that were mutated in 3 or more probands were also selected. **Results:** We identified rare heterozygous variants in KIAA1217 in 3 unrelated probands. In a separate cohort of 35 patients with CSD from DISCO (Deciphering Disorders Involving Scoliosis & COMorbidities) study, we identified compound heterozygous variants in a single proband and heterozygous variants in 5 probands. In total, we identified 10 variants in 9 patients, 3 of which have not been described in gnomAD and one is a stopgain variant. Collectively, patients had cervical segmentation abnormalities, cardiac and central nervous system abnormalities. **Discussion:** KIAA1217 is involved in embryonic skeletal development (The Gene Ontology project). It is also found to be required for normal intervertebral disc development and GWAS studies identified its association with lumbar disc herniation. Mice homozygous for a gene-trapped allele, display malformations of the notochord and caudal vertebrae and may exhibit caudal tail kinks. DECIPHER describes a patient with a 10.44 Mb (97 genes) *de novo*, heterozygous deletion affecting KIAA1217 in a proband with hemivertebrae, abnormality of the pinna, hypoplasia of the radius, low-set ears, and renal agenesis. **Conclusion:** Based on our findings and on what is known about the biology of this gene, we suggest that KIAA1217 pathogenic variants can cause autosomal dominant CSD with or without cardiac and CNS anomalies.

1045T

Diagnostic yield and clinical relevance of exome screening in patients with early-onset scoliosis (EOS). S. Zhao^{1,2}, W. Chen^{1,2,3,4}, Z. Yan^{1,2,3}, W. Li^{5,6,7}, Y. Zhang^{1,2,3}, Z. Akdemir⁸, Y. Zhao^{1,2}, L. Wang^{1,2,3}, R. Du⁴, M. Lin^{1,2,3}, Y. Ye^{1,2}, J. Lin^{1,2,3}, J. Posey⁴, P. Liu^{4,8}, G. Qiu^{1,2,9}, F. Zhang^{5,6,7}, Z. Wu^{2,9,10}, J. Lupski^{11,12}, N. Wu^{1,2,4,9}. **Deciphering Disorders Involving Scoliosis and COMorbidities (DISCO) study.** 1) Department of Orthopedic Surgery, Peking Union Medical College Hospital, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing 100730, China; 2) Beijing Key Laboratory for Genetic Research of Skeletal Deformity, Beijing 100730, China; 3) Graduate School of Peking Union Medical College, Beijing 100005, China; 4) Department of Molecular and Human Genetics, Baylor College of Medicine; Houston, TX 77030, USA; 5) Obstetrics and Gynecology Hospital, State Key Laboratory of Genetic Engineering at School of Life Sciences, Institute of Reproduction and Development, Fudan University, Shanghai 200433, China; 6) Key Laboratory of Reproduction Regulation of NHFPC, Collaborative Innovation Center of Genetics and Development, Fudan University, Shanghai 200433, China; 7) Shanghai Key Laboratory of Female Reproductive Endocrine Related Diseases, Shanghai 200433, China; 8) Baylor Genetics Laboratory, Houston, TX 77021, USA; 9) Medical Research Center of Orthopedics, Chinese Academy of Medical Sciences, Beijing 100730, China; 10) Department of Central Laboratory, Peking Union Medical College Hospital, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing 100730, China; 11) Departments of Pediatrics, Baylor College of Medicine, Houston, TX 77030, USA; 12) Texas Children's Hospital, Houston, TX 77030, USA.

Scoliosis, defined as a lateral spine curvature more than 10°, is a descriptive clinical diagnosis rather than an etiological diagnosis. The age of onset of scoliosis is a key factor in determining the therapeutic strategy. Early onset scoliosis (EOS), onset age before 10, conveys significant health risk to affected children. The etiology of EOS includes vertebral anomalies (CS I, vertebral malformation; CS II, segmentation defect; CS III, mixed condition), neuromuscular defects, and various syndromes, presenting a complex phenotypic context. Although genetic factors are believed to underlie these conditions, genetic testing is not routinely ordered for children with EOS. To investigate the molecular diagnosis rate by whole exome sequencing (WES) among EOS and its associated syndromes, and the prevalence of clinical actionable variants including both primary and incidental findings, we consecutively recruited 500 patients with EOS in Peking Union Medical College Hospital as part of the Deciphering Disorders Involving Scoliosis and COMorbidities (DISCO) study (<http://www.discostudy.org/>), 110 of whom have familial samples. WES was performed on the patients and available family members. Variant calling and annotation was executed by an in-house developed analytic pipeline, followed by data interpretation and reporting according to ACMG guidelines. A total of 106 (21.2%) EOS patients received a molecular diagnosis, including 56 (11.2%) TBX6-associated scoliosis (TACS) patients and 50 (10.0%) patients with reported pathogenic variants. The diagnostic rate varies between scoliotic subgroups: 62/126 (49.2%) in CS I, 8/45 (18%) in CS II, 21/289 (7.3%) in CS III, 3/4 in Neurofibromatosis, 3/10 in neuromuscular scoliosis, 3/3 in osteogenesis imperfecta, 3/3 in achondroplasia, and 3/20 in idiopathic EOS. In 106 patients with molecular diagnoses, 19 (18%) yielded findings of significant clinical relevance, defined by preventive surgical measures (PLOD1, RYR1, COL5A2), multi-system symptomatic treatment (EBP, MYH7, JAG1, PTPN11), and targeted treatment (COL1A1, NF1). In conclusion, WES allows the identification of molecular etiology in 21.2% patients with EOS, and detection of clinically actionable mutations emphasizes the necessity of genetic testing for EOS patients.

1046F

Structural alteration of glycosaminoglycan side chains and spatial disorganization of collagen networks in the skin of patients with musculocontractural Ehlers-Danlos syndrome caused by CHST14/D4ST1 deficiency. T. Koshio^{1,2}, T. Hirose³, N. Takahashi³, P. Tangkawattana⁴, J. Minaguchi³, S. Mizumoto⁵, S. Yamada⁵, N. Miyake⁶, A. Hatamochi⁷, J. Nakayama⁸, T. Yamaguchi^{1,2}, K. Takehana³. 1) Medical Genetics, Shinshu University School of Medicine, Matsumoto, Japan; 2) Center for Medical Genetics, Shinshu University Hospital, Matsumoto, Japan; 3) Laboratory of Microanatomy, School of Veterinary Medicine, Rakuno Gakuen University, Ebetsu, Hokkaido, Japan; 4) Department of Veterinary Anatomy, Faculty of Veterinary Medicine, Khon Kaen University, Khon Kaen, Thailand; 5) Department of Pathobiochemistry, Faculty of Pharmacy, Meijo University, Nagoya, Japan; 6) Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan; 7) Department of Dermatology, Dokkyo Medical University, Mibu, Japan; 8) Department of Molecular Pathology, Shinshu University Graduate School of Medicine, Matsumoto, Japan.

Musculocontractural Ehlers-Danlos syndrome (EDS) due to CHST14/D4ST1 deficiency (mcEDS-CHST14) is a recently delineated type of EDS caused by biallelic loss-of-function mutations in *CHST14*, resulting in depletion of dermatan sulfate (DS). Clinical characteristics of mcEDS-CHST14 consist of multiple malformations and progressive fragility-related manifestations including skin hyperextensibility and fragility. Skin fragility is postulated to result from impaired assembly of collagen fibrils caused by alteration of glycosaminoglycan (GAG) chain of decorin-proteoglycan (PG) from DS to chondroitin sulfate (CS). This systemic investigation of skin pathology of patients with mcEDS-CHST14 includes immunostaining of decorin and transmission electron microscopy-based cupromeronic blue staining to visualize GAG chains. Collagen fibrils were dispersed in the affected papillary to reticular dermis in contrast to regularly and tightly assembled in the controls; and also appeared to have perpendicular arrangement to the affected epidermis in contrast to the parallel arrangement along with the control epidermis. Affected GAG chains were shown to be linear stretching from the outer surface of collagen fibrils, in contrast to those of the controls showing to be curved attaching onto the outer surface of collagen fibrils. Decorin-PG appeared as filamentous forms in the affected skin, whereas thick bundles intermingled with collagen fibers in the controls. This is the first observation of compositional alteration, from DS to CS, of the GAG side chains to cause structural alteration of GAG side chains, which results in spatial disorganization of collagen networks, presumably disrupting recently uncovered ring-mesh structure of GAG side chains surrounding collagen fibrils. McEDS-CHST14 is a critical model to delineate importance of DS as GAG side chains of decorin-PG assembling collagen fibrils in the maintenance of connective tissues.

1047W

A unique COL1A2 mutation in a teenage male with patellar dislocations. E.M. Pereira, E. Guzman. New York-Presbyterian, Morgan Stanley Children's Hospital, Columbia University Medical Center, Department of Pediatrics, Division of Clinical Genetics, NY, NY.

Our patient presented to clinical genetics to rule out a connective tissue disorder. This 16-year-old male is the product of a non-consanguineous union between parents of Dominican descent. After an unremarkable prenatal course, he was noted to have bilateral congenital hip dislocations which resolved over time. At 3 years of age he sustained a femur fracture after an accidental trauma which required a spica cast for 3 months. As a teenage he suffered a left knee injury and subsequently had 4 patellar dislocations (one spontaneously). After corrective surgery at 15 years old there have been no recurrences. The patient had a normal echocardiogram. An ophthalmologic exam was scheduled. Developmentally, the patient had a history of speech delay and early intervention services were in place till he was 5 years old; currently he is in regular education. His family history is significant for a mother, maternal half-brother, and paternal half-sister with minor flexibility but no dislocations. The patient's physical exam did not reveal skin hyperextensibility. There was atrophic scarring around his left knee at the surgical site. He had grey tinged sclerae but otherwise unremarkable features. His Beighton score was 0, but there was mild joint hypermobility. Given his neonatal history and subtle physical findings, chromosomes, microarray, and a collagenopathy panel were sent. Results were positive for a likely pathogenic *COL1A2* variant: c.398G>T giving p.Gly133Val which affects the highly conserved glycine residue important in fibrillary collagen stability. *COL1A2* related disorders have been associated with phenotypes ranging from autosomal dominant forms of osteogenesis imperfecta and 2 forms of Ehlers-Danlos syndrome (cardiac valvular and arthrochalasia type). Though he does have grey tinged sclerae and had a femur fracture at a young age, he has been playing competitive high contact sports in high school and has not sustained further fractures. His normal echocardiogram rules out Ehlers-Danlos syndrome, cardiac valvular type. He does have a history of congenital bilateral hip dislocation, but does not meet clinical criteria for Ehlers-Danlos syndrome, arthrochalasia type (aEDS). Additionally, patients with aEDS have a mutation in exon 6 of *COL1A1* or *COL1A2* while our patient's variant is in exon 9. Lifestyle changes may be warranted given this *COL1A2* variant, but more studies (including familial variant testing) would be important for further classification.

1048T

Homozygous mutation in *PLOD3*, encoding lysyl hydroxylase 3, causes recessive dystrophic epidermolysis bullosa with abnormal anchoring fibrils and deficiency in type VII collagen. H. Vahidnezhad^{1,5}, L. Youssefian¹, A.H. Saeidianan¹, A. Touati¹, S. Pajouhanfar¹, S. Aristodemou², L. Liu², T. Baghdadli³, A. Kariminejad⁴, S. Zeinali⁵, J.A. McGrath⁶, J. Uitto¹. 1) Department of Dermatology and Cutaneous Biology, Sidney Kimmel Medical College, Thomas Jefferson University, Philadelphia, PA, USA; 2) Viapath, St Thomas' Hospital, London, UK; 3) Department of Orthopedic Surgery, Imam Khomeini Hospital, Tehran University of Medical Sciences, Tehran, Iran; 4) Kariminejad-Najmabadi Pathology & Genetics Center, Tehran, Iran; 5) Molecular Medicine Department, Biotechnology Research Center, Pasteur Institute of Iran, Tehran, Iran; 6) Department of Medical and Molecular Genetics, St. John's Institute of Dermatology, King's College London (Guy's Campus), London, UK.

Epidermolysis bullosa (EB), a group of heritable skin fragility disorders, manifests with blistering and erosions leading to ulcers with poor healing tendency. As many as 20 different genes have so far been associated with different forms of EB. Tissue separation in the dystrophic EB (DEB) which is characteristically caused by mutations in *COL7A1* is below the lamina densa within the upper papillary dermis at the level of anchoring fibrils. Anchoring fibrils, which consist of type VII collagen, can be reduced in number, morphologically altered or entirely absent in DEB as a result of mutations in *COL7A1*. In this study, 363 distinct consanguineous EB families were examined by whole exome sequencing and genome-wide homozygosity mapping. A pathogenic homozygous mutation in the *PLOD3* was detected in a patient with clinical impression of mild DEB and multiple joint contractures. Immunofluorescence microscopy and Western blot showed complete absence of *PLOD3* in the patient. Interestingly, skin immunostaining with different antibodies against proteins of basement membrane zone showed markedly attenuated amounts of type VII collagen. Residual VII collagen was present both in the roof and the floor of the blister. In addition, the level of cleavage was shown to be below the lamina densa by transmission electron microscopy. Furthermore, only a few, thin wisp-like structures with no clearly discernable anchoring fibrils were present. Also, several fibroblasts were present close to the dermal-epidermal junction embedded within densely packed individual collagen bundles, and these cells displayed dilated endoplasmic reticulum and prominent vesicles close to the cell membrane. The latter findings are consistent with fibrotic reaction, a feature of DEB. In summary, our study demonstrated the consequences of homozygous *PLOD3* mutation in a patient with skin fragility with DEB-like phenotype with sub-lamina densa blistering, and perturbation of anchoring fibrils with deficiency in type VII collagen. In addition, there was evidence of fibrosis in the upper papillary dermis, features of DEB. Collectively, these findings suggest a novel gene harboring mutations in the mild dystrophic forms of EB, and suggest an additional mutant gene associated with skin fragility disorders in the spectrum of EB.

1049F

Reduced elastin deposition in a *PI4K2A*-related cutis laxa patient. E.C. Lawrence¹, E. Morava^{2,3}, F. Liu¹, M. Mohamed⁴, M.G. Waugh⁵, R.A. Wevers¹, Z. Urban¹. 1) Human Genetics, University of Pittsburgh, Pittsburgh, PA; 2) Haywards Genetics Center, Tulane University, New Orleans, LA; 3) Department of Pediatrics, University Medical Centre Leuven, Belgium; 4) Translational Metabolic Laboratory, Department of Laboratory Medicine, Radboud University Nijmegen Medical Center, the Netherlands; 5) Lipid and Membrane Biology Group, University College London, UK.

A recent whole exome sequencing study identified a recessive p.R275W mutation in the phosphatidylinositol 4-kinase type 2 α (*PI4K2A*) gene in an individual with cutis laxa, choreoathetoid movement disorder, dysmorphic features and intellectual disability. This mutation reduced *PI4K2A* activity to 22.5% of normal. The goal of this study is to identify the molecular mechanisms that may connect this gene mutation to the disease phenotype by examining elastic fiber biosynthesis and signalling pathways influenced by phosphatidylinositides. We Sanger sequenced an additional 115 probands with Cutis laxa of unknown genetic origin on the 9 coding exons of *PI4K2A*. Two additional probands had amino acid mutations, one mutation also likely to decrease enzyme activity (p.R328Q) and another (p.L262F) found in a water molecule-mediated hydrogen bond. The original patient and 3 control skin fibroblasts lines were grown to harvest RNA and protein for assays. By immunostaining, *PI4K2A* was localized to EEA1-positive endosomal compartments and was absent in the Golgi apparatus and lysosomes in both patient and control cells. The staining with the EEA1 and *PI4K2A* also showed colocalization in both patient and control samples showing that the mutation did not alter the localization of the *PI4K2A*. Quantitative RT-PCR detected increased expression of *ELN*, decreased expression of *FN1* and *FBLN4* and normal expression of *FBLN5*, *LOX* and *LOXL1*. Immunostaining showed largely normal levels and distribution of elastin, but diminished fibrillin-1 microfibrils. Colorimetric FASTIN assays demonstrated significant reduction of insoluble elastin in patient cells. In response to WNT3A, none of the cells showed a significant stabilization of beta catenin. Whereas normal cells responded by a robust increase in AKT phosphorylation to WNT3A treatment, the mutant cells did not. AKT phosphorylation, however, was increased the same way in patient and control cells in response to insulin. The p.R275W mutation does not interfere with the correct localization of *PI4K2A*, and likely acts through a change in enzyme activity. Impaired elastin deposition in mutant cells is associated with decreased fibrillin microfibril density. It remains to be determined whether the apparent loss of responsiveness to WNT3A in mutant cells contributes to the observed alterations in elastic fiber production and we are working on the additional amino acid changes to determine how they are involved in the disease phenotype.

1050W

A pathogenic deletion of the *VPS33B* gene leading to a case of attenuated arthrogryposis-renal dysfunction-cholestasis (ARC) syndrome. S.S. Rabin-Havt¹, K. Gallagher², P. Levy², S. Klugman¹, D.H. Lee², R. Marion². 1) Dept of Obstetrics & Gynecology and Women's Health, Montefiore Medical Center/ Albert Einstein College of Medicine, Bronx, NY; 2) Department of Pediatrics, Montefiore Medical Center/ Albert Einstein College of Medicine; 3) Department of Medicine, Montefiore Medical Center/ Albert Einstein College of Medicine.

Introduction: The increasing clinical use of exome sequencing (ES) is providing a window into the phenotypic spectrum of known syndromes. The *VPS33B* gene plays a crucial role in intracellular trafficking (Zhou & Zhang, 2014). Variants can disrupt epidermal barrier formation (Gruber et al., 2017) and have been implicated in arthrogryposis-renal dysfunction-cholestasis syndrome (MIM 208085) and, rarely, autosomal recessive keratoderma-ichthyosis-deafness syndrome (ARKID). ARC syndrome, an autosomal recessive disorder typically fatal in infancy, presents with arthrogryposis, renal tubular acidosis, and cholestatic jaundice often accompanied by ichthyosis, platelet dysfunction, brain and cardiac anomalies, and deafness (Zhou & Zhang, 2014). Rare cases of attenuated ARC syndrome, presenting later and with milder symptoms who have survived into childhood have been described (Smith et al., 2012). A 2017 paper reported on 3 patients with ARKID who presented with palmoplantar keratoderma (PPK), ichthyosis, sensorineural deafness, and variable associated symptoms (Gruber et al., 2017). **Case:** We report on an 11-year-old female who was referred to Dermatology-Genetics clinic. She presented with pruritic ichthyosis (since 6 months of age), bilateral hearing loss (since 2 years of age), severe intellectual disability, and failure to thrive. Pregnancy, birth, and family history were unremarkable and there was no reported consanguinity. Exam revealed proportional growth delay, microcephaly, rocker bottom feet, severe PPK, and ichthyosis. **Results:** Chromosomal microarray and connexin deafness panel were negative. ES revealed two variants of the *VPS33B* gene: the first, a partial gene deletion that includes the starting codon (a new pathogenic variant), the second, c.1225+5G>C, a previously described likely pathogenic variant predicted to cause abnormal gene splicing (previously reported in patients with attenuated ARC syndrome). After the diagnosis of *VPS33B*-related disorder was made, she was found to have elevated aminotransferase and renal tubular acidosis. **Conclusion:** This rare presentation of attenuated ARC syndrome with phenotypic overlap with ARKID syndrome suggests that these conditions represent a phenotypic spectrum. Diagnosis by ES led to presymptomatic identification of hepatic and renal involvement, and ES will continue to elucidate natural history and improve management. We propose renaming this clinical spectrum *VPS33B*-Related Disorder.

1051T

Probing a novel familial giant lipomatosis syndrome with genetics. J. de Vegvar¹, V. Aggarwal¹, D. Goldstein¹, K. Anyane-Yeboah¹, H. Depaz², N. Jouayed², F. Figurasin², R. Engha². 1) Columbia University, New York, NY; 2) Harlem Hospital, New York, NY.

Abstract: Introduction: Lipomas are slow-growing adipose tumors that occur more frequently in mature adults aged 40 to 60 years old. Most lipomas are small and weigh only several grams. A giant lipoma is defined as at least 10 cm in one dimension and weighs at least 100 grams. Cases of giant lipomas have been reported, but a clinical syndrome of a solitary giant encapsulated lipoma with a strong familial component has not yet been described.

Goals and hypothesis: We studied a 42-year-old man born in St. Croix with macrocephaly, intellectual disability and a giant lipoma. This same condition has been traced through three generations of the patient's family including his grandmother, two aunts, uncle and cousin. We hypothesized that this condition is caused by an unknown hereditary factor. **Methods:** We employed whole exome sequencing of the proband to determine the disease causing variant.

Results: A number of rare variants with significance were found. **Conclusion:** This case shows that solitary giant lipomas can present with a genetic family history. In this case, the pattern of inheritance fits an autosomal dominant model because it is three generations, and affects males and females equally. The presence of macrocephaly and intellectual disability in the affected individuals suggests that this is likely to be syndromic rather than an isolated condition. The results also show that there are many genetic conditions with unknown molecular basis, even with exome testing. Whole genome sequencing is now being used to identify the disease causing variant.

1052F

Exome sequencing identifies semidominant inheritance of a new *GNMB* mutation in an extended pedigree with amyloidosis cutis dyschromica.

A. Onoufriadis¹, A. Nanda², W. Scott³, C. Tierney³, C-K. Hsu^{3,4}, G.E. Orchard⁵, A.L. Guy⁶, L. Liu⁶, K. Tomita¹, A. Sherif¹, J.Y.W. Lee¹, M.A. Simpson⁷, E. Calonje⁶, J.A. McGrath¹. 1) St John's Institute of Dermatology, School of Basic and Medical Biosciences, King's College London, London SE1 9RT, United Kingdom; 2) As'ad Al-Hamad Dermatology Center, Al-Sabah Hospital, Kuwait City 13001, Kuwait; 3) Department of Dermatology, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan 70101, Taiwan; 4) International Research Center of Wound Repair and Regeneration (iWRR) National Cheng Kung University, Tainan 70101, Taiwan; 5) Dermatopathology Laboratory, Viapath, St Thomas' Hospital, London SE1 7EH, United Kingdom; 6) National diagnostic EB laboratory, Laboratory, Viapath, St Thomas' Hospital, London SE1 7EH, United Kingdom; 7) Department of Genetics, School of Basic and Medical Biosciences, King's College London, Guy's Hospital, London SE1 9RT, United Kingdom.

Amyloidosis cutis dyschromica (ACD) is an autosomal recessive disorder characterized by prepubertal onset of reticular hyperpigmentation with hypopigmented spots, and amyloid deposition in the upper dermis. Recently, bi-allelic mutations in *GNMB* (encoding a glycoprotein involved in melanosome maturation) have been described. We undertook whole-exome sequencing in three affected individuals with ACD from an extended consanguineous Bedouin pedigree with nine affected members. Focusing on homozygous protein altering variants with a frequency <0.01, we identified a single previously unreported splicing variant (c.700+5G>T) in *GNMB*. Co-segregation analysis showed homozygosity for this variant in all affected individuals but also revealed two individuals (the mother and one offspring) who were heterozygous and who had a milder/intermediate skin phenotype (less hyperpigmentation, fewer areas of hypopigmentation). We confirmed the splicing defect by reversion-transcription, with loss of the canonical donor splice site and use of a cryptic one, and introduction of a new premature stop codon. Moreover, qPCR in RNA from whole skin biopsies in three individuals with different genotypes of the *GNMB* mutation, showed approximately 35% reduction in *GNMB* expression in a heterozygous affected member, and almost no expression in a homozygous individual when compared to a wild-type unaffected family member. Immunofluorescence analysis showed markedly reduced *GNMB* in homozygote skin, patchy reduction in melanocyte numbers and basal keratinocyte melanin, as well as positive amyloid labelling in the papillary dermis. Electron microscopy in homozygote skin revealed numerous dermal colloid bodies (dead basal keratinocytes). In contrast, heterozygote skin showed a slight reduction in the number of melanocytes and reduced melanin but no amyloid deposition or colloid bodies. In summary, these data expand the spectrum of mode of inheritance in ACD, highlighting semidominant inheritance rather than just autosomal recessive transmission, and identify a new mutation in *GNMB*.

1053W

A novel KIT mutation in a family with expanded syndrome of Piebaldism.

S. Majid¹, I. Hamadah², M. Chisti², M. Haider², H. Dosssari¹, R. Humaidan¹, B. Meyer¹. 1) King Faisal Specialist Hospital & Research Center, Riyadh, Saudi Arabia; 2) Department of Dermatology, King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia.

Mutations in the KIT gene have been associated rare autosomal dominant genodermatosis characterized by depigmentation of skin and white forelocks. KIT gene which encodes the cell surface transmembrane tyrosine kinase receptor for KIT ligand represents critical factor for melanoblast migration, proliferation, differentiation, survival and a decrease in melanogenesis. We screened a family of four affected's consistent with autosomal dominant pattern of inheritance using Next Generation Sequencing in a comprehensive gene panel incorporating 194 dermatological genes. While our patients presented with classical phenotype of piebaldism, additional features in the form of sensorineural deafness and heterochromia iridis were observed in some patients. Our results from next generation sequencing derm panel suggests novel missense mutation in the tyrosine kinase domain of the KIT gene, responsible for the severe form of piebaldism in our patients. Our findings expand the mutation spectrum of KIT-associated piebaldism disorders and highlight the clinical utility of next generation sequencing derm panel for the diagnosis in patients with suspected genodermatosis.

1054T

Analysis options for human whole genome sequencing. T. Stokowy^{1,2,3}, A. Supernat⁴, O.V. Vidarsson⁵, R. Holdhus¹, T. Fiskerstrand⁶, G. Houge⁶, V.M. Steen^{1,2,3}. 1) Department of Clinical Science, University of Bergen, Bergen, Hordaland, Norway; 2) NORMENT & K.J. Jebsen Centre for Psychosis Research, Department of Clinical Science, University of Bergen, Norway; 3) Dr. E. Martens Research Group for Biological Psychiatry, Department of Medical Genetics, Haukeland University Hospital, Bergen, Norway; 4) Laboratory of Cell Biology, Intercollegiate Faculty of Biotechnology, University of Gdańsk and Medical University of Gdańsk, Poland; 5) Computational Biology Unit, Institute of Informatics, University of Bergen, Norway; 6) Department of Medical Genetics, Haukeland University Hospital, Bergen, Norway.

Testing of patients with hereditary disorders is in progress of shifting from single gene assays to gene panel sequencing, whole-exome sequencing (WES) and whole-genome sequencing (WGS). Therefore, we decided to summarize analysis options for human whole genome sequencing from a clinical perspective. As the first step, we performed a comprehensive literature search and review of 512 articles related to WGS applications. Subsequently, we tested DeepVariant, a new TensorFlow machine learning-based variant caller, and compared to GATK 4.0 and SpeedSeq, using 30X WGS data of the well-known NA12878 DNA reference sample. According to our comparison, DeepVariant reached the highest F-Score (i.e. harmonic mean of recall and precision) both for SNVs and indels (0.98 and 0.94 respectively). Performance of SNV calling using the three mentioned tools was comparable (GATK and SpeedSeq reached lower F-Scores, however still equal to 0.98 after rounding to two decimal places). On the other hand, DeepVariant was more precise in indel calling (GATK and SpeedSeq reached F-Scores 0.90 and 0.84, respectively). We concluded that the DeepVariant tool has great potential and usefulness for medical genetics. We used methods mentioned above to disclose causes of three monogenic disorders: keratolytic winter erythema (CTSB, duplication of enhancer), severe Penttinen syndrome (PDGFRB, de novo substitution) and immunodeficiency syndrome (IL, deletion). The functional data which support our findings (RNA-Seq, epigenetics and successful drug testing in patient-derived cell lines) will be presented during the meeting.

1055F

Novel *Dsg1* mutation in focal palmoplantar keratoderma. T. Nomura¹, M. Takeda¹, T. Miyauchi¹, S. Suzuki¹, J.T. Peh¹, T. Uesugi², H. Shimizu¹. 1) Department of Dermatology, Hokkaido University Graduate School of Medicine, Sapporo, Japan; 2) Uesugi Dermatology Clinic, Sapporo, Japan.

Inherited palmoplantar keratodermas (PPKs) are clinically and genetically heterogeneous genodermatoses, characterized by hyperkeratosis of the palms and soles. Based on the distribution pattern of hyperkeratotic lesions, PPKs are clinically classified into four types; diffuse, striate, focal, and punctate. To date, more than 20 genes have been reported to be responsible for PPKs. Although the majority of heterozygous mutations in the gene encoding desmoglein 1 (*DSG1*), a key molecule for epidermal adhesion and differentiation, cause striate PPK, mutations causing haploinsufficiency of *DSG1* have also been identified in one and three families with diffuse and focal PPKs, respectively. Here we report a Japanese family with focal PPK caused by a *DSG1* mutation. 10-year-old dizygotic twins presented with focal skin thickening of the palms and soles that had appeared since their birth. Palmoplantar skin of their father and paternal grandfather was similarly affected. None of the affected individuals showed a striate pattern of hyperkeratosis on their palmoplantar regions. Histology of the affected plantar skin showed marked hyperkeratosis, parakeratosis, highly eosinophilic cytoplasm of keratinocytes, and widen intercellular spaces and disadhesion of keratinocytes. Sanger sequencing revealed that the patients are heterozygous for a previously unreported nonsense mutation in *DSG1*, whereas their unaffected mother is wild-type for the mutation. These findings collectively led us to diagnose the patients as having focal PPK due to a *DSG1* mutation. In conclusion, this study further highlights the clinical heterogeneity of *DSG1*-associated PPKs. Although mutation analysis of *DSG1* has been performed primarily for striate PPK, our results suggest that *DSG1* screening should also be considered routinely in focal PPK.

1056W

Genetic heterogeneity in Osteogenesis Imperfecta, NGS based study.

K. Sałacińska¹, L. Rutkowska¹, I. Pinkier¹, D. Salachna¹, A. Rusinska², E. Jakubowska-Pietkiewicz³, A. Jamsheer³, M. Rybak-Krzyszowska⁴, L. Jakubowski¹, A. Gach¹. 1) Department of Genetics, Polish Mother's Memorial Hospital Research Institute, Lodz, Poland; 2) Department of Propaedeutics of Paediatrics and Metabolic Diseases, University Teaching Hospital, Lodz, Poland; 3) Department of Medical Genetics, University of Medical Science, Poznan, Poland; 4) Department of Obstetrics and Perinatology, Jagiellonian University Medical College, Cracow, Poland.

IntroductionOsteogenesis Imperfecta (MIM 166200) is a rare genetic disorder of connective tissue characterized by numerous bone fractures, gray or blue sclera, tooth abnormalities, short stature and skeletal deformity. The diversity of clinical features and its severity varies between patients, even within a single family, ranging from severe perinatal lethal to a mild form. It is estimated that about 1500 different mutations in only 2 genes linked to proper protein structure and amount of collagen type I are responsible for about 90% of OI cases. Remaining 10% is caused by mutations of numerous genes directly influencing collagen type I protein biosynthesis or related molecular pathways. The aim of this project was to determine genetic heterogeneity versus clinical variability observed in patients affected by OI. **Material and methods**Next Generation Sequencing technique using custom panel dedicated for re-sequencing of 34 genes with confirmed and probable significance in pathogenesis of OI was performed. The study included 102 patients, aged 1-44 years, presenting a broad spectrum of clinical manifestation. Twenty-five of them were familial cases. Based on bioinformatic analysis, selected variants were verified with Sanger sequencing in a group of 69 patients. **Results**We have identified *COL1A1* (MIM 120150) mutations in 40 patients, *COL1A2* (MIM 120160) in 23 patients and mutations in 3 non-collagenous genes in 6 patients. In one patient pathogenic variants in both collagen type 1 genes were identified. A total of 36 different mutations in *COL1A1* and 20 in *COL1A2* were found. Respectively 14 and 11 have not been reported in dedicated Osteogenesis Imperfecta Variant Database. A novel identified variant in *COL1A2* occurred in two family members and two nonconsanguineous patients. In a 25-year old patient with moderate form of OI we identified novel compound heterozygous mutations in a *BMP1* gene (MIM 112264). Both *BMP1* mutations affect highly conserved residues across distant species localized in CUB1 domain. **Discussion**Obtained results extend variant spectrum of *COL1A1* and *COL1A2* mutations. Phenotype variability between patients within the same type of OI or possessing the same mutation is observed both inter- and intrafamilial. Thus far, mutations in *BMP1* gene have been identified within metalloprotease, CUB2 and CUB3 domains, signal peptide, 3'UTR sequence but not in CUB1 domain. Our study broadens insights into genotype-phenotype correlation in OI.

1057T

MALTA (MYH9 Associated eLasTin Aggregation) syndrome: Germline variants in MYH9 cause rare sweat duct proliferations and irregular elastin aggregations.

E. Fewings¹, M. Ziemer², K. Hörtnage³, K. Reichert³, A. Larionov⁴, J. Redman¹, M.A. Goldgraben¹, H. Firth⁴, T. Ha⁵, J. Schaller⁶, D. Adams⁷, E. Rytina⁸, M. van SteENSEL^{9,10}, M. Tischkowitz¹. 1) University of Cambridge, Cambridge, United Kingdom; 2) Department of Dermatology, Venerology and Allergology, University Hospital, Leipzig, Germany; 3) Center for Genomics and Transcriptomics and Praxis für Humangenetik Tübingen, Tübingen, Germany; 4) East Anglian Medical Genetics Service, Cambridge University Hospitals National Health Service Foundation Trust, Cambridge Biomedical Campus, Hills Road, Cambridge, Cambridgeshire, CB2 0QQ, UK; 5) Department of Dermatology, Addenbrooke's Hospital, Cambridge, Cambridgeshire, CB2 0QQ, UK; 6) Dermatopathologie Duisburg, Duisburg, Germany; 7) Experimental Cancer Genetics, Wellcome Trust Sanger Institute, Hinxton, Cambridge, Cambridgeshire, CB10 1SA, UK; 8) Department of Histopathology, Addenbrooke's Hospital, Cambridge, Cambridgeshire, CB2 0QQ, UK; 9) Skin Research Institute of Singapore, Singapore; 10) Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore.

Background: In the present study we investigate several patients with an irregular distribution of elastin fibres and benign sweat duct proliferations who were initially diagnosed with Rombo or Nicolau-Balus syndrome. Analysis of the germline mutations within these individuals provides evidence that these disorders represent a single entity, characterised by variants in non-muscle myosin gene *MYH9* [MIM: 160775]. Pathogenic variants in *MYH9* have been previously described in patients with thrombocytopaenia and giant platelets. **Methods:** DNA was extracted from blood of six affected individuals from three families with a history of sweat duct proliferations. Whole exome sequencing was performed across these individuals to identify rare, protein-affecting variants that co-segregate in affected individuals. The *MYH9* gene, which contained variants in these three families, was Sanger sequenced in a further two affected families. The functional impact of identified variants on known binding sites was analysed *in silico* using 3D protein modelling and the conservation of affected amino acids was assessed across different myosin classes. **Results:** Predicted protein-affecting variants in *MYH9* were identified in all tested sweat duct proliferation syndrome families, co-segregating with the phenotype. None of the affected individuals presented with any phenotypic features of the previously described *MYH9*-related platelet disorder (MRPD). Four out of the five identified variants were located within the myosin head ATP binding pocket. The amino acids modified by these variants are conserved across eight different classes of myosin. **Discussion:** We provide evidence that sweat duct proliferations with abnormal elastin aggregations are associated with germline pathogenic variants in *MYH9*. We therefore suggest the name MALTA (MYH9 Associated eLasTin Aggregation) syndrome to reflect this underlying genetic basis. The identified *MYH9* variants are located in different gene regions and appear to have a different effect on myosin function to those previously associated with *MYH9*-related platelet disorder (MRPD). Understanding the genetic basis of MALTA syndrome will aid the diagnosis and management of this disease by providing a way of differentiating its associated skin lesions from histopathologically similar but clinically aggressive sweat duct tumours. Importantly, our findings point to a novel disease mechanism for non-muscle myosins in tumorigenesis and skin diseases.

1058F

Characterization of vascular phenotypic variability in a non-isogenic mouse model for Marfan syndrome. I.G. Gyuricza¹, R.B. Souza¹, G.R. Fernandes¹, L.E. Farinha-Arcieri¹, I.H.J. Koh², L.V. Pereira¹. 1) Department of Genetics and Evolutionary Biology, University of São Paulo, São Paulo, Brazil; 2) Department of Surgery, Paulista School of Medicine, Federal University of São Paulo, São Paulo, Brazil.

Marfan syndrome (MFS) is autosomal dominant connective tissue disorder caused by mutations in *FBN1* gene, which encodes the extracellular matrix protein Fibrillin-1. The main manifestations of MFS include the skeletal and cardiovascular tissues, the latter being the leading cause of mortality. Despite its complete penetrance, there is a wide inter and intra familial clinical variability among patients, which makes the prognosis more difficult and suggest the existence of modifier genes. Based on that, our group created a new MFS mice strain by crossing C57BL/6 mice heterozygous for *Fbn1* mutation ($mg\Delta^{loxprneo}$) and 129/Sv wild type (WT) mice, generating a model with higher genetic and phenotypic variability. The main objective of this study was to characterize the vascular phenotypic variability in the mixed B6/129 $mg\Delta^{loxprneo}$ model. We analyzed histological aorta sections from $mg\Delta^{loxprneo}$ and WT mice, and measured a set of features such as: medial and adventitia aortic wall thickness, number of vascular smooth muscle cells (VSMC) and elastic fibers integrity. Using Principal Component Analysis (PCA), we could observe that $mg\Delta^{loxprneo}$ and WT mice presented wide phenotypic variability, and the main aortic feature that distinguishes MFS from WT mice is the elastic fibers integrity. Despite the fact that all the $mg\Delta^{loxprneo}$ mice showed reduced elastic fibers integrity, those with thicker medial and adventitia aortic wall, as well as higher number of VSMCs, did not develop aneurysms. These results suggest that the adaptive events of aortic wall might be important for the maintenance of vascular function and that genes involved in these pathways should be further analyzed as potential protective genes. In conclusion, the B6/129 $mg\Delta^{loxprneo}$ strain models the clinical variability of MFS and can be useful for identifying the molecular and physiological mechanisms underlying cardiovascular diseases. In addition, the identification of these mechanisms can improve the patient's prognosis and support the development of new therapeutic targets.

1059W

A somatic activating NRAS mutation associated with kaposiform lymphangiomatosis. S.F. Barclay¹, K.W. Inman², V.L. Luks², J.B. McIntyre³, A. Al-Ibraheemi^{2,7}, A.R. Perez-Atayde², S.J. Fishman^{4,7}, A.I. Alomari^{5,7}, G. Chaudry^{6,7}, C.C. Trenor, III^{6,7}, D. Adams^{6,7}, H.P.W. Kozakewich^{2,7}, K.C. Kurek¹. 1) Departments of Pathology & Laboratory Medicine and Medical Genetics, Alberta Children's Hospital Research Institute and Cumming School of Medicine, University of Calgary, Calgary, AB; 2) Department of Pathology, Boston Children's Hospital, Boston, MA; 3) Translational Laboratory, Tom Baker Cancer Centre, Department of Oncology, Cumming School of Medicine, University of Calgary, Calgary, AB; 4) Department of Surgery, Boston Children's Hospital, Boston, MA; 5) Division of Interventional Radiology, Boston Children's Hospital, Boston, MA; 6) Department of Hematology/Oncology, Boston Children's Hospital, Boston, MA; 7) Vascular Anomalies Center, Boston Children's Hospital, Boston, MA.

Kaposiform lymphangiomatosis (KLA) is a rare, aggressive, and systemic disorder of lymphatic proliferation, occurring primarily in children with rare reports in adults. Even with multimodal treatments, KLA has a poor prognosis and high mortality rate secondary to coagulopathy, effusions and systemic involvement. We hypothesized that, as has recently been found for other vascular anomalies, KLA may be caused by somatic mosaic mutations during vascular development. To explore this hypothesis, we performed exome sequencing of tumour samples from five individuals with KLA, along with samples from uninvolved control tissue in three of the five. We identified a somatic activating *NRAS* (c.182A>G, p.Q61R) mutation in four of the five KLA tumour samples, at levels ranging from 3-14%, that was absent from the control tissues. We subsequently used digital PCR (dPCR) to validate the exome findings and to screen four additional KLA samples. In total we identified the *NRAS* p.Q61R mutation in lesional tissue from 8/9 KLA individuals. The *NRAS* p.Q61R mutation is a known 'hotspot' mutation, frequently identified in several types of human cancer, especially melanoma. It results in a constitutively active *NRAS* protein, leading to unchecked cellular growth and proliferation. KLA therefore joins a growing group of vascular malformations and tumours caused by somatic activating mutations in the RAS/PI3K/mTOR signalling pathways. This discovery will expand treatment options for these high risk patients as there is potential for use of targeted RAS pathway inhibitors.

1060T

Novel ABCA12 mutations in Japanese patients with harlequin ichthyosis.

M. Takeda¹, T. Nomura¹, S. Suzuki¹, T. Miyauchi¹, Y. Fujita¹, N. Tabata², M. Akiyama³, H. Shimizu¹. 1) Department of Dermatology, Hokkaido University Graduate School of Medicine, Sapporo, Hokkaido, Japan; 2) Department of Dermatology, Japanese Red Cross Sendai Hospital, Sendai, Miyagi, Japan; 3) Department of Dermatology, Nagoya University Graduate School of Medicine, Nagoya, Aichi, Japan.

Harlequin ichthyosis (HI) is the most severe form of autosomal recessive congenital ichthyosis and has a high mortality rate. HI is clinically characterized by thick plate-like scales with deep fissures, severe ectropion, eclabium, rudimentary ears, excessive loss of body fluids, hypohidrosis, hyper/hypothermia, respiratory distress, infection, and growth retardation. Loss-of-function mutations in the gene encoding adenosine triphosphate-binding cassette subfamily A member 12 (ABCA12), a lipid transporter that is crucial for epidermal barrier formation, have been identified as the genetic cause of the disease. In this study, we genetically analyzed two HI siblings from a Japanese family. Two girls of the ages of 2 years and 1 year who presented with generalized erythroderma and marked scales were referred to us for genetic testing. Although they showed typical HI phenotypes at their birth, oral etretinate had alleviated the severity of their skin symptoms. Whole-exome and Sanger sequencing revealed them to be compound heterozygous for previously unreported two mutations (c.2122-1G>A and c.4663_4677del) in ABCA12, whereas the father and mother are heterozygous for the former and the latter mutation, respectively. cDNA sequencing analysis demonstrated that the splice-site mutation c.2122-1G>A yields a single aberrant splice variant, r.2122_2332del (p.Ala708Hisfs*19). The deletion mutation c.4663_4677del results in the absence of 5 amino acids within the first ATP-binding cassette (p.Leu1555_Gly1559del). The deleted amino acids are conserved among species, suggesting the importance of these amino acid residues in ABCA12 lipid transporter activity. In summary, this study further expands our understanding of genotype-phenotype correlations in HI and highlights the importance of the early introduction of oral retinoids in the management of HI.

1061F

Intranasal desmopressin treatment for massive subcutaneous hematoma in five patients with musculocontractural Ehlers-Danlos syndrome.

N. Fujita^{1,3}, M. Ishikawa¹, E. Kise¹, T. Yamaguchi^{1,2}, T. Yoshinaga¹, T. Takano^{1,2}, T. Kosho^{1,2}. 1) Center for Medical Genetics, Center for Medical Genetics, Matsumoto, Japan; 2) Department of Medical Genetics, Shinshu University School of Medicine, Matsumoto, Japan; 3) Problem-Solving Oriented Training Program for Advanced Medical Personnel: NGSD (Next Generation Super Doctor) Project, Matsumoto, Japan.

Musculocontractural Ehlers-Danlos syndrome (mcEDS) is a rare subtype of Ehlers-Danlos syndrome (EDS) caused by impaired synthesis of dermatan sulfate. Massive subcutaneous hematoma is one of the most serious complications in mcEDS, occurring with mild traumas, spreading with severe pain, and sometimes resulting in hemorrhagic shock. Intranasal desmopressin treatment is used for patients with mild hemophilia and von Willebrand disease through increased activities in factor VIII and von Willebrand factor. According to several reports describing the usefulness of intravenous or intranasal desmopressin in patients with various types of EDS (Stine and Becton, J Pediatr Hematol/Oncol 19: 156–158, 1997; Mast et al., Br J Haematol 144: 230–233, 2008), we tried the treatment for a patient with mcEDS suffering from recurrent massive subcutaneous hematoma (Kosho et al., Am J Med Genet A 138A: 282–287, 2005; Kosho et al., Am J Med Genet A, 152A: 1333–1346, 2010). In this report, we described safety and efficacy of intranasal desmopressin treatment (150µg/episode) for a total of five patients (1 male, 4 females; the mean age 18.2 years) receiving the treatment to date in Shinshu University Hospital. Clinical information was retrospectively reviewed from medical records of these patients. The mean age of the first episode of massive subcutaneous hematoma was 8 years (range, 4 to 15). Most episodes were associated with obvious triggers (falling down, fighting). Hematoma developed at the legs in all and at the heads in four. Surgical drainage was required in three and blood transfusion in two. Elevated activities of factor VIII and von Willebrand factor were measured in all after administration of intranasal desmopressin. Four continue the treatment without adverse effects, whereas one child discontinued it because of numbness in the arms. Except for extraordinary situations such as a traffic accident, no patients developed severer episodes requiring surgical drainage or transfusion under the treatment. Most realized that the treatment suppressed growth of hematoma and shortened the suffering time. In view of these findings, intranasal desmopressin treatment on patients with mcEDS would be safe at least in adolescent or adult patients and effective in preventing massive subcutaneous hematoma.

1062W

Characterization of unusual DMD cases. *F. Barthelemy*^{1,3}, *R. T. Wang*^{2,3}, *E. Douine*^{2,3}, *A. Eskin*^{2,3}, *H. Lee*^{2,4}, *V. Arboleda*⁴, *N. Khanlou*⁴, *P. B. Shieh*^{2,5}, *M. C. Micelli*^{1,3}, *S. F. Nelson*^{2,3,4}. 1) MIMG, UCLA, 90095, CA; 2) Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA; 3) Center for Duchenne Muscular Dystrophy, University of California Los Angeles, Los Angeles, CA, USA; 4) Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA; 5) Department of Neurology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA.

Duchenne muscular dystrophy (DMD) is a relatively common X linked recessive progressive muscle dystrophy affecting approximately 1 in 5000 male births. Most mutations are new in the family due to a high rate of de novo mutations in DMD. Typically, diagnosis is established at age 3-5, and with progressive loss of muscle function, loss of ambulation occurs around age 10-13 years. There are a wide range of mutations in the DMD locus, and a portion of disease variability is accounted for by variations in dystrophin expression, and is linked to specific mutation type. Thus, Comprehensive DNA diagnostics are standard for providing prognostic information. Further, novel genetic treatments require specific knowledge of mutation type. However, even with intense molecular diagnostics, there remain patients without a clear DNA mutation. Here we explore the use of RNA_SEQ from muscle tissue to clarify mutational consequence for rare mutations in DMD and to identify otherwise cryptic intronic DNA mutations. Further, we explore unusual DMD patients with out of frame mutations and a high rate of revertant fibers where dystrophin is demonstrated to be restored in the muscle membrane, and explore the molecular mechanisms responsible. We argue that muscle biopsy should remain a commonly performed test in Duchenne/Becker muscular dystrophy.

1063T

EPG5 variants with modest functional impact result in an ameliorated and primarily neurological phenotype in a 3-year old patient with Vici syndrome. *M.S. Kane*¹, *J. Zhao*¹, *J.A. Muskett*¹, *A. Diplock*², *S. Srivastava*², *N. Hauser*¹, *J.F. Deeken*^{1,3}, *J.E. Niederhuber*^{1,4}, *W.E. Smith*⁵, *D. Ebrahimi-Fakhari*², *T. Vilboux*¹. 1) Inova Translational Medicine Institute, Inova Health System, Falls Church, VA; 2) Department of Neurology, Boston Children's Hospital, Harvard Medical School, Boston, MA; 3) Department of Medicine, Virginia Commonwealth University, Richmond, VA; 4) Department of Oncology and Surgery, Johns Hopkins University School of Medicine, Baltimore, MD; 5) Department of Pediatrics, The Barbara Bush Children's Hospital, Maine Medical Center, Portland, ME.

Congenital disorders of autophagy are an emerging class of multisystem disorders with significant neurological involvement. *EPG5*-associated Vici syndrome is a prototypical congenital disorder of autophagy and presents with the cardinal features of agenesis of the corpus callosum, bilateral cataracts, hypertrophic and/or dilated cardiomyopathy, combined immunodeficiency, and skin, hair, and retinal hypopigmentation in the majority of cases. The vast majority of *EPG5* variants leading to Vici syndrome are null alleles with only a few missense variants reported to date. We have previously reported a missense variant in *EPG5* that leads to aberrant splicing with a high percentage of the resulting transcripts encoding premature termination codons. Here we report on a 3-year old patient with a molecular diagnosis of Vici syndrome attributed to compound heterozygous *EPG5* variants [NM_020964.2:c.772G>T, p.Glu258* and NM_020964.2:c.5943-9_5943-5del]. The clinical presentation of this patient deviates notably from the classical Vici phenotype with a lack of hypopigmentation, cataracts, immunodeficiency, cardiomyopathy or failure to thrive. Neurological manifestations, however, are consistent within the known spectrum of Vici syndrome and are significant for early-onset global developmental delay and post-natal microcephaly with a thin corpus callosum. Based on the ameliorated and primarily neurological phenotype, we hypothesized that the functional impact of the *EPG5* variants present would be milder with a higher amount of residual *EPG5* expression. Analyses of *EPG5* mRNA in the patient and his parents were performed to examine expression level and splicing; mRNA from a healthy donor and a patient with classic Vici syndrome were also included. Despite the nonsense mutation, *EPG5* mRNA levels were modestly decreased in the patient and half the transcripts were properly spliced. In contrast, we observed a 50% reduction in mRNA expression in classic Vici syndrome and proper splicing in only 30% of the residual mRNA. These results support a model of disease severity which correlates to the dosage of wild-type *EPG5* expression.

1064F

Search for molecular biomarkers of spinal muscular atrophy. M.A. Maretina^{1,2}, N.A. Tsyanova², A.A. Egorova¹, T.E. Ivashchenko¹, V.S. Baranov^{1,2}, A.V. Kiselev¹. 1) D.O. Ott Research Institute of Obstetrics, Gynecology and Reproductology, Saint-Petersburg, Russian Federation; 2) Saint-Petersburg State University, Saint-Petersburg, Russian Federation.

Spinal muscular atrophy (SMA) is a severe autosomal recessive neuromuscular disorder caused by mutations within the *SMN1* gene. This gene has a nearly identical centromeric copy – the *SMN2* gene – that produces low amount of full-length *SMN* (*FL-SMN*) transcripts. The majority of transcripts produced by the *SMN2* gene lack exon 7 (*Δ7-SMN*) and encode non-functional protein. Correction of *SMN2* gene splicing is one of the most promising SMA therapeutic strategies. Development of new therapies for SMA requires reliable biomarker whose score significantly differs between patients, carriers and healthy individuals and a precise technique for detection of this biomarker. Several potential biomarkers of SMA have been examined in this work: a percentage of *FL-SMN* transcripts from total *SMN* transcripts; a ratio of *FL-SMN* to *Δ7-SMN* transcripts and a level of SMN protein relative to overall proteins level. Such methods as real-time quantitative RT-PCR, semiquantitative RT-PCR and quantitative fluorescent RT-PCR have been compared for most accurate determination of *SMN* transcripts level, while SMN protein level has been determined by means of ELISA. The analysis has been performed on blood cell samples obtained from SMA patients, SMA carriers and healthy individuals. Percentage of *FL-SMN* transcripts detected by means of semiquantitative RT-PCR and quantitative fluorescent RT-PCR has been shown to differ significantly between three examined groups. In order to further evaluate developed biomarkers SMA patient-derived fibroblasts have been transfected with several antisense oligonucleotides (ASO) able to correct *SMN2* exon 7 splicing. Significant increase in percentage of *FL-SMN* transcripts have been observed in SMA fibroblasts treated with ASO compared to intact cells. These data indicate that percentage of *FL-SMN* transcripts can serve as a suitable biomarker for SMA treatment efficacy determination. This work was supported by Russian Foundation for Basic Research grant 18-315-00258 mol_a. Marianna Maretina is supported by President of Russian Federation scholarship (SP-822.2018.4).

1065W

A rare variant in *ABCC1* is associated with altered APP processing in a familial case of late-onset Alzheimer's disease. W.M. Jepsen¹, M. De Both¹, A.L. Siniard¹, A. Henderson-Smith¹, K. Ramsey², R. Caselli³, G. Serrano⁴, T.G. Beach⁴, M.J. Huettelman¹. 1) Neurogenomics, Translational Genomics Research Institute, Phoenix, AZ; 2) Center for Rare Childhood Disorders, TGen, Phoenix, AZ; 3) Mayo Clinic, Phoenix, AZ; 4) Banner Sun Health Research Institute, Sun City, AZ.

The study of germline mutations in familial, early-onset Alzheimer's disease (AD) has been invaluable in unraveling the molecular pathways of AD, yet no effective therapy exists to halt or reverse the disease at any age of onset. Other than mutations in the *TREM2* locus, little is known about the genetics of heritable late-onset AD (LOAD). By identifying rare mutations in familial LOAD cases, it may be possible to identify novel, targetable pathways to alter disease progression. We analyzed a familial case of late-onset AD – in which 6 of 10 siblings were diagnosed with AD – and have identified a rare, germline mutation in *ABCC1* (chr16:16216007 A>G, p.Y1189C) found in 3 affected females, but not in 2 unaffected siblings. Genetic material for the remaining 5 siblings was unavailable. Previous studies have shown *Abcc1* double K/O significantly decreases Abeta clearance across the BBB in an AD mouse model, and that an amino acid change at our position of interest (p.Y1189) alters the substrate specificity of the human protein. Here, we show that *ABCC1*(WT) overexpression significantly decreases Abeta1-40 and 1-42 levels in BE(2)-m17 cells in culture, with a reduced effect in the Y1189C mutant line. We also show that our *ABCC1* cell lines are differentially affected by both thiethylperezine (TEP, an *ABCC1*-transport activator) and Reversan (an *ABCC1* inhibitor). 25uM TEP decreases extracellular levels of Abeta1-40 in both *ABCC1* cell lines, but Abeta1-42 only decreases in the WT line. 0.75uM Reversan resulted in cellular detachment of the *ABCC1*(WT) line with no effect on the empty vector or Y1189C line, suggesting that *ABCC1* may influence extracellular matrix proteolysis, and that this effect is dramatically reduced in the Y1189C mutant. RNA-seq of the cell lines revealed that mRNA levels of *TIMP3*, a membrane metalloprotease inhibitor, was significantly reduced in the *ABCC1*(WT) cell line versus the control line, with a smaller effect in the Y1189C line. The mutation also resulted in a higher ratio of extracellular Abeta to soluble APPalpha. This suggests that the mutant skews APP processing towards the beta-secretase pathway, or specifically hinders Abeta transport. Taken together, our data show that the Y1189C mutation likely increases the amyloid burden in our AD family, and to our knowledge, this is the first germline mutation in *ABCC1* to ever be associated with human disease, and only the second identification of a late-onset familial Alzheimer's disease gene.

1066T

Strikingly high frequency of p.R132X variant identified in *MMACHC* gene of Indian children presenting with late-onset cobalamin C defect. S.V. Attiri¹, R. Kaur¹, A.G. Saini¹, N. Sankhyani¹, M. Faruq², A. Patial¹, S. Singh¹. 1) Pediatrics, Postgraduate Institute of Medical Education & Research (PGIMER), Chandigarh, India; 2) Genomics and Molecular Medicine Unit, Institute of Genomics and Integrative Biology (IGIB), New Delhi, India.

Introduction: Cobalamin C (CblC) defect is the most common inborn error of cobalamin metabolism caused by defects in *MMACHC* gene leading to impaired conversion of dietary vitamin B₁₂ into methylcobalamin and adenosylcobalamin. Variants in *MMACHC* cause the accumulation of methylmalonic acid and homocysteine and decreased methionine synthesis. **Objective:** To identify disease-causing variants in *MMACHC* gene in children with Cbl defect. **Methods:** A total of 18 children diagnosed with Cbl defect on the basis of elevated plasma and urine homocysteine, urine methylmalonic acid, propionylcarnitine (C3) and low methionine were enrolled. Only the exonic regions of *MMACHC* gene were screened for the pathogenic variants by Sanger sequencing. **Results:** Out of 18 children with suspected cbl defect, 13 were found to have CblC (*MMACHC*) defect. Median (IQR) age at the enrolment was 11 years (6-18). Male to female ratio was 10:3. The spectrum of clinical features included gait disturbances (11/13), intellectual disability (10/13), global developmental delay (8/13), speech issues (8/13), behavior problems (7/13), acute encephalopathy (6/13), seizures (5/13) and megaloblastic anemia (4/13). Vitamin B₁₂ and folate levels were normal in all the cases. The median (IQR) plasma total homocysteine was 111 µmol/L (81-135), urine homocysteine 6.0 mmol/mol Cr (4.7-38.5), methionine 8.22 µmol/L (6.9-10) and urine MMA was 813 mmol/mol Cr (89-1640). The mean (range) C3 level was 4.7 µmol/L (1.0-9.1). The consanguinity was present in 3 families. Remarkably, sequencing of *MMACHC* revealed 12/13 (92%) cases with homozygous pathogenic p.R132X (c.394C>T) and one with homozygous pathogenic p.L116P (c.347T>C) variant. Parents and some of the siblings were found to have this variant in heterozygous state. Stable mRNA is expected to be produced from *MMACHC* harboring p.R132X variant since it is located within 50 nt of penultimate exon-exon junction. **Conclusion:** Our study presents the largest clinical series of children with late-onset CblC defect from India. Finding of common homozygous p.R132X variant in 92% children from unrelated families suggests founder effect. Association of p.R132X variant with late-onset disease suggests that stable mRNA is generated from *MMACHC* resulting in residual function protein. Such premature termination generating variants could be ideal target to prioritize readthrough therapies in individualized fashion to treat basic defect in Cbl defects and other Mendelian disorders.

1067F

Spectrum of *CBS* gene variants in Indian children with classical homocystinuria: A study from tertiary care hospital. R. Kaur¹, S.V. Attiri¹, A.G. Saini¹, N. Sankhyani¹, M. Faruq², V.L. Ramprasad³, S. Sharda³, A. Patial¹, S. Singh¹. 1) Department of Pediatrics, Postgraduate Institute of Medical Education and Research (PGIMER), Chandigarh, India; 2) Genomics and Molecular Medicine Unit, Institute of Genomics and Integrative Biology (IGIB), New Delhi, India; 3) MedGenome Labs Ltd., Bengaluru, Karnataka, India.

Introduction: Classical homocystinuria, characterized by elevated levels of homocysteine in blood and urine can be caused by pathogenic variants in *CBS* gene, the most common cause responsible for isolated homocystinuria. **Objectives:** Characterization of children with classical homocystinuria through clinical, biochemical and genetic assessment. **Methods:** A total of 17 children with urine homocysteine ≥4 mmol/mol Creatinine on two or more occasions, high or normal methionine, and normal methylmalonic acid with strong clinical symptoms of classical homocystinuria; were enrolled from Neurology and Genetic units of Advanced Pediatrics Centre, PGIMER, Chandigarh. Written informed consent was obtained in all the cases. Plasma total homocysteine was analyzed in all the cases. The coding regions of *CBS* gene were amplified through PCR and screened for variants through Sanger sequencing. **Results:** Out of 17 suspected cases, genetic testing for *CBS* gene confirmed the diagnosis in 14 cases. Median age at enrolment was 9 years (Interquartile range, IQR= 6-11 years). Male to female ratio was 4:3. Ten children were on treatment at the time of enrolment. The clinical features in the children with confirmed diagnosis included vision issues (12/14), lens subluxation/dislocation (11/14), intellectual disability (9/14), developmental delay (8/14), skeletal abnormalities (8/14), behaviour problems (6/14), spontaneous thrombosis (3/14), stroke (3/14) and seizures (2/14). One family had two affected individuals and consanguinity was reported in four families. The median (IQR) plasma total homocysteine was 117.72 µmol/litre (54-145), mean (range) urine homocysteine was 46.8 mmol/molCreatinine (18-84), median (IQR) methionine was 33.7 µmol/litre (16.4-229). Fifteen different pathogenic/ likely pathogenic variants were found in 14 cases and 12/15 variants were in highly conserved catalytic domain of CBS enzyme. Five out of 14 variants were novel (three frameshift deletions, one nonsense, one splice site variant) and were predicted to be deleterious by Mutation Taster software. Thirteen cases were found to be either homozygous or compound heterozygous for potentially pathogenic variants; however one case was found to have pathogenic variant in one allele only. **Conclusion:** In this series, varied genotypes observed in *CBS* gene resulted in classical homocystinuria with broad clinical heterogeneity in children from India.

1068W

The molecular profile of degenerative cerebellar ataxias and some unique, novel genotypes from India: A journey from short PCR to NGS based diagnostics over last decade.

S. Aggarwal^{1,2}, S. Balakrishnan¹, M. Muthu Lakshmi², A. Das Bhowmik², V. Venugopal², A.K. Meena³, K. Rukmini Mridula³, P. Ranganath^{1,2}, A. Dala^{1,2}. 1) Department of Medical Genetics, Nizam's Institute of Medical Sciences, Hyderabad, India; 2) Diagnostics Division, Centre for DNA fingerprinting & Diagnostics, Hyderabad, India; 3) Department of Neurology, Nizam's Institute of Medical Sciences, Hyderabad, India.

This study reports the molecular profile of patients presenting with degenerative cerebellar ataxia at a tertiary Medical Genetics facility in South India over last ten years. The study is retrospective from 2008-2016 and prospective from 2017 onwards. Patients satisfying inclusion criteria were recruited and triaged for molecular testing on basis of clinical presentation, pedigree information, neuroimaging and other ancillary laboratory findings. Molecular testing involved short PCR for Spinocerebellar ataxia(SCA) type 1,2,3,6,7 and TP-PCR for Friedreich ataxia(FA) in the retrospective cohort. The prospective cohort in addition underwent second tier testing for SCA 8,12,17,36 by short PCR and selected cases with autosomal recessive pattern or syndromic/complex presentations were evaluated using phenotype guided tests or Next generation sequencing. 165/557(30%) patients tested positive for SCA 1,2,3 with equal prevalence of the three subtypes. 452 patients underwent testing for SCA6 & 7, leading to a diagnosis of SCA7 in six, and SCA6 in one. None of the 29 individuals tested for SCA 8,12,17 and 36 had a positive result. 106/379(28%) individuals tested positive for Friedreich ataxia, indicating FA to be the single most common diagnosis in our cohort. Phenotype based targeted testing yielded a diagnosis of SCA 34 in one patient and Huntington disease in another. One patient each homozygous for SCA 1 and 3 were found and one individual with FA due to homozygous deletion of *FXN* was identified in the cohort. Targeted exome sequencing interrogating the known disease causing OMIM genes yielded final diagnosis in 6/7 patients, including 3 cases of ataxia with oculomotor apraxia, and one each of SCA 23,29 and Dystroglycanopathy. WES revealed a putative pathogenic variant in *PCDH12* gene in two sibs with a complex phenotype of cerebellar ataxia, dystonia and exudative retinopathy. To conclude, FA and SCA 1,2,3 account for a significant proportion of the degenerative cerebellar ataxias in our cohort, while the other SCAs are extremely rare. NGS based testing in selected cases showed a high diagnostic yield and revealed a putative novel gene for ataxia-dystonia-retinopathy phenotype in one sibpair. This data provides interesting clues for designing a diagnostic algorithm suited for our population.

1069T

A biallelic frameshift variant in a novel gene results in ataxia and dysarthria in humans. F. Aslam, S. Naz. School of Biological Sciences, University of the Punjab, Lahore, Punjab, Pakistan.

Hereditary ataxias are a heterogeneous group of disorders that are characterized by uncoordinated body movements. Ataxias may be manifested with gait abnormalities, dysarthria, apraxia, dysphagia and other pyramidal and extrapyramidal signs. More than forty genes have been implicated in the etiology of recessively inherited ataxias and others remain to be discovered. We recruited a consanguineous Pakistani family with two affected siblings. Patients suffered from a movement disorder but had no specific diagnosis for their disorder. We performed clinical assessments and molecular analyses which included magnetic resonance imaging and whole-exome sequencing to delineate the underlying pathology of the disorder. Whole-exome sequencing data were evaluated for all homozygous exonic and splice-site variants which were predicted to be pathogenic and had a frequency of less than 0.01 in public databases. Segregation analyses were performed for the variants by Sanger sequencing. Effect of variant on RNA expression was evaluated by gene specific PCR on an oligo dT primed cDNA library which was constructed from patient RNA, extracted from blood. Clinical evaluations of the patients revealed that both affected individuals suffered from abnormalities of gait, limb incoordination and dysarthria. These conditions were consistent with a diagnosis of ataxia. Magnetic resonance imaging conducted for the oldest affected individual did not reveal brain abnormalities. Whole-exome data analyses identified a frameshift variant of a gene which has not been previously associated with a human disorder though knockout of its orthologous gene is known to cause movement related phenotypes in mice. Patients were homozygous for the variant, obligate carriers were heterozygous and an unaffected sibling did not have the variant. The variant was absent in all public databases and from the DNA of ethnically matched controls. cDNA analysis demonstrated that the mutant RNA escapes nonsense mediated decay, at least in blood. If the mutant mRNA is translated, half of the important functional domains of the protein will be missing. Our work has identified a gene which encodes a protein participating in a conserved molecular pathway in both mice and humans. This study is funded by HEC 2877, Pakistan.

1070F

De novo mutation screening in childhood-onset cerebellar atrophy identifies gain of function mutations in the calcium channel *CACNA1G*. V. Cantagrel^{3,4}, J. Chemin^{1,2}, K. Siquier-Pernet^{3,4}, G. Barcia^{3,4}, A. Ahmad^{1,2}, D. Medina-Cano^{3,4}, S. Hanein⁵, N. Altin^{3,4}, L. Hubert⁶, C. Bole-Feyso⁶, C. Fourage^{7,9}, P. Nitschke⁷, J. Thevenon⁸, M. Rio^{4,9}, P. Blanc^{4,9}, C. Vidal⁵, N. Bahi-Buisson^{3,10,11}, I. Desguerre^{3,11}, A. Munnich^{3,9}, S. Lyonnet^{3,9,10}, N. Boddaert^{3,12,13}, E. Fassi¹⁴, M. Shinawi¹⁴, H. Zimmerman¹⁵, J. Amiel^{3,9,10}, L. Faivre⁸, L. Colleaux^{3,4}, P. Lory^{1,2}. 1) IGF, CNRS, INSERM, University of Montpellier, Montpellier, France; 2) LabEx 'Ion Channel Science and Therapeutics', Montpellier, France; 3) Paris Descartes - Sorbonne Paris Cité University, Imagine Institute, Paris, France; 4) Developmental brain disorders Laboratory, INSERM UMR 1163, Paris, France; 5) Translational Genetics, INSERM UMR 1163, Imagine Institute, Paris, France; 6) Paris Descartes-Sorbonne Paris Cité University, Imagine Institute, Genomic Core Facility, 75015 Paris, France; 7) Paris-Descartes Sorbonne Paris-Cité University, Imagine Institute, Bioinformatics Core Facility, 75015 Paris, France; 8) Centre de Génétique et Centre de Référence "Anomalies du Développement et Syndromes Malformatifs", Hôpital d'Enfants, CHU Dijon, Dijon, France; 9) Service de Génétique, Necker Enfants Malades University Hospital, APHP, Paris; 10) Laboratory of embryology and genetics of congenital malformations, INSERM UMR1163, Paris, France; 11) Service de neurologie pédiatrique, Necker Enfants Malades University Hospital, APHP, Paris; 12) Pediatric Radiology Department, Necker Enfants Malades University Hospital, APHP, Paris, France; 13) Image - Institut Imagine, INSERM UMR1163 and INSERM U1000, Université Paris Descartes, Hôpital Necker Enfants Malades, Paris, France; 14) Division of Genetics and Genomic Medicine, Department of Pediatrics, Washington University School of Medicine, St. Louis, Missouri, USA; 15) Division of Genetics, Department of Pediatrics, University of Mississippi Medical Center, 2500N State St, Jackson, MS, 39216, USA.

Cerebellar atrophy is a key neuroradiological finding usually associated with cerebellar ataxia and cognitive development defect in children. Unlike the adult forms, early onset cerebellar atrophies are classically described as mostly autosomal recessive conditions and the exact contribution of de novo mutations to this phenotype has not been assessed. In contrast, recent studies pinpoint the high prevalence of pathogenic de novo mutations in other developmental disorders such as intellectual disability, autism spectrum disorders and epilepsy. Here, we investigated a cohort of 47 patients with early onset cerebellar atrophy and/or hypoplasia using a custom gene panel as well as whole exome sequencing. De novo mutations were identified in 35% of patients while 27% had mutations inherited in an autosomal recessive manner. Understanding if these de novo events act through a loss or a gain of function effect is critical for treatment considerations. To gain a better insight into the disease mechanisms causing these cerebellar defects, we focused on *CACNA1G*, a gene not yet associated with the early-onset form. This gene encodes the Cav3.1 subunit of T-type calcium channels highly expressed in Purkinje neurons and deep cerebellar nuclei. In total, we identified four patients with de novo *CACNA1G* mutations. They all display severe motor and cognitive impairment, cerebellar atrophy as well as variable features such as facial dysmorphisms, digital anomalies, microcephaly and epilepsy. Three subjects share a recurrent c.2881G>A/p.Ala961Thr variant while the fourth patient has the c.4591A>G/p. Met1531Val variant. Both mutations drastically impaired channel inactivation properties with significantly slower kinetics (5 times) and negatively shifted potential for half-inactivation (> 10 mV). In addition, these two mutations increase neuronal firing in a cerebellar nuclear neuron model and promote a larger window current fully inhibited by TTA-P2, a selective T-type channel blocker. This study highlights the prevalence of de novo mutations in early-onset cerebellar atrophy and demonstrates that Ala961Thr and Met1531Val are gain of function mutations. Moreover, it reveals that aberrant activity of Cav3.1 channels can markedly alter brain development and suggests that this condition could be amenable to treatment.

1071W

Late onset hereditary dominant episodic ataxia in French Canadians.

K. Choquet^{1,2}, M. Renaud¹, M. Tetreault³, S. Provost^{4,5}, M.-J. Dicaire¹, R. La Piana¹, R. Massie¹, C. Chalk⁶, A.-L. Lafontaine^{1,6}, M.-P. Dubé^{4,5,7}, A. Duquette³, B. Brais^{1,2}. 1) Montreal Neurological Institute, Montreal, Québec, Canada; 2) Department of Human Genetics, McGill University, Montreal, Québec, Canada; 3) Department of Neurosciences, Université de Montréal, Centre de Recherche du Centre Hospitalier de l'Université de Montréal (CRCHUM), Montréal, Québec, Canada; 4) Institut de Cardiologie de Montréal, Université de Montréal, Montréal, Québec, Canada; 5) Centre de Pharmacogénomique Beaulieu-Saucier, Université de Montréal, Montréal, Québec, Canada; 6) Montreal General Hospital, McGill University, Montreal, Québec, Canada; 7) Faculté de Médecine, Université de Montréal, Montréal, Québec, Canada.

Episodic ataxias (EA) are a heterogeneous group of inherited neurological disorders characterized by recurrent attacks of ataxia. Eight forms have been described to date based on causal genes, presumed gene loci or clinical characteristics. Mutations in *KCNA1* (EA1) and *CACNA1A* (EA2) account for the majority of EA cases worldwide. All forms described to date follow a dominant mode of transmission with variable penetrance. We recruited a large French Canadian cohort of 65 undetermined EA cases belonging to 37 families, expecting that because of their common origin some would share the same founder mutation. We excluded families with mutations in known EA genes. We characterized the clinical phenotypes of EA, including the age at onset (AAO) of first episodes, AAO of the progressive ataxia and SARA scores. All affected patients had recurrent attacks of ataxia, often triggered by physical activity or alcohol. Variable penetrance was observed in all families. The mean AAO was 54.4 years (30-72) for the first episodes and 60 years (45-72) for the progressive ataxia. Ataxia was mild to moderate with a mean SARA score of 8.9/40 (1-18). We established that 50% of affected patients have diminished vibration sense in their lower extremities. Brain MRI showed moderate cerebellar atrophy in permanently ataxic cases. In the largest studied family, we observed ataxia with mean AAO of 56.4 years (40-64) for first episodes and 59.5 years (45-67) for progressive ataxia, but with major variability in the severity and duration of symptoms (mean SARA score: 6/40 (0-18)). To uncover the genetic cause(s) of disease, we first performed whole exome sequencing (WES) in the three largest families, which failed to identify potentially pathogenic variants that co-segregated with the disease. This led to whole genome sequencing (WGS) in two affected cases in each of the three families, which again did not share any rare variant. To prioritize the thousands of variants identified within families, we launched linkage analysis using SNP genotyping in the largest family (9 affected and 11 unaffected) in the hope to uncover a locus that will allow targeted analysis of the WGS data sets. In conclusion, this new French Canadian cohort of late-onset exercise-induced EA illustrates that WES and WGS alone may not be sufficient to uncover mutations for clinically variable autosomal dominant conditions.

1072T

107 Tunisian patients with autosomal recessive ataxia: A clinical and genetic study. M. Hammer¹, H. Nehdi², G. El Euch Fayache², J. Ding¹, JR. Gibbs¹, R. Amouri², F. Hentati², A. Singleton¹. 1) National Institutes of Health, Bethesda, MD; 2) National Institute of Neurology, Tunis, Tunisia.

Background Autosomal recessive cerebellar ataxia (ARCA) comprises an extensive group of heterogeneous neurological disorders where balance is lost without muscle weakness. Whole exome sequencing presented a significant impact in the identification of disease causing genes driving the number to over 50. **Objective** We aim at identifying the genetic origin of ARCA in a cohort of 107 Tunisian patients. **Methods** This study comprised 79 patients belonging to 32 families with autosomal recessive pattern of inheritance and 28 patients with no family history. The strategy consisted in: 1) Performing a TP-PCR to screen for GAA trinucleotide repeat in *FXN* responsible for Friedreich ataxia 2) Screening with direct sequencing for the frame-shift deletion (c.744delA) in *TTPA* associated with ataxia with vitamin E deficiency and for the mutations reported in Tunisia at the time of the study in *APTX*, *SETX* and *SACS* genes, respectively responsible for ataxia with oculomotor apraxia type 1, ataxia with oculomotor apraxia type 2 and ataxia of Charlevoix-Saguenay 4) Whole exome sequencing for the remaining cases. **Results** The (GAA)_n expansion in *FXN* was present at the homozygous state in 7/107 patients. Only two siblings (2/107) were found to be homozygous for the c.744delA mutation in *TTPA*. Screening for mutations reported in Tunisia in *APTX*, *SETX* and *SACS* revealed two mutations in each gene in a total of 16/107 patients. Exome sequencing allowed the identification of two known *SETX* mutations not previously reported in Tunisia as well as known mutations in other genes associated with ARCA and spastic paraplegia (*SLC25A46*, *APOB*, *SPG20* and *SPG11*) in 13/107 affected individuals. Novel probably disease-causing variants were also found in common ARCA genes (*SACS*, *SETX*, *GRM1*, *ADCK3*, etc.) and in genes associated with different autosomal recessive disorders where ataxia is only a clinical feature (*POLR3A*/ Hypomyelinating leukodystrophy 7, *KCNJ10*/ SeSAME syndrome, *PEX16*/ Zellweger syndrome). Most importantly, strong novel candidate genes were identified such as *MAP1A* and *CHD6* which lead to an ataxic phenotype in homozygous mutated mice. **Conclusion** We used different techniques to investigate the genetic origin of ataxia in a cohort of 107 patients but only found the probable disease-causing mutation in 72 individuals. Whole genome sequencing might be the next step to take in hope to unravel the genetic cause of ataxia in the remaining cases.

1073F

Next-generation sequence enabled diagnosis of family aggregated ataxia in an endogamic Colombian population. J.A. Tejada Moreno¹, C.A. Villegas Lanau^{1,2}, L. Madrigal Zapata², A.Y. Baena², O. Campo Nieto¹, L. Rishishwar^{3,4}, E.T. Norris^{3,4,5}, A.T. Chande^{3,4,5}, I.K. Jordan^{3,4,5}, G. Bedoya Berrio². 1) Molecular Genetics Research Group, University of Antioquia, Medellin, Colombia; 2) Neurosciences Research Group, University of Antioquia, Medellin, Colombia; 3) IHRC-Georgia Tech Applied Bioinformatics Laboratory, Atlanta, Georgia, USA; 4) PanAmerican Bioinformatics Institute, Cali, Valle del Cauca, Colombia; 5) School of Biological Sciences, Georgia Institute of Technology, Atlanta, Georgia, USA.

Neurodegenerative diseases are a class of severe pathologies, including dementia and numerous movement disorders. Diagnosing many of these disorders is a complex and resource-intensive process and can involve multiple invasive procedures. Furthermore, clinical tests of this kind often do not result in any conclusive diagnosis. This is an even bigger challenge in a limited resource environment such as Colombia. Numerous recent studies have shown the advantages of adopting next-generation sequencing (NGS) based diagnostic assays for quicker and more precise diagnosis of patients with previously inconclusive diagnoses. Our group is developing approaches for NGS enabled diagnosis of neurodegenerative diseases in the Colombian population of Antioquia. Historically, the population of Antioquia has shown high levels of endogamy and corollary high prevalence of autosomal recessive diseases. Here, we discuss our efforts to apply NGS based diagnosis to autosomal ataxia (AA) in five multigenerational families from Antioquia. Standard clinical and neurodevelopmental tests performed on the probands from these families have not been sufficient alone for unambiguous diagnoses. In an effort to yield more precise diagnoses for these patients, we used whole exome sequencing (WES) of the AA probands and two immediate family members (parent(s) and/or sibling). We evaluated 367 autosomal recessive and 118 autosomal dominant AA candidate genes compiled from the literature for potential AA causal single nucleotide variants. Family pedigrees were first evaluated to characterize AA cases as autosomal recessive or dominant, and WES variants were screened accordingly by comparing affected and unaffected family members. We also used a number of additional variant filters following the American College of Medical Genetics and Genomics guidelines. Potential causal variants were considered in the context of the clinical, pathophysiological and neuropsychological characteristics of these patients in order to converge on the best NGS enabled diagnosis for each patient. While further work is needed to establish the mode of action of the putative causal variants, we have shown the viability and power of NGS based tests in Colombia for neurodegenerative disease diagnosis.

1074W

Hotspot variants in the *GRIK2* kainate receptor gene are causative for neurodevelopmental disorders with diverse phenotypes. G. T. Swanson¹, J. R. Stolz¹, G. L. Carvill², B. Keren³, C. Mignot⁴, P. R. Mark⁴. 1) Pharmacology, Northwestern University Feinberg School of Medicine, Chicago, IL; 2) Neurology, Northwestern University Feinberg School of Medicine, Chicago, IL; 3) Département de Génétique, Hôpital Pitié-Salpêtrière, Paris, France; 4) Spectrum Health Medical Genetics, Grand Rapids, MI.

Ionotropic glutamate receptors (iGluRs) are critical signaling molecules in the central and peripheral nervous systems, and missense variants in genes encoding several iGluR subunits have been shown to cause autism and related neurodevelopmental disorders (NDDs). We recently reported the first missense variant (p.A657T) in a gene encoding a kainate receptor (KAR) subunit subtype of the iGluRs, *GRIK2* (Glutamate receptor, ionotropic, kainate 2; GluK2 protein) (Guzmán et al., *Neurol. Genet.*, 2017). This *de novo* variant was determined to be causative for a non-syndromic NDD with intellectual disability, speech deficits and ataxia as primary clinical features. The p.A657T substitution occurs in the pore-forming third membrane (M3) domain, which has been characterized as a "hotspot" of NDD-causing mutations in other iGluR subunit genes. The biophysical properties of GluK2(A657T) subunit-containing receptors were different from wildtype GluK2 in several substantive ways that included a greatly enhanced sensitivity to glutamate, slowed decay kinetics and destabilization of the desensitized state of the receptor. This study prompted us to determine if pathogenic variants in the same hotspot region of *GRIK2* or other KAR subunit genes were implicated in NDDs more broadly. In this report, we will present clinical data from two additional children with *de novo* *GRIK2* variants that occur in the exon encoding the M3 and the adjacent linker between M3 and the S2 segment of the ligand-binding domain (M3-S2 linker). These children had highly divergent NDD phenotypes: severe pediatric epilepsy [GluK2(p.T660K) variant] and autism [GluK2(p.I668T) variant]. Introduction of the analogous mutations into the GluK2 subunit also had varied effects on the expression and biophysical properties of the receptors. GluK2(T660K) receptors were very similar to GluK2(A657T) in their slowed decay kinetics and destabilized desensitized states, whereas GluK2(I668T) had profoundly reduced peak amplitudes and rapid desensitization. These data demonstrate that variants in the *GRIK2* KAR subunit gene in the M3/M3-S2 hotspot region alter channel function and can be causative for a spectrum of NDDs, including intellectual disability, autism, epilepsy and movement disorders such as ataxia. KAR subunit genes should therefore be added to the growing list of synaptic signaling molecules whose disruption can cause NDDs of variable presentation.

1075T

Functional characterization of biallelic *SZT2* variants in a patient with seizures and brain abnormalities. J. Calhoun¹, I. Bellinski¹, L. Kinsley¹, M. C. Schultz², A. Ramamurthy¹, E. Bryant¹, J. O'Connor², J. Millichap², E. Gerard¹, G. Carvill¹. 1) Northwestern University Feinberg School of Medicine, Chicago, IL; 2) Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL.

Compound heterozygous *SZT2* variants were first implicated in autosomal recessive epileptic encephalopathy with recurrent seizures, developmental delay, macrocephaly, and corpus callosum abnormalities. *SZT2* encodes a large gene product with no discernible structural similarities to other known proteins. Due to this lack of overt functional protein architecture, the function of *SZT2* in the cell remained elusive for several years; though a *Szt2* mouse model was originally reported to modify seizure susceptibility. Recently *SZT2* was shown to form part of the KICSTOR complex responsible for recruiting the GATOR complex to the lysosome where it acts as a negative regulator of mTORC1 (mechanistic target of rapamycin complex 1) in the context of amino acid deprivation. We identified a proband with compound heterozygous *SZT2* missense variants of unknown significance (VUS) in a patient with adult-onset epilepsy, macrocephaly, and MRI abnormalities. While the majority of published patients are reported with truncating variants, suggesting complete loss of function for *SZT2*, the patient described here carried missense variants in an uncharacterized portion of *SZT2*. We hypothesized that these missense variants may be full or partial loss-of-function *SZT2* variants and may lead to a milder phenotype than the null allele which is associated with early onset epilepsy with poor outcome. We examined whether primary dermal fibroblasts from the proband displayed altered regulation of mTOR activity in response to amino acid deprivation. Primary dermal fibroblasts from the proband display robust mTORC signaling by enhanced phosphorylation of S6K upon amino acid deprivation, consistent with a defect in the amino acid response. These results are consistent with diminished or lost *SZT2* function leading to breakdown of amino acid sensing in knockout models of *SZT2*. We are developing a cellular platform to functionally test individual *SZT2* variants which will allow us to characterize previously reported and novel *SZT2* missense variants. Analysis of additional variants provide an opportunity to define the functional domains of *SZT2*, understand how mTORC is recruited to the lysosome, unravel potential genotype-phenotype correlations and identify novel therapeutic targets.

1076F

An intronic repeat expansion in familial ALS. *M.M. Course¹, K. Gudsnuik¹, K. Winston¹, N. Desai¹, J. Ross², P.A. Dion², G.A. Rouleau², P.N. Valdmanis¹.* 1) Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA, USA; 2) Montreal Neurological Institute, McGill University, Montreal, QC, Canada.

Amyotrophic lateral sclerosis (ALS) is a devastating neurodegenerative disease, characterized by rapidly progressive muscle weakness and early death. No effective treatments are available, and our understanding of the disease remains nascent. Our lab seeks to uncover genetic mutations underlying the 40-50% of familial ALS cases with no identified causes. By applying linkage analysis to a large family with ALS, we identified a haplotype that segregates with the disease. While the family already carries a pathogenic variant in *FUS* known to cause ALS, this mutation cannot fully explain the penetrance of the phenotype. It is fully explained, though, once coupled with our disease haplotype. The haplotype region includes an intronic, human-specific tandem repeat expansion in the gene *WDR7*, which exhibits greater repeat copy length in patients with ALS as compared to unaffected individuals, both within our family and in 350 sporadic ALS patient samples as compared to 270 control samples from the Coriell Institute. We therefore hypothesize that this expansion is a genetic cause or modifier of ALS. So far we have observed that this repeat can expand up to 5520 bp, while the ancestral copy is only 105 bp. By sequencing 144 multiplexed patient samples using long-read PacBio sequencing, we were able to obtain the exact number of repeats in each allele, as well as the exact sequence of each repeat copy. The repeat also boasts a unique hairpin structure, which we have confirmed forms microRNAs. These microRNAs, in turn, are predicted to target gene products involved in ALS pathobiology, and we are now working to understand the mechanism by which this repeat expansion leads to ALS.

1077W

Coexistence of schwannomatosis and glioblastoma in two families. *C. Deiller^{1,2}, C. Goizet¹, C. Zordan¹, J. Tinat¹, H. Loiseau³, T. Fabre⁴, C. Delleci⁵, J. Cohen⁶, M. Vidaud^{6,7}, B. Parfait^{6,7}, J. Van-Gils¹.* 1) CHU Bordeaux, Service de Génétique Médicale, Bordeaux, France; 2) Laboratoire MRGM, INSERM U1211, Univ. Bordeaux, Bordeaux, France; 3) CHU Bordeaux, Service de neurochirurgie, Bordeaux, France; 4) CHU Bordeaux, Service d'orthopédie, Bordeaux, France; 5) CHU Bordeaux, Service de Médecine Physique et Réadaptation, Bordeaux, France; 6) Service de Génétique et Biologie Moléculaires, Hôpital Cochin, Assistance Publique-Hôpitaux de Paris, Paris, France; 7) EA7331, Université Paris Descartes, Sorbonne Paris Cité, Faculté de Pharmacie de Paris, Paris, France.

Schwannomatosis is a rare affection predisposing to multiple peripheral neurologic tumors. Approximately, one third of patients with schwannomatosis are carriers of a germline mutation in *LZTR1* (Leucin Zipper Transcription Regulator 1). Tumorigenesis in schwannomatosis responds to a somatic 5-hit/3-step mechanism resulting in a loss of function (LOF) of *LZTR1* and the contiguous genes of locus 22q11.2q12.2. Effectively, *LZTR1* is mapped on 22q11.2 and centromeric to *SMARCB1* also implicated in the determinism of schwannomatosis and *NF2*, responsible for neurofibromatosis type 2. The 5-hit/3-step consists in germline LOF of *LZTR1* first allele followed by two somatic events: a loss of heterozygosity of the locus 22q11.2q12.2 with the deletion of *LZTR1*, *SMARCB1* and *NF2*, and mutation of the remaining allele of *NF2*. On a somatic point of view, *LZTR1* mutations are known to drive with a significant frequency glioblastoma (GB) development. Pathophysiology of GB also requires biallelic loss of function of *LZTR1*. We report here two families in which segregate both multiple schwannomas and GB. In each family, we describe a proband with schwannomatosis due to a germline heterozygous *LZTR1* mutation, and a relative deceased from an apparently isolated GB. Molecular analysis of GB sample from the affected relative showed the presence of the familial mutation. These observations suggest a potential link between schwannomatosis and the GB development.

1078T

Huntington disease in Africa: Emerging genetic and phenotypic differences. A. Krause¹, D.G. Anderson¹, J. Levesley¹, F. Baine-Savanhu¹, M. Ciosi², D. Monckton², R. Margolis³. 1) Human Genetics, National Health Laboratory Service and University of the Witwatersrand, Johannesburg, Gauteng, South Africa; 2) Institute of Molecular, Cell and Systems Biology, University of Glasgow, Glasgow, Scotland; 3) Departments of Psychiatry and Neurology, Program in Cellular and Molecular Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA.

Huntington disease (HD) is an inherited, progressive neurodegenerative disease. Most patients with an HD phenotype have a trinucleotide repeat expansion in the *HTT* gene. However, an HD phenocopy (HDL2) due to a trinucleotide expansion in the *JPH3* gene, accounts for 30% of confirmed HD diagnoses among black South Africans, with the highest number of cases reported number worldwide. Although the clinical phenotype of HDL2 is reported to be broadly similar to HD, some differences have been proposed, based on case studies. We report preliminary findings of a cross-sectional study to characterize the clinical phenotype of HDL2 from a neurological, neuropsychological, haematological and radiological perspective; and to compare it to HD. The eyes signs were indistinguishable in a comparison of 13 HD and 9 HDL2 patients, suggesting that oculomotor signs are not a feature that distinguishes HDL2 from HD as initially proposed. There was no significant difference in the total score between the two diseases or in any sub-score parameters of ocular pursuit, saccade initiation or saccade velocity. Early reports suggested that the presence of acanthocytes in the peripheral blood smear might distinguish patients with HDL2 from those with HD. No acanthocytes were seen in 12 South African cases studied. In MRI studies, using semiautomated volumetric analysis and morphometry, the pattern of volume loss was similar in HDL2 and HD but tended to be greater in the HDL2 group. Significantly more thalamic volume loss was seen in HDL2 relative to HD and controls when controlled for age, repeat length and disease duration. HD in African patients has molecular differences from HD in European patients - *HTT* expansions occur on different haplotype backgrounds. Further, the CAG/CCG repeat tract in *HTT* shows significantly higher sequence diversity among Africans. The most common disease allele in African ancestry HD patients has an atypical structure compared to European patients. A unique *cis*-acting modifier, the CCGCCA cassette, appears to reduce the average age of diagnosis in African ancestry HD patients by 6.6 years. Studies of HD and HDL2 in Africa provide novel insights into the clinical spectrum of Huntington disease and provide unique insights into the pathophysiology of the disease. Further novel molecular insights are being gained from studies of the genetic diversity of this monogenic disease in Africa.

1079F

Novel clinical and genetic insights into dysfunction of the ASC-1 complex. A. Marais¹, A.M. Bertoli-Avella¹, A. Narravula¹, U. Altunoglu², A.M. Alhashem^{3,4}, S. Mohamed⁵, A. Alghamdi⁶, P. Willems⁶, E. Tsoutsou⁷, H. Fryssira⁷, R. Pons⁷, R. AL Marzooq⁸, A.R. Al Madhoob⁹, M. Calvo Del Castillo¹, Z. Yüksel¹, K.K. Kandaswamy¹, M. Werber¹, A. Rolfs^{1,9}, P. Bauer¹. 1) Centogene AG, Rostock, Germany; 2) Istanbul University, Istanbul Medical Faculty, Medical Genetics Department, Istanbul, Turkey; 3) Division of Genetics and Metabolic Medicine, Department of Pediatrics, Prince Sultan Military Medical City Riyadh, Saudi Arabia; 4) Department of Anatomy and Cell Biology, College of Medicine, Alfaisal University, Riyadh, Saudi Arabia; 5) Pediatrics Department, Prince Sultan Military Medical City, Riyadh, Saudi Arabia; 6) GENDIA, Genetic Diagnostic Network, Belgium; 7) First Department of Pediatrics, "Aghia Sophia" Children's Hospital, University of Athens, Greece; 8) MBBS, Salmaniya Medcial Complex, Ministry of Health, Manama, Kingdom Of Bahrain; 9) Albrecht-Kossel-Institute for Neuroregeneration, Medical University Rostock, Rostock, Germany.

The nuclear activating signal cointegrator 1 (ASC-1) complex has been suggested to participate in RNA processing events and in transcriptional coactivation, but specific downstream targets remain elusive. It is composed of the four subunits ASC-1 (encoded by *TRIP4*), ASCC1 (*ASCC1*), ASCC2 (*ASCC2*) and ASCC3 (*ASCC3*). Homozygous variants in *TRIP4* (n=4 families) and in *ASCC1* (n=4 families) were recently associated with severe neuromuscular phenotypes. Upon whole exome sequencing (WES) of a clinically heterogeneous patient cohort, we identified the *TRIP4* variants c.512G>A (p.Cys171Tyr) and c.1678+1_1678+2insC as well as the *ASCC1* variants c.626+1G>A and c.897G>A (p.Trp299*) in homozygosity in one individual each; Sanger sequencing confirmed these WES-derived findings. The two patients with homozygous *TRIP4* variants (currently aged 7 months and 14 years, respectively) shared muscular hypotonia, joint contractures, scoliosis, microretrognathia and high palate; bone fractures were not observed. The homozygous *ASCC1* variants were identified in an aborted fetus and in an infant that died at 27 days of age from massive pleural effusion. Both were characterized by decreased fetal movements, thin ribs, and fractures of long bones (humeri and femora). The apparent differences in severity were followed up by a meta-analysis of all cases known to date (n=12 for *TRIP4* vs. n=11 for *ASCC1*). This revealed that variants in *ASCC1* are more often associated with congenital bone fractures (p=0.011, two-sided Fisher's exact test) and with perinatal death (p=0.003). We also noticed that all known pathogenic *ASCC1* variants are predicted to entail complete absence of the ASCC1 protein, while this was not true for any of the pathogenic *TRIP4* variants (presence of ASC-1 protein species with (i) a single amino acid exchange, (ii) an internal in-frame deletion, or (iii) an altered C-terminus). Our findings broaden the mutational spectrum for ASC-1 complex-related disease, and suggest that the degree of severity depends on the mutated subunit and/or the mutational consequences. We are currently following up on genetic findings for *ASCC2* and *ASCC3* to see whether variants in these two genes can be incorporated into the above concept.

1080W

Loss-of-function variants in the gene *IRF2BPL* are associated with neurological phenotypes. L.D.M. Pena¹, P.C. Marcogliese², R.C. Spillmann¹, N. Stong³, J.A. Rosenfeld², M.K. Koenig⁴, D. Ortiz⁵, K. Riley¹, G. Mirzaa^{6,9}, D. Hemelsoet⁷, S. Yamamoto^{2,11,12,13}, M.F. Wangler^{2,11,12}, B. Lee², S.F. Nelson⁸, V. Shashi¹, H.J. Bellen^{2,10,11,12}, UD-ProZA Members, UDN Members. 1) Division of Medical Genetics, Department of Pediatrics, Duke University School of Medicine, Durham, North Carolina; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas; 3) Institute for Genomic Medicine, Columbia University, New York, New York; 4) Division of Child & Adolescent Neurology, Department of Pediatrics, University of Texas Health Sciences Center, Houston, Texas; 5) Children's Hospital of Pittsburgh, University of Pittsburgh, Pittsburgh, Pennsylvania; 6) Center for Integrative Brain Research, Seattle Children's Research Institute; 7) Department of Medical Genetics, Ghent University Hospital, Ghent, Belgium; 8) Department of Human Genetics, David Geffen School of Medicine, University of California - Los Angeles, Los Angeles, California; 9) Department of Pediatrics, University of Washington, Seattle, Washington; 10) Howard Hughes Medical Institute, Houston Texas; 11) Program in Developmental Biology, Baylor College of Medicine, Houston, Texas; 12) Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, Texas; 13) Department of Neuroscience, Baylor College of Medicine, Houston, Texas.

The Interferon Regulatory Factor 2 Binding Protein-Like (*IRF2BPL*) is a single-exon gene with unknown function or disease association in humans. We report seven individuals with neurodevelopmental phenotypes and heterozygous variants in *IRF2BPL* identified by exome sequencing. Current age ranges between 2-43 years, and 5/7 had normal early childhood development. The five cases with nonsense variants had neurological regression that started with deterioration of gait and loss of intentional movements during childhood. Subsequently they developed dysarthria, loss of speech, dysphagia, spasticity, incontinence, and a progressive movement disorder. Although brain MRI was normal in childhood, the two oldest cases had diffuse cerebral atrophy. All had abnormal EEGs and eventually developed epilepsy. The two cases with missense variants had a generally milder course with epilepsy and developmental delay, but no developmental regression or a movement disorder. The variants are *de novo* for all who had parental testing. *IRF2BPL* is a very intolerant gene with an RVIS score of 9.3%, a pLI score of 0.97, and constraint to missense variants ($z = 4.73$). All variants were absent from the gnomAD and ExAC databases. The CADD score for all variants is over 34, among the most deleterious. The DOMINO tool yielded a high score that predicts that monoallelic variants in *IRF2BPL* would very likely cause disease. The fly homolog of *IRF2BPL* is the uncharacterized conserved gene *pits*. *Pits* is robustly expressed in the nucleus of neurons. Complete loss of *pits* is lethal early in development. Partial knock-down with RNA interference in neurons leads to neurodegeneration, revealing requirement for this gene in proper neuronal function and maintenance. The nonsense variants in *IRF2BPL* identified in patients behave as severe loss-of-function alleles in flies, while ectopic expression of the missense variants leads to a range of phenotypes. We present a new, undiagnosed neurological phenotype with variants in *IRF2BPL*. The bioinformatics signature of the gene and variants supports pathogenicity. Modeling in the fruit fly revealed that *pits* is important for development and neuronal integrity over time, and showed that the nonsense variants are strong loss-of-function. Additional model organism work is ongoing to clarify gene function and inform the clinical phenotype.

1081T

Multi locus variation in a neurodevelopmental disease cohort with high identity-by-descent. J. Punetha¹, E. Karaca^{1,2}, J.E. Posey¹, Z. Coban Akdemir¹, X. Song¹, D. Pehlivan^{1,3}, S. Isikay⁴, T. Tos⁵, T. Harel^{1,6}, S. Jhangiani⁷, Y. Bayram^{1,8}, J. Fatih¹, D. Muzny¹, R.A. Gibbs¹, J.R. Lupski^{1,3,7,9}, Baylor-Hopkins Center for Mendelian Genomics. 1) Baylor College of Medicine, Houston, TX; 2) Department of Genetics, University of Alabama at Birmingham, Birmingham, Alabama, USA; 3) Texas Children's Hospital, Houston, TX, USA; 4) Department of Physiotherapy and Rehabilitation, Hasan Kalyoncu University, School of Health Sciences, Gaziantep, Turkey; 5) Department of Medical Genetics, Dr. Sami Ulus Research and Training Hospital of Women's and Children's Health and Diseases, Ankara, Turkey; 6) Department of Genetic and Metabolic Diseases, Hadassah-Hebrew University Medical Center, Jerusalem 91120, Israel; 7) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA; 8) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York; 9) Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA.

Multilocus pathogenic rare variation refers to the presence of pathogenic variant alleles in two or more disease-causing genes. Recent studies from our lab (Karaca *et al*, Genet Med. 2018) have shown that multilocus variation can explain some cases of apparent phenotypic expansion, with an overall rate of multiple molecular diagnoses in 12% of molecularly diagnosed neurodevelopmental cases (13/108 families), including 42% (8/19) of cases with apparent phenotypic expansion. In this study, we have performed whole exome sequencing (WES) and analysis on an additional 47 families with neurodevelopmental disorders (NDDs). Our preliminary analyses have identified two families with multilocus variation presenting with homozygous rare variants in genes linked to both autosomal recessive (AR) and autosomal dominant (AD) diseases. The first case (BH8450-1) has likely pathogenic variants in two known disease-causing genes, *RTN2* (AD, MIM#604805) and *PNKP* (AR, MIM#613402) in the same AOH (absence of heterozygosity) region of ~21Mb on Chr19. The second case (BH8426-1) has a likely pathogenic variant in *WDR81* (AR, MIM#610185) and a homozygous missense variant in *TUBB6* (c.736C>A, p.L246I) which has been recently linked to disease (AD, MIM#617732). In both cases, we see examples of AR inheritance in AD described diseases. The overall rate of molecular diagnoses, i.e. cases for which WES was diagnostic, for our NDD cohort is 69% (121/175 families). Of these, the overall rate of multiple molecular diagnoses in our molecularly diagnosed NDD cohort is 12.4% (15/121 families), which is higher than that observed in an outbred clinical cohort as reported by Posey *et al*, NEJM 2017 (4.9%, 101/2076 families, $p=0.003$). Of the 15 families with multiple molecular diagnoses in our NDD cohort, 13 families were identified to have dual molecular diagnoses and 6/13 families showed an AR+AR inheritance model. Higher rates of presumed identity-by-descent as evidenced by AOH in this cohort aided in the identification of homozygous variants that are potentially pathogenic. Additional studies including segregation, functional analyses, and sequencing of additional families with NDDs are being carried out. Development of tools to utilize structured phenotype ontologies to determine the relative contributions to phenotype of each identified molecular diagnosis will further assist in identifying multilocus variation.

1082F

Genetic background of unexplained full-term cerebral palsy: Genomic analysis identifies candidate pathogenic variants in 9 of 17 patients.

Y. Takezawa¹, A. Kikuchi¹, K. Haginoya^{1,2,3}, T. Niihori⁴, Y. Numata-Uematsu^{1,3}, T. Inui^{2,3}, S. Yamamura-Suzuki², T. Miyabayashi², M. Anza^{2,3}, S. Suzuki-Muramoto², Y. Okubo^{2,3}, W. Endo^{2,3}, N. Togashi², Y. Kobayashi³, A. Onuma⁵, R. Funayama⁵, M. Shirota⁶, K. Nakayama⁵, Y. Aoki⁴, S. Kure¹. 1) Department of Pediatrics, Tohoku University School of Medicine, Sendai, Japan; 2) Department of Pediatric Neurology, Miyagi Children's Hospital, Sendai, Japan; 3) Department of Pediatric Neurology, Takuto Rehabilitation Center for Children, Sendai, Japan; 4) Department of Medical Genetics, Tohoku University School of Medicine, Sendai, Japan; 5) Division of Cell Proliferation, United Centers for Advanced Research and Translational Medicine, Tohoku University Graduate School of Medicine, Sendai, Japan; 6) Division of Interdisciplinary Medical Sciences, United Centers for Advanced Research and Translational Medicine, Tohoku University Graduate School of Medicine, Sendai, Japan.

Background and Objective: Cerebral palsy is a common, heterogeneous neurodevelopmental disorder that causes movement and postural disabilities. Recent studies have suggested cerebral palsy have more genetic abnormalities than expected. We hypothesized that two simple criteria, i.e., full-term births and nonspecific brain MRI findings, are keys to detecting pathogenic variants in cerebral palsy cases due to the following: (i) preterm infants are susceptible to multiple environmental factors and therefore demonstrate an increased risk of cerebral palsy and (ii) brain MRI assessment is essential for excluding environmental causes and other particular disorders. **Methods:** A total of 107 patients—all full-term-births—without specific findings on brain MRI were identified among 897 patients diagnosed with cerebral palsy who were followed at our center. DNA samples were available for 17 of the 107 cases for trio whole-exome sequencing and array comparative genomic hybridization. We prioritized variants in genes known to be relevant in neurodevelopmental diseases and evaluated their pathogenicity according to the American College of Medical Genetics guidelines. **Results:** Pathogenic/likely pathogenic candidate variants were identified in 9 of 17 cases (52.9%) within eight genes: CTNNB1, CYP2U1, SPAST, GNAO1, CACNA1A, AMPD2, STXBP1, and SCN2A. Five identified variants had previously been reported. No pathogenic copy number variations were identified. The AMPD2 missense variant and the splice-site variants in CTNNB1 and AMPD2 were validated by in vitro functional experiments. **Conclusion:** Our study showed the possibility that a subset of CP patients categorized by two simple criteria (full-term births and nonspecific brain MRI findings) include pathogenic/likely pathogenic variants despite several limitations. Further genetic investigations of larger cohorts of CP patients are required to help support our results or identify other CP subsets that may likely include genetic abnormalities. This could eventually aid the further characterization of the molecular mechanisms of CP and optimizing its therapeutic interventions.

1083W

AMPD2-related pontocerebellar hypoplasia type 9: A report of 12 new Egyptian families and further delineation of the spectrum. M.S. Zaki¹, M.S. Abdel-Hamid², M.Y. Issa¹, H.M. Elbendary³, G.M.H. Abdel-Salam¹.

1) Clinical Genetics Department, Human Genetics and Genome Research Division, National Research Centre, Cairo, Egypt; 2) Medical Molecular Genetics Department, Human Genetics and Genome Research Division, National Research Centre, Cairo, Egypt; 3) Child Health Department, Medical Research Division, National Research Centre, Cairo, Egypt.

Pontocerebellar hypoplasia (PCH) is a rare inherited heterogeneous group of disorders with a progressive neurodegeneration of prenatal onset. Currently, there are 11 partially overlapping clinical subtypes with 15 known causative genes. PCH type 9 is an autosomal recessive disorder characterized by severe microcephaly, severe psychomotor delay, spasticity, intractable seizures. The brain MRI shows hypogenesis of the corpus callosum, cerebellar hypoplasia, ventral pontine degeneration, figure 8 shape of hypoplastic midbrain and variable degrees of cortical atrophy. Herein, we report 13 patients from 12 unrelated Egyptian families with PCH9. They were 7 males and 6 females, their aged from 3 months to 2 6/12y and parental consanguinity was recorded in all families. Patients were assigned as candidates for target *AMPD2* gene sequencing based on clinical and neuro-radiological presentation. All had progressive microcephaly, severe global developmental delay, spasticity, intractable generalized and myoclonic seizures and the specific neuroimaging findings including variable degrees of the cerebellum and brain stem hypoplasia, hypogenesis of corpus callosum and cortical atrophic changes. Mutational analysis of *AMPD2* gene identified 8 novel mutations and 4 previously reported in the literature. This study emphasizes the importance of clinico-radiological characterization of PCH9 linked to *AMPD2* gene and further expand the mutational spectrum. To our knowledge, this is the largest reported series of PCH9 from the same ethnic group.

1084T

Opposite functional consequences of a novel de novo variant in the voltage-sensing domain of Kv1.2 (KCNA2) in a case of infantile epilepsy. M. Kaneko¹, A. Pantazis², G. Raca^{1,3}, W. Mitchell^{4,5}, R. Olcese^{2,6}, S. Saitta^{1,3}. 1) Center for Personalized Medicine, Children's Hospital Los Angeles, Los Angeles, CA; 2) Department of Anesthesiology & Perioperative Medicine, University of California Los Angeles, Los Angeles, CA; 3) Department of Pathology, Keck School of Medicine at University of Southern California, Los Angeles, CA; 4) Department of Neurology, Children's Hospital Los Angeles, Los Angeles, CA; 5) Department of Pediatrics, Keck School of Medicine at University of Southern California, Los Angeles, CA; 6) Department of Physiology, University of California Los Angeles, Los Angeles, CA.

Using an exome-based panel of 166 epilepsy genes, we found a heterozygous *de novo* c.906T>G missense change in *KCNA2*, in a female infant with onset of generalized seizures at age 5 months. Clusters of seizures were provoked by fever or overheating, overlapping with a Dravet phenotype. *KCNA2* encodes the Kv1.2 voltage-gated potassium channel, expressed in the nervous system. This p.Phe302Leu change in the S4 segment of Kv1.2's voltage-sensing domain (VSD) was not previously reported in either general population (gnomAD) or in clinical variant (ClinVar) databases. Amino-acid sequences of the S4 segment are identical from human through *Drosophila* and 6 heterozygous pathogenic variants in S4 have been reported, with varied patient phenotypes from infantile epileptic encephalopathy to spastic paraplegia. The functional effects of S4 variants range from loss of function, gain of function, or combined loss and gain of function. Genotype-phenotype correlation and molecular mechanisms underlying the functional effects of *KCNA2* variants are still poorly understood. We characterized the functional consequences of p.Phe302Leu (F302L) using the highly conserved rat Kv1.2 homolog expressed in *Xenopus* oocytes by cut-open oocyte voltage clamp and voltage-clamp fluorometry. Channels with F302L activated faster and at more negative membrane potentials, a gain-of-function effect. Co-expressing wild-type and F302L subunits (1:1) to simulate heterozygous expression resulted in left-shifted conductance suggesting a dominant effect. Voltage-clamp fluorometry showed that VSDs from F302L channels also activated at more negative potentials and F302L stabilized the active state of Kv1.2, facilitating channel opening at more hyperpolarized potentials. Studies of the slow-inactivation process during prolonged depolarization showed that F302L inactivates at more hyperpolarized potentials with a loss-of-function effect. Our data indicate that the functional consequences of F302L may depend on neuronal excitability: in slow-firing neurons a gain-of-function effect on channel activation would predominate suppressing excitability, while in fast-firing neurons, increased Kv1.2 inactivation would favor excitability. As new variants are identified, biophysical assays are useful adjuncts to test the effects on an affected protein. Knowledge of the functional consequences of a variant can help identify specific treatments personalized to a patient's genotype.

1085F

The usefulness of additional clinical evaluation in patients with complex genetic testing results. E.R. Elias¹, K. Chatfield². 1) Pediatrics and Genetics, Children's Hospital Colorado, Aurora; 2) Pediatric Cardiology, Children's Hospital Colorado, Aurora.

Introduction: Testing with Next Generation Sequencing panels have established new diagnoses for complex patients. Interpreting rare variants or VUS's can be difficult, however, and guidelines are often unclear regarding further workup. A case is reported of a patient with several unrelated genetic abnormalities, demonstrating the value of additional evaluation in clarifying the significance of variant findings. **Case History:** This 9 yr old boy has a history of pontocerebellar hypoplasia, partial agenesis of the corpus callosum (molar tooth sign), strabismus, amblyopia, and a feeding disorder. He has severe intellectual disability. On physical exam he is thin, and microcephalic. He has dysmorphic features including dysplastic pinnae, and esotropia. His orthopedic deformities include windswept lower extremities and a moderate scoliosis. **Genetic testing** demonstrated a **TUBA1A (c.1169G>A, p.Arg390His) pathogenic variant**. He also has a variant of uncertain significance in **FLNA, (c.3814C>T p.Arg1272Cys)**. This variant is predicted to impact secondary protein structure. Lastly, he is heterozygous for a duplication within 9q34.3. The duplicated region involves exons 1-3 of the **INPP5E gene**. **Discussion:** The TUBA1A gene encodes the tubulin alpha 1A protein, which is involved in neuronal migration and brain development. Variants in TUBA1A are associated with brain malformations involving the corpus callosum, brainstem and cerebellum. Patients present with microcephaly, intellectual disability, and epilepsy. An EEG was moderately abnormal, although this patient does not have clinical seizures. Mutations of X-linked FLNA encoding Filamin A, have been found in a broad spectrum of phenotypes. As aortic abnormalities may be seen in patients with FLNA mutations, an echocardiogram was done, which demonstrated mild dilation of the aortic root, with an hour-glass appearance of the aortic root/ ascending aorta. The INPP5E gene encodes inositol polyphosphate-5-phosphatase. Pathogenic variants in INPP5E cause autosomal recessive Joubert syndrome. This child has a Molar Tooth sign on MRI, but the significance of a duplication of part of the INPP5E gene is not clear. It is comforting for the family to have a partial explanation for his phenotype. Furthermore, test results led to additional evaluations showing cardiac and EEG abnormalities, which require future medical oversight. Additional testing of multiple genetic findings has led to an improvement in this patient's care.

1086W**Genetic investigation of Nemaline myopathy in Middle Eastern patients.**

A. Haghighi¹, Y. Nilipour², S. Nafissi³, A. Haghighi⁴, R.J. Desnick⁵, R. Kornreich⁵.

1) Harvard Medical School, Boston, MA; 2) Department of Pathology, Pediatric Pathology Research Center, Mofid Children Hospital; 3) Tehran University of Medical Sciences; 4) Toronto General Hospital, University of Toronto; 5) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York.

Nemaline myopathy is a rare genetic muscular disease clinically characterized by muscle weakness and hypotonia, with nemaline rods in skeletal muscle fibers. Nemaline myopathy is both genetically and phenotypically heterogeneous disease, for which a number of disease genes have been identified. We investigated the clinical features and genetic cause of nemaline myopathy in pathologically diagnosed patients from Middle Eastern families. Half of the families were consanguineous (first-cousins). We applied a targeted next-generation sequencing (NGS) myopathy panel to identify single nucleotide variants (SNVs), insertions/deletions, and copy number variations in these patients. In addition, we reviewed the previously reported cases with the identified variants and attempted to suggest genotype–phenotype correlations. Sequence analysis revealed rare homozygous SNVs in NEB (in 66% of families) and ACTA1 (in 34% of families). Segregation analysis showed that the other affected family members were homozygous for the identified variants, whereas parents were heterozygous and healthy siblings were either heterozygous or wild type. All NEB variants were predicted to cause loss of function of the gene products, whereas the ACTA1 variants were missense. Five of the identified disease-causing variants were novel. These results suggest a) NEB and ACTA1 to be the most prevalent cause of nemaline myopathy in Middle Eastern patients, and b) targeted NGS panel is a rapid and cost-effective approach to analyze the molecular basis of nemaline myopathy.

1087T**Whole exome sequencing identifies novel mutation in ChAT gene for a lethal perinatal presentation of hydrops and arthrogryposis multiplex congenita.** *D. Khattar, N. Smith, R. Hopkin.* CCHMC, Cincinnati, OH.

Arthrogryposis refers to multiple congenital contractures consequent of decreased intrauterine fetal movement. Etiology includes genetic causes such as chromosomal or single gene disorders as well as non-genetic causes including teratogen exposures and intrauterine space constraints. The cause remains unknown in more than half of the cases. Accurate diagnosis can aid in prognosis, risk assessment for future pregnancies and evaluation for preimplantation genetic diagnosis and counseling. We describe a case with recurrent pregnancy losses complicated by hydrops and arthrogryposis. Proband was a 36-year old previously healthy woman who was being evaluated for an abnormal second trimester ultrasound with fetal joint contractures, skin and scalp edema, bilateral pleural effusion and abdominal ascites. Her obstetric history was significant for one healthy live term birth, one first-trimester loss, and two mid-trimester losses which included a 26-week singleton neonate with hydrops and arthrogryposis and 22 week triplets, where baby B was affected with a similar presentation. No drug exposures, infections, or illness were identified. First trimester screening was normal. Amniocentesis was performed and alpha-fetoprotein level, chromosomes and microarray were sent, all of which returned normal. At 22 weeks gestation, the pregnancy was complicated by PPROM with subsequent neonatal demise. Postnatal evaluation and tissue culture of facia lata was performed. Whole exome sequencing was done as a trio with parental samples. Compound heterozygous mutations were identified in the ChAT gene, one maternally inherited (c.1663G>T) and the other was paternally inherited (c.1811T>A), consistent with a diagnosis of presynaptic congenital myasthenic syndrome (CMS). These variants occurred at a highly conserved nucleotide and were predicted to be deleterious by multiple software programs including PolyPhen, SIFT and MutationTaster. Conclusion: Biallelic mutations in ChAT result in a deficiency of choline acetyltransferase which impairs presynaptic neuromuscular transmission. It is characterized by variable presentation of apnea, hypotonia and respiratory distress with an onset anytime from birth to childhood. To our knowledge, this is the first description of a case of severe perinatal CMS presenting as fetal losses with hydrops and arthrogryposis. Future directions include investigating potential treatment options including cholinergic agonists in the perinatal period.

1088F

Expanding the phenotype of chromosome 17q21.31 duplication syndrome: Severe epilepsy and self-injuring behavior as prominent features. C.M. Lourenco¹, Y.B.S. Valeta¹, L. El-Kadre¹, R. Calvo¹, R. Bogas¹, D. Sas², L. Lima². 1) Clinical Genetics, Centro Universitario Estacio de Ribeirao Preto, Ribeirao Preto, Sao Paulo, SAO PAULO, Brazil; 2) Genotyping - Diagnósticos Genéticos, Botucatu, Sao Paulo, Brazil.

Background: 17q21.31 microdeletion syndrome (Koolen-de Vries syndrome), is a rare copy number variants associated in humans with intellectual disability, friendly behavior, congenital malformations. Recently, a novel microduplication syndrome involving the reciprocal 17q21.31 was described and to **date only 10 cases were reported**. is not so well characterized. Here we describe the 11th patient with 17q21.31 duplication in a Brazilian family. **Methods/Results:** Clinical features and genetic data are reported, and compared with previously reported patients with 17q21.31 microduplication. It was possible to show that duplication arose de novo. Intellectual skills showed moderate mental retardation with progressive cognitive decline not reported/described in other patients so far. Patient showed poor social interaction, self-mutilating behavior and refractory epilepsy, besides relationship difficulties, reminiscent of autistic spectrum disorders. **Conclusion:** Patient clinical characterization helps to further delineate and expand the clinical spectrum of 17q21.31 microduplication syndrome; progressive cognitive decline was not previously described and can be related to the recent findings showing 17q21.31 duplication as the genetic basis of a novel entity marked by prominent tauopathy, leading to early-onset dementia with an Alzheimer disease clinical phenotype.

1089W

Prodromal Parkinson's disease in patients with 22q11.2 microdeletion syndrome: Preliminary results. G.M. Repetto¹, P. Chana^{2,3}, T. Cordova⁴, A. Cuiza⁴, R. Fritsch⁵, C. Jurí⁶, V. Kramer⁷, A. Ocampo⁴, C. Ornstein⁵, K. Villanueva⁴. 1) Universidad del Desarrollo, Santiago, Chile; 2) Centro de Trastornos del Movimiento, Santiago, Chile; 3) Dept of Neurology, Universidad de Santiago, Chile; 4) Instituto de Ciencias Biomedicas, Universidad de Chile; 5) Dept of Psychiatry, Facultad de Medicina, Universidad de Chile; 6) Dept of Neurology, Facultad de Medicina, P. Universidad Catolica de Chile; 7) Positron-Med, Santiago, Chile.

Parkinson's disease (PD) is the second most common neurodegenerative disorder. The classical motor symptoms are preceded by non-motor, prodromal manifestations by a decade; prodromes include anosmia, autonomic dysfunction and REM-sleep behavior disorder (RBD). Chromosome 22q11.2 microdeletion syndrome (22q11DS) was recently recognized as a cause of PD, particularly of early onset (<45 years of age). The aim of this study is to assess the presence of prodromal and motor PD findings in adults with 22q11DS, using the International Parkinson and Movement Disorder (MDS) research criteria, namely: motor testing with UPDRS scale, autonomic symptoms with COMPASS-21 scale, olfactory function with "Sniffin' Sticks", RBD-13 screening questionnaire, one full-night domiciliary video-polysomnographic (v-PSG) record and PET-CT with the presynaptic dopamine transporter (DAT) tracer ¹⁸F-PR04-MZ. Data on clinical and molecular manifestations of 22q11DS, a cognitive evaluation using WISC-IV and a clinical psychiatric and MINI interview, including use of antipsychotic medications was obtained. We report the results on the first 12 participants: 6 males and 6 females, median age 29.5 years (range 18-49 years), none had PD. Data were available on deletion size for 9 patients, 6 had the common LCR-AD 3Mb deletion, and 3 had an AB 1.5Mb deletion. Ten had a history of repaired open (1) or submucous (9) cleft palate. Full-scale IQ median was 74 (range 56-87). Three patients had psychosis; two were receiving aripiprazole and one, clozapine. UPDRS score average was 10 (range 1-16, normal 0); three patients had olfactory scores in the anosmia range, none had orthostatic hypotension. Home v-PSG performed in 7 participants, showed no objective sign of REM-sleep without atonia or behavioral RBD. PET-CT results showed increased, symmetric ¹⁸F-PR04-MZ DAT signaling compared to age and gender-matched controls (130% of controls in the caudate nucleus, 115% in putamen). These initial results suggest the presence of anosmia and motor disturbances. The observed increase in dopamine signaling on PET-CT may be related to haploinsufficiency of COMT, located within the deletion region and involved in dopamine metabolism. Caution should be taken when analyzing these results, due to the limited sample size and the lack of validated screening tools for patients with cognitive deficits. Funded by Fondecyt-Chile #1171014 and Fundacion Puelma.

1090T

Clinical and genomic characterization of seven novel *PLP1* deletion cases reveals large-deletion syndrome specific to female cases. H. Hijazi¹, G.M. Hobson², L. Bernardini³, S.S. Mar⁴, M.A. Manning^{5,6}, A. Hanson-Kahn^{7,8}, C. Gonzaga-Jauregui⁹, P. Simons⁴, K. Sperle², F.S. Coelho^{10,11}, J.A. Lee¹², P. Fang¹³, S.W. Cheung¹³, K. Inoue¹⁴, D. Pehlivan^{1,15}, J.R. Lupski^{1,16,17,18}, C.M.B. Carvalho¹. 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Nemours Biomedical Research, Alfred I. duPont Hospital for Children, Wilmington, DE; 3) Cytogenetics Unit; "Casa Sollievo della Sofferenza" Foundation, San Giovanni Rotondo (FG), Italy;; 4) Department of Neurology, Washington University School of Medicine, St. Louis, MO; 5) Stanford University, Department of Pediatrics and Medical Genetics, Palo Alto, CA; 6) Stanford University, Department of Pathology, Palo Alto, CA; 7) Stanford University, Department of Genetics, Palo Alto, CA; 8) Lucile Packard Children's Hospital, Department of Pediatrics, Division of Medical Genetics, Palo Alto, CA; 9) Regeneron Genetics Center, Regeneron Pharmaceuticals, Inc., Tarrytown, NY; 10) Programa de Pós-Graduação em Genética Departamento de Biologia Geral – UFMG, Belo Horizonte, MinasGerais, Brazil; 11) Centro de Pesquisas René Rachou – FIOCRUZ, Belo Horizonte, MG, Brazil; 12) Molecular Diagnostic Laboratory, Greenwood Genetic Center, Greenwood, SC; 13) Baylor Genetics Laboratories, Houston, TX; 14) Department of Mental Retardation and Birth Defect Research, National Institute of Neuroscience, National Center of Neurology and Psychiatry, Kodaira, Japan; 15) Section of Neurology, Department of Pediatrics, Baylor College of Medicine, Houston, TX; 16) Human Genome Sequencing Center, BCM, Houston, TX; 17) Department of Pediatrics, BCM, Houston, TX; 18) Texas Children's Hospital, Houston, TX.

PLP1 is a disease-causing gene in which variation can result in X-linked Pelizaeus–Merzbacher disease (PMD, MIM# 312080) or Spastic Paraplegia type 2 (SPG2, MIM# 312920). Pathogenic variation in *PLP1* includes point mutations, whole-gene duplications and deletions. The frequency, as well as phenotypic effect of each of these lesions, vary extensively within same-sex individuals as well as affected males and females owing to X-chromosome inactivation (XCI) in the latter. Microdeletions involving *PLP1* are the rarest form of all lesions (<10% of phenotypic cases) and have been associated with a complicated form of SPG2. We describe the genomic features of 7 novel *PLP1*-deletion events and expand on the clinical picture associated with large genomic deletions encompassing this gene. Three of the reported cases elucidate a severe infantile-onset neurological phenotype that is specific in its presentation to female cases. The minimal region of overlap between deletion events identified in similarly manifesting female cases was investigated for potential modifier regions. Moreover, the size of the deletion appears to play a role in early-onset penetrance of disease in the majority of female cases. XCI testing was performed revealing random patterns in blood, with the exception of the largest-deletion case in which skewed inactivation (>90%) of the deletion-bearing paternal X-chromosome was confirmed. Two of the described deletions are unique in their genetic content as one is a hemizygous deletion spanning more genes in addition to *PLP1* than has been observed in reported males thus far, and another is a heterozygous deletion encompassing an additional disease-causing gene, *PIH1D3*, in an affected female. Breakpoint-junction sequencing and examination of genomic architecture via a novel approach of identifying low copy repeats (LCRs) and repetitive elements revealed an overrepresentation of deletion breakpoints that lie within these classes of repeat structures/genome architectures, specifically LCRs. Although previous studies suggested a mechanism of non-homologous end joining, genomic characteristics of the majority of these *PLP1* deletions reveal micro-homology at the breakpoints, as observed for *PLP1* copy-number gain events, and suggest the potential involvement of replicative mechanisms and template switching in their formation.

1091F

The morbidity of neurofibromatosis 1: Rates and nature of hospitalizations in a population-based cohort in Denmark. J.J. Mulvihill¹, L. Kenborg², A.K. Duun-Heniksen², V.M. Riccardi³, K. Rugbjerg², C. Pedersen², K. Doser², A. Kroyer², P.E. Bidstrup², S.O. Dalton², C. Johansen², K.K. Andersen², J.R. Ostergaard⁴, H. Hove⁵, S.A. Sorensen⁶, J.F. Winther^{2,7}. 1) Pediatrics, University of Oklahoma, Oklahoma City, OK; 2) Danish Cancer Society Research Center, Copenhagen, DK; 3) The Neurofibromatosis Institute, La Crescenta, CA; 4) Pediatrics, Centre for Rare Diseases, Aarhus University Hospital, Aarhus, DK; 5) Clinical Genetics, Rigshospitalet University Hospital, Copenhagen, DK; 6) Neurogenetics, Institute of Cellular and Molecular Medicine, The Panum Institute, University of Copenhagen, Copenhagen, DK; 7) Clinical Medicine, Aarhus University, Aarhus, DK.

Neurofibromatosis 1 (NF1) is a Mendelian disorder with increased mortality, predisposition to certain cancers, neurocognitive and neurologic manifestations, and, according to case series, occasional manifestations in nearly every organ system. To quantitate the disease burden to assay the full morbidity, we analysed the rates and nature of hospitalizations for individuals NF1 in all of Denmark. We matched all 2,515 persons with NF1 in the Danish National Patient Register (DNPR) (1977–2013) and two national Centres for Rare Diseases (1995–2013) to 24,629 general population comparisons from the Danish Civil Registration System. International Classification of Diseases versions 8 and 10 identified patients and allowed careful exclusion of neurofibromatosis 2, but gene testing was not systemically done. Based on hospital discharge diagnoses from the DNPR, hospitalization rates for non-psychiatric diagnoses were used to calculate a rate ratio (RR) and an absolute excess risk (AER). In Cox proportional hazard models, a hazard ratio (HR) was estimated for 12 main diagnostic groups according to age and cancer status. The RR for a first hospital admission among NF1 persons in 146 selected diagnostic categories was 1.81 (95% confidence interval 1.73–1.90), in another words, nearly double rate in the general population. The overall AER of 304 per 10,000 person-years (95% CI, 274–334) was high in all 12 diagnostic groups; 64% of the excess was accounted for by just five groups [benign neoplasms (14%), malignant neoplasms (14%), disorders of the central nervous system and sense organs (13%), the respiratory (11%) and the digestive systems (11%)]. An increased HR for any hospitalization was observed for all age groups, with or without an associated cancer, and highest for children. Individuals with NF1 spent twice as many days in hospital as did the general population, with an overall mean of 9.5 versus 4 hospitalizations by age 60 years; children under 10 years averaged 3.5 hospitalizations compared to 1 in the population. In this population-based, nation-wide cohort, NF1 persons had an overall greater likelihood of hospitalization, with more frequent and longer hospitalizations involving problems in every organ system, and the highest burden in children.

1092W

Disease-associated sites in *PCDH19* cluster in between cadherin domains depleted from the general population. A. Roctus¹, S. Iqbal^{2,3}, E. Pérez-Palma⁴, L. Smith¹, D. Lal^{2,3,4,5}, A. Poduri¹. 1) Epilepsy Genetics Program, Department of Neurology, Boston Children's Hospital and Harvard Medical School, Boston, MA; 2) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA; 3) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 4) Cologne Center for Genomics, University of Cologne, Cologne, Germany; 5) Cleveland Clinic, Cleveland, OH.

Rationale: Protocadherin-19 (*PCDH19*) encodes a transmembrane protein with six cadherin domains, which creates an extensive adhesive interface. Over 100 pathogenic variants have been identified in patients with a female-restricted form of infantile-onset epilepsy. In contrast to typical X-linked diseases, hemizygous males are spared. Our aim was to characterize structural determinants associated with *PCDH19*-related epilepsy variants. **Methods:** Variants were ascertained from the literature with the search parameters "PCDH19" AND "Epilepsy" (PubMed). Subsequently, we assembled a 3-dimensional structural model of the *PCDH19* protein, mapped all reported missense *PCDH19* variants on the model, and compared the spatial distribution of patients' variants to those reported in the population database (gnomAD). Furthermore, we determined whether the patient variants affect buried amino acids or amino acids at the surface. Finally, we performed an enrichment analysis comparing amino acid positions with patient vs. population missense variants with respect to the predicted functional domains of the protein model, such as the calcium-binding sites. **Results:** We mapped 106 missense variants reported in literature to 77 locations on the 3-dimensional protein model. We show that the missense patient variants are present in distinct regions from population variants ($p < 0.0001$, Fisher's exact test). Moreover, 47/77 (61%) of the missense patient variants are enriched in regions where control individuals from gnomAD show a paucity of variants ($p < 0.0001$, Fisher's exact test). Among those variants, 62% (29/47) are enriched in calcium binding sites in between the cadherin domains ($p = 0.038$, Fisher's exact test). **Conclusions:** Mapping all known disease-associated variants on a computationally derived structural model of *PCDH19* reveals that missense variants in *PCDH19*-related epilepsy cluster at regions distinct from those where general population variants are located, specifically, between the cadherin domains at the adhesive interface of the protein. The present results provide further support to the idea that it is the abnormal binding of these cadherin domains between cells that mediate at least some of the neurological dysfunction seen in patients. We thus demonstrate that mapping genetic variants onto a 3-dimensional protein structure can facilitate variant interpretation and also provide molecular insight into the pathogenesis of disease.

1093T

Genetic analysis of SMN in SMA pedigrees. F. Song, YY. Cao, YJ. Qu, JL. Bai, MM. Cheng. Capital Institute of Pediatrics, Beijing, China.

Spinal muscular atrophy (SMA) pathogenic gene, *SMN1* and modified gene, *SMN2* co-locate at 5q13. This region is complex and many repeats are there, easily leading to deletion or gene conversion. *SMN1* homozygous deletion and compound heterozygous mutation can lead to SMA. Currently the SMA genetic studies focused on disease population, but no family studies were reported. In this study, we aim to identify the distribution of *SMN* in SMA pedigrees. One hundred and fifty-three SMA pedigrees, which were got the genetic diagnosis from the department of genetics in Capital Institute of Pediatrics, were recruited. Multiplex Ligation-dependent Probe Amplification (MLPA) was used to analyze copy numbers of the *SMN1* and *SMN2*. The gene conversion was determined by the copy numbers of exon 7 and exon 8 of *SMN1* and *SMN2*. In 153 SMA pedigrees, 146 (95.4%) probands were homozygous deletion of *SMN1*, and 7 (4.6%) probands carried *SMN1* compound heterozygous mutations. Both parents with *SMN1* heterozygous deletion were found in 128 families (Classic deletion), while only one parent with *SMN1* heterozygous deletion were found in 18 families (2+0), whose children were all *SMN1* homozygous deletion. Moreover, in 7 pedigrees with *SMN1* compound heterozygous mutations, both parents carried *SMN1* heterozygous deletions and gene variations, respectively. Conversion analysis showed that in 21/128 (16.4%) pedigrees, at least one family member had SMN conversion; three families (3/18, 16.7%) had SMN conversion in 18 pedigree with 2+0 type; no SMN conversion were found in 7 pedigree with *SMN1* compound heterozygous mutations. Therefore, 15.7% (24/153) of SMA pedigrees had SMN conversion. Moreover, among 306 SMA carriers from 153 SMA pedigrees, there were 25 cases (2+0 and point variant) could not be detected by MLPA, leading to false negative rate of 8.9% (25/306). Therefore, individuals with type of 2+0 and point variant of *SMN* should be considered in SMA carrier screening; on the other hand, SMN conversion is still a factor of SMA, which is not be ignored. **Acknowledgments** This study was supported by the National Key Research and Development Program of China (2016YFC0901505), National Natural Science Foundation of China (81500979, 81501850), the Natural Science Foundation of Beijing Municipality (5163028, 7172039), CAMS Initiative for Innovative Medicine (CAMS-I2M), Central research institutes of basic research and public service special operations (2016ZX310182-6).

1094F

Variant pathogenicity evaluation through the community-driven Inherited Neuropathy Variant Browser. *D. Bis¹, C. Saghira¹, D. Stanek², A. Strickland¹, D.N. Herrmann³, M.M. Reilly⁴, S.S. Scherer⁵, M.E. Shy⁶, S. Zuchner¹.* 1) Department of Human Genetics and Hussman Institute of Human Genomics, University of Miami, Miami, USA; 2) DNA Laboratory, Department of Paediatric Neurology, 2nd Faculty of Medicine, Charles University in Prague and University Hospital Motol, Prague, Czech Republic; 3) Department of Neurology, University of Rochester, Rochester, New York, USA; 4) MRC Centre for Neuromuscular Diseases, UCL Institute of Neurology, Queen Square, London, UK; 5) Department of Neurology, University of Pennsylvania, Philadelphia, Pennsylvania, USA; 6) Department of Neurology, University of Iowa, Iowa City, Iowa, USA.

Charcot-Marie-Tooth disease (CMT) is an umbrella term for inherited neuropathies affecting an estimated 1 in 2500 people. Over 120 CMT and related genes have been identified and clinical gene panels often contain more than 100 genes. Such a large genomic space will invariably yield variants of uncertain clinical significance (VUS) in nearly any person tested. This rise in number of VUS creates major challenges for genetic counseling. Additionally, fewer individual variants in known genes are being published as the academic merit is decreasing, and most testing now happens in clinical laboratories, which typically do not correlate their variants with clinical phenotypes. For CMT, we aim to encourage and facilitate the global capture of variant data to gain a large collection of alleles in CMT genes, ideally in conjunction with phenotypic information. The Inherited Neuropathy Variant Browser provides user-friendly open access to currently reported variation in CMT genes. Geneticists, physicians, and genetic counselors can enter variants detected by clinical tests or in research studies in addition to genetic variation gathered from published literature, which are then submitted to ClinVar bi-annually. Active participation of the broader CMT community will provide an advance over existing resources for interpretation of CMT genetic variation.

1095W

Improvements in arginase 1 deficiency-related disease manifestations following plasma arginine reduction with pegzilarginase: Early phase 2 results. *R.T. Zori¹, G.A. Diaz², J.L. Merritt II³, A. Schulze⁴, G.M. Enns⁵, M.C. McNutt⁶, E.L. Teles⁷, K.C. Patki⁸, J.E. Wooldridge⁹, S.P. Batzios⁹.* 1) University of Florida, Gainesville, FL, USA; 2) Icahn School of Medicine at Mt. Sinai, NY, USA; 3) University of Washington, Seattle, WA, USA; 4) University of Toronto and The Hospital for Sick Children, Toronto, ON, Canada; 5) Stanford University, Stanford, CA, USA; 6) UT Southwestern Medical Center, Dallas, TX, USA; 7) Centro Hospitalar de São João, Porto, Portugal; 8) Aeglea BioTherapeutics, Austin, TX, USA; 9) Great Ormond Street Hospital NHS Trust, London, UK.

Background: Arginase 1 Deficiency (ARG1-D) is a rare, progressive, autosomal recessive disease characterized by hyperargininemia that causes spasticity, impaired mobility, and developmental delay, typically presenting in early childhood. Current methods to reduce plasma arginine are inadequate and problematic. Pegzilarginase is a pegylated, cobalt-substituted human arginase 1 that has been shown to lower plasma arginine in patients with ARG1-D. **Methods:** Data are from an ongoing Phase 1/2 clinical trial and Phase 2 Open Label Extension study of intravenously (IV) administered pegzilarginase in patients with ARG1-D. In addition to evaluations of safety, pharmacokinetics and effects on plasma arginine and related guanidino compounds (GCs), clinical outcomes assessments were incorporated at baseline and follow-up, including 6-minute walk test (6MWT), Berg Balance Scale (BBS), and Gross Motor Function Measure (GMFM-66). Safety assessments include physical examinations, laboratory tests, anti-drug antibody (ADA) evaluations and ECGs. **Results:** All enrolled patients had disease manifestations including hyperargininemia, neuromotor, and neurocognitive deficits. IV pegzilarginase, added to standard disease management, rapidly lowered plasma arginine into the normal range in all patients and was well tolerated except for a single infusion associated reaction in one patient. This infusion reaction was managed with dose interruption, medication, and a slower infusion rate without further adverse events. Sustained reductions in arginine, GCs, and improvements in several clinical tests of neuromotor function were observed in the first two patients during repeat-dose treatment. Dietary protein intake also increased in these two patients. **Discussion:** These data from the Phase 1/2 clinical trial and Phase 2 Open Label Extension study demonstrated the first evidence of clinically relevant treatment effects in neuromotor function with pegzilarginase beyond what can be achieved with standard disease management. Pegzilarginase produced marked and sustained reduction of plasma arginine and GCs in patients with ARG1-D. Additional insights are expected from recently enrolled adult and pediatric patients, as well as from longer term dosing in previously enrolled patients. .

1096T

Genetic causes of epilepsy with migrating focal seizures in infancy in a new large cohort of patients. G. Barcia^{1,2}, S. Gobin¹, N. Chermaly³, M. Kuchenbuch⁴, E. Azouguene¹, Z. Assouline¹, C. Nava¹¹, V. Ciorna⁵, A. Macaya⁵, L. Lambert⁶, D. Doummar⁷, T. Billette de Villemeur⁸, F. Dubois-Teklal⁹, J. Lefranc¹², N. Bahi-Buisson³, I. Desguerre³, A. Munnich^{1,2}, J. Steffann^{1,2}, A. Munnich^{1,2}, N. Boddaert^{10,2}, M. Milh¹³, S. Auvin¹⁴, J.P. Bonnefont^{1,2}, R. Nabbout^{3,2}. 1) Department of medical genetics, Hôpital Necker Enfants Malades, APHP, Paris, France; 2) INSERM1163 Imagine Institute, Université Paris Descartes, Paris, France; 3) Department of pediatric neurology, Hôpital Necker Enfants Malades, APHP, Paris, France; 4) Department of clinical neurophysiology and pediatrics, INSERM1099, Hôpital Universitaire de Rennes, France; 5) Pediatric neurology research group, Val d'Hebron Research Institute, Univeritat Autònoma de Barcelona, Barcelona, Spain; 6) Department of medical genetics, Hôpital Femme Mère Enfant, Metz-Thionville, France; 7) Neuropediatrics Unit, Hôpital Trousseau, Paris, France; 8) Department of Clinical Genetics, Hôpital d'Enfants, CHU Nancy, Nancy, France; 9) Neuropediatrics Unit, CHU de Grenoble, Grenoble, France; 10) Neuroradiology Unit, Hôpital Necker Enfants Malades, APHP, Paris, France; 11) Inserm U1127, CNRS UMR 7225, Institut du Cerveau et de la Moëlle Épine, Paris, France, AP-HP, GH Pitié-Salpêtrière, Department of Genetics, Paris, France; 12) Pediatrics and Medical Genetics Department, CHU Brest, Brest, France; 13) Child Neurology Unit, INSERM GMGF UMR S910 Marseille, France; 14) Child neurology Unit, Hôpital Robert Debré, APHP Paris, France.

Introduction: Epilepsy with migrating focal seizures in infancy (EMFSI) is a severe epileptic syndrome characterized by focal seizures beginning in the first months of life, migrating from one cortical region to another, highly pharmac-resistant, and associated with a severe cognitive outcome. In about 50% of patients, EMFSI is due to de novo KCNT1 gain-of-function mutations. In a few patients, EMFSI is linked to mutations in different genes (SCN1A, SMC1A, GABRB3, SCN8A, SCN2A, QARS, SLC25A22, SLC12A5, TBC1D24, PLCB1) according to de novo or autosomal recessive inheritance models. **Methods:** We reviewed clinical and molecular data of 26 new patients with EMFSI. We analyzed KCNT1 by Sanger sequencing and performed targeted next generation sequencing (NGS) analysis in patients without KCNT1 mutations. **Results:** We identified a causal mutation in 23/26 patients (88%). Fourteen patients (54%) carried missense KCNT1 mutations. The majority of KCNT1 mutations lies on already known hotspots. All occurred de novo, except for p.Arg398Gln which was inherited from a father affected by autosomal dominant nocturnal frontal lobe epilepsy (ADNFLE) and present in two paternal uncles, one being asymptomatic and the other having presented a single tonic-clonic seizure. In 9 patients (35%), mutations were identified in different genes: PIGA (n:2), SCN8A (n:2), KCNA1 (n:1), KCNC1 (n:1), CHRNA2 (n:1), ALDH7A1 (n:1), QARS (n:1). PIGA mutations were associated with cerebellar atrophy and with a better response of seizures to ketogenic diet; QARS mutations with a severe ponto-cerebellar hypoplasia. **Conclusions:** We confirm in a new large cohort that KCNT1 mutations are the major cause of EMFSI, with mutations in 54% of patients. The majority of KCNT1 mutations lies on specific hotspots. No genotype-phenotype correlation is possible, since a same mutation can be linked to a spectrum of conditions ranging from EMFSI to asymptomatic carrier, even in the same family. Moreover, these results underline that EMFSI is a genetically heterogeneous condition since we identified pathogenic variants affecting different genes in 35% of patients. In particular, we described for the first time PIGA, KCNA1, KCNC1, CHRNA2, and ALDH7A1 mutations in EMFSI. These results underline that targeted NGS is an appropriated strategy in patients with EMFSI. Moreover, they have relevant implication for antiepileptic treatment and genetic counseling.

1097F

A novel variation in X-linked inherited IQSEC2 from a three-generation Korean family with Lennox-Gastaut syndrome and mental retardation. M. Choi^{1,2}, J. Min¹, J. Lee¹, S. Jun¹, Y. Lee¹, J. Yoon¹, S. Jeon¹, I. Byeon³, J. Kang³, J. Yang⁴, N. Kim^{1,2}. 1) Genomics Structure Research Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon, South Korea; 2) Department of Functional Genomics, Korea University of Science and Technology (UST), Daejeon, 34141, Republic of Korea; 3) Department of Pediatrics and Medical science, Chungnam National University Hospital, College of Medicine, Chungnam National University, Daejeon, 35015, Republic of Korea; 4) Korean BioInformation Center (KOBIC), Korea Research Institute of Bioscience and Biotechnology (KRIBB), Daejeon, 34141, Republic of Korea.

Lennox-Gastaut syndrome (LGS [MIM 606369]) is a childhood onset epilepsy that makes up 10% of childhood epilepsies. Of LGS patients, 80% also have intellectual disability. To identify causative genes related to LGS, we collected data from a three-generation Korean family with LGS, in which one family members had LGS, two had mental retardation, and two were normal. LGS patients have features such as generalized slow spike and wave discharges in electroencephalography, multiple types of seizures, and developmental delay. Analyses of whole-exome data from the family revealed a mutation of guanine to adenine on nucleotide 1048 of *IQSEC2* [MIM 300522] (c.1048G>A). This variation changes a non-polar hydrophobic alanine residue to a polar uncharged threonine at residue 350 (p.A350T). This variation is within highly evolutionarily conserved IQ-like motif, indicating decrease of calmodulin-binding capacity or AMPA transmission. The hemizygous variation in the male proband was a maternally inherited X-linked variation from heterozygosity in the maternal grandmother and mother. These findings indicate that the variation of X-linked *IQSEC2* triggers both LGS and mental retardation dependent on sex and variation type. We first report a novel variation of X-linked inherited *IQSEC2* for LGS and mental retardation that could be a useful marker to diagnose LGS and mental retardation.

1098W

Exome sequencing identifies molecular diagnosis in children with drug-resistant epilepsy that associates with risk of sudden death. B. Chung, M. Tsang. The University of Hong Kong, Hong Kong, NA, China.

Objective: Sudden unexpected death in epilepsy (SUDEP) is relatively uncommon in children but the risk increases and persists into adulthood. This life-threatening condition is strongly related to drug-resistant epilepsy. We aim to identify the underlying genetic causes of a children cohort with drug-resistant epilepsy, and whether the findings can provide information to stratify the risk of sudden death. **Methods:** We include subjects with drug-resistant epilepsy with onset before 18 years old managed in the Department of Paediatrics & Adolescent Medicine, the University of Hong Kong. Clinical chromosomal microarray (CMA) and then whole exome sequencing (WES) was performed using genomic DNA from the subjects. In the first-tier analysis of the exome data, we aimed to identify disease-causing mutations in 546 genes known to cause epilepsy. For negative cases, we proceeded to exome-wide analysis. Rare coding variants were interrogated for pathogenicity based on the American College of Medical Genetics and Genomics (ACMG) guideline. **Results:** We recruited 50 subjects (male=22). We identified six pathogenic or likely pathogenic mutations, giving a diagnostic yield of 12% (6/50). Mutations were found in six different genes: *SCN8A*, *SCN1A*, *MECP2*, *CDKL5*, *DEPDC5*, and *CHD2*. The *CDKL5* variant was found to be a mosaic mutation. One variant of unknown significance (VUS) in *KCNT1* was found in a subject with compatible clinical features. Four variants were *de novo*, and three variants were inherited from father. Interestingly, an incidental finding of pathogenic *SCN5A* mutation contributed to Brugada syndrome, was also noted in the subject with *SCN1A* mutation. **Significance:** Genetic diagnosis can help to consolidate the clinical diagnosis, to facilitate phenotypic expansion, and to influence management options for seizure control in our patients. Importantly, in our study, a significant portion of the genetic findings are known to be associated with an increased risk of SUDEP. Since the genetic variants can be inherited, the risk of SUDEP can go beyond the patients to family members without epilepsy but carry the variants. This information is important for the comprehensive genetic counselling of these families.

1099T

A candidate gene for progressive myoclonic epilepsy in a Malian family.

M.E. Dembélé¹, L. Cissé¹, S.H. Diallo^{2,3}, T. Coulibaly^{1,2}, S. Diallo³, S. Diarra⁴, F. Kané¹, A. Simaga^{2,5}, C. Grunseich⁴, M. Keita^{2,6}, K.H. Fischbeck⁴, C.O. Guinto^{1,2}, Y. Maiga^{2,3}, G. Landouré^{1,2,4}, The H3Africa Consortium. 1) Service de Neurologie, CHU du Point G, Bamako, Mali; 2) Faculté de Médecine et d'Odontostomatologie, USTTB, Bamako, Mali; 3) Service de Neurologie, CHU de Gabriel Touré, Bamako, Mali; 4) Neurogenetics Branch, NINDS, NIH, Bethesda, MD; 5) Institut d'Ophtalmologie Tropicale de l'Afrique, Bamako, Mali; 6) Service d'ORL, CHU de Gabriel Touré, Bamako, Mali.

Introduction: Progressive myoclonic epilepsy (PME) is a group of disorders characterized by seizures and other neurological and non-neurological features. Several genes and loci have been associated to the disease; however, reports in sub-Saharan Africa are rare. **Objectives:** Characterize patients with PME, and identify the underlying genetic defect. **Methods:** After consent, patients went through a thorough clinical examination. Electroencephalography (EEG), brain imaging and blood chemistries were done to consolidate our diagnosis. DNA was extracted for genetic analysis. **Results:** We identified a consanguineous family in which three siblings presented symptoms consistent with PME at around age 7. Symptoms started with hand tremor followed by jerk movement and difficulty with walking. Symptoms worsened overtime, and in addition they presented problems with speech and progressive cognitive decline. Neurological examination found cerebellar ataxia, and hearing and visual impairment. EEG showed generalized spike-waves and slow basal rhythm and photosensitivity in both patients, and brain imaging was consistent with cerebellar atrophy. In addition, unusual features including extrapyramidal symptoms were noticed. Epilepsy gene panel testing (126 genes + mtDNA) was negative. Exome sequencing in two affected individuals has showed three variants in three different genes. These variants are not present in SNP databases including dbSNP, Exac and 1000Genomes. None of these genes have been previously involved in any disease, and two of them are highly expressed in the brain. Of latter, one has a role in the axonal function while SNPs in the other were associated with neurodegenerative disorders. In addition, the mutated residues were conserved through a wide range of species. **Conclusion:** Our findings suggest two novel candidate genes causing PME. Further studies are underway to evaluate these variants in extended family members and the Malian population to select the putative variant that will be studied in cell culture.

1100F

Pyridoxal phosphate-dependent neonatal epileptic encephalopathy:

Clinical case. T.V. Kozhanova^{1,2}, S.S. Zhilina^{1,2}, T.V. Mescheryakova¹, E.G. Luk'yanova¹, K.V. Osipova¹, S.O. Ayvazyan¹, A.G. Prityko^{1,2}. 1) St.Luka Scientific and Practical Center of Specialized Medical Care for Children, Department of Health, Moscow; 2) Pirogov Russian National Research Medical University, Moscow, Russia.

Pyridoxine-dependent epilepsy (OMIM # 266100) is severe epileptic encephalopathy with an autosomal recessive type of inheritance, characterized by neonatal seizures resistant to traditional antiepileptic therapy, and a therapeutic response to the administration of pharmacological doses of vitamin B6. Pyridoxine-dependent epilepsy is caused by homozygous or compound heterozygous mutations in *ALDH7A1* gene. *ALDH7A1* gene is mapped on chromosome 5q31 and consists of 18 coding regions (exons). The article presents a clinical case of a patient with severe infantile generalized idiopathic epilepsy with status seizures, muscular dystonia syndrome and development delay. The observation included phenotypic analysis: the nature of seizures, cognitive and behavioral disorders, video electroencephalography, magnetic resonance imaging of the brain, features of the perinatal period, and targeted exome sequencing of the genes associated with epileptic encephalopathy (NGS). The previously not described nucleotide heterozygous variants in the *ALDH7A1* gene, which led to stop codon in the 82 position of the protein (p.Arg82Ter) and amino acid substitution in the 399 position of the protein (p.Glu399Gln) were detected by NGS. This clinical observation demonstrates the importance of DNA diagnostics by using targeted exome sequencing to find molecular defects, especially in severe neonatal drug-resistant seizures. The patient should be given therapeutic doses of vitamin B6 for relief of seizures in the case of determining mutations in the *ALDH7A1* gene.

1101W

Infantile catastrophic epilepsy caused by a *de novo* tryptophan to arginine mutation in a transmembrane region of ATP1A1.

A. Lindstrand^{1,2}, E.E. Akkuratov³, D.C. Jans³, F. Taylan¹, P.F. Kinoshita⁴, I. Nennesmo⁵, M. Andersson⁶, S.J.D. Karlsh⁷, H. Brismar^{3,8}, A. Aperia⁹, S. Ygberg^{9,10}. 1) Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden; 2) Department of Clinical Genetics, Karolinska University Hospital, Stockholm, Sweden; 3) Science for Life Laboratory, Department of Applied Physics, Royal Institute of Technology, Stockholm, Sweden; 4) Molecular Neuropharmacology Laboratory, Department of Pharmacology, Institute of Biomedical Science, University of São Paulo, São Paulo, São Paulo, Brazil; 5) Department of Clinical Pathology and Cytology, Karolinska University Hospital, Stockholm, Sweden; 6) Science for Life Laboratory, Department of Physics and Swedish e-Science Research Center, KTH Royal Institute of Technology, Stockholm, Sweden; 7) Department of Biomolecular Sciences, Weizmann Institute of Science, Rehovot 7610001, Israel; 8) Science for Life Laboratory, Department of Women and Children's Health, Karolinska Institutet, Stockholm, Sweden; 9) Neuropediatric Unit, Karolinska University Hospital, Stockholm, Sweden; 10) Centre for Inherited Metabolic Diseases (CMMS), Karolinska University Hospital, Stockholm, Sweden.

Epileptic encephalopathies are severe brain disorders that result in developmental delay and sometimes early death. To date, more than a hundred genes have been implicated, many of which encode ion channels. Here, we report a *de novo* variant (c.2791T>C; p.W931R) in exon 20 of *ATP1A1* in a girl with treatment refractory epilepsy and profound hypo-magnesemia who died at 10 months from a respiratory infection. *ATP1A1* encodes the ubiquitous alpha1 subunit of Na⁺/K⁺-ATPase, the ion pump that creates the sodium and potassium gradients across the cell membrane and is responsible for generating the membrane potential. It uses more than 50% of brain's energy consumption. No previous links between *ATP1A1* variants and epilepsy have been described. However, mutations have been identified in the Na⁺/K⁺-ATPase alpha3 subunit (*ATP1A3*) in two children with catastrophic early life epilepsy. In the affected child, autopsy revealed a partial loss of primarily hippocampal neurons and by high resolution confocal microscopy on the autopsy material indicates that ATP1A1 was still expressed in the membrane of the remaining hippocampal neurons. Next, we showed that expression of the W931R alpha 1 in HEK cells led to significantly decreased viability of the cells. Since W931 is located in a trans-membrane region, we then used molecular dynamics simulations to characterize possible differences in interactions to non-specific bulk lipids. In this way, W931R alpha1 was found to induce asymmetric protein-lipid interactions and to accumulate water in pockets between the mutant residue and the sodium-binding sites. The findings indicate that the mutant alpha1 disturbs the hydrophobic core of the membrane and is less favorably anchored to its lipid environment. Finally, imaging of fluorescently tagged wild type and W931R alpha1, expressed in cultured hippocampal neurons and HEK cells, show that W931R Na⁺/K⁺-ATPase is, to a lesser extent than wild type Na⁺/K⁺-ATPase, inserted into the membrane. Taken together, these findings are likely to be the cause of impaired control of membrane potential and neuronal activity. In summary, our data indicate that a unique *de novo* variant in *ATP1A1* is the likely cause of catastrophic early life epilepsy. The exchange of a non-polar hydrophobic residue to an arginine residue in a trans-membrane region may represent a novel disease-causing mechanism in the field of ion pumps and channels.

1102T

Towards the elucidation of the genetic architecture of autosomal dominant lateral temporal epilepsy. C. Nobile¹, E. Dazzo¹, A. Giallonardo², C. Di Bonaventura², P. Pulitano³, O. Mecarelli⁴, A. Gambardella⁴, P. Striano⁵, R. Michelucci⁶, Genetics Commission Italian League Against Epilepsy. 1) CNR-Neuroscience institute, Padova, Italy; 2) Department of Neurological Sciences, University of Rome "Sapienza," Roma, Italy; 3) Department of Neurology and Psychiatry, University of Rome "Sapienza", Roma, Italy; 4) Institute of Neurology, University Magna Graecia, Catanzaro, Italy; 5) Muscular and Neurodegenerative Disease Unit, Institute "G. Gaslini," University of Genoa, Genova, Italy; 6) IRCCS-Institute of Neurological Sciences, Bellaria Hospital, Bologna, Italy.

Autosomal dominant lateral temporal epilepsy (ADLTE [MIM 600512]) is a genetically heterogeneous focal epilepsy characterized by auditory seizures. Disease-causing mutations have been discovered in three genes, *LGI1* (MIM 604619), *RELN* (MIM 600514), and *MICAL-1* (MIM 607129), which altogether account for the disorder in 50-60% of ADLTE families. To fully elucidate the genetic architecture of ADLTE, we investigated eight well-studied ADLTE families without mutations in known genes by combining genome-wide single nucleotide polymorphism-array linkage analysis and whole exome sequencing. Filtering exome-derived rare variants included localization to positive linkage peaks in the corresponding families, gnomAD frequency < 0.001, and gene residual variation intolerance score < 50%. A relatively short list of candidate genes was finally evaluated based on their expression pattern, function, and implication in other epileptic or neurodevelopmental disorders in each affected family. Using this approach, we identified several potentially pathogenic variants that might be implicated in ADLTE. Studies of their functional effects and of additional families will reveal which of them is actually implicated in the aetiology of ADLTE, improving our knowledge of the genetic basis of ADLTE.

1103F

Autophagy and awakening grand mal seizures in JME. M. Tanaka^{1,2}, I.E. Martínez-Juarez³, R.M. Durón⁴, V.H. Nguyen^{1,5}, A. Ochoa³, A. Jara-Prado³, M.E. Alonso³, M. López Ruiz⁶, M.T. Medina⁷, L. Guilloto⁸, E.M. Yacubian⁸, R. Silva⁹, C. Patterson^{1,2}, J.N. Bailey^{1,10}, A.V. Delgado-Escueta^{1,2}. 1) Epilepsy Genetics/Genomics Lab, VAGLHS, Los Angeles, CA; 2) Department of Neurology, David Geffen School of Medicine at University of California, Los Angeles, CA; 3) Instituto de Neurología y Neurocirugía "Manuel Velasco Suarez", Mexico City, Mexico; 4) Facultad de Ciencias de la Salud, Universidad Tecnológica Centroamericana UNITEC, Tegucigalpa, Honduras; 5) Chapman University, Irvine, CA; 6) General Hospital of Mexico, Mexico City, Mexico; 7) Universidad Nacional Autónoma de Honduras, Tegucigalpa, Honduras; 8) Universidad Federal de Sao Paulo (UNIFESP-EPM), Sao Paulo, Brazil; 9) Hospital de Especialidades Nuestra Señora Reina de la Paz, San Miguel, El Salvador; 10) Epidemiology, Fielding School of Public Health, University of California, Los Angeles, CA.

To investigate disease mechanisms in Juvenile Myoclonic Epilepsy, we have studied autophagy for a rare variant found in a 37-member, four-generation Mexican family, who have 6 epilepsy affected members (five with Childhood Absence Epilepsy (CAE) with eyelid myoclonias evolving to JME, one with absence-tonic-clonic grand mal only) and two with only EEG spike waves. An extremely rare variant (R638W, frequency 0.00000921 at ExAC browser) in *IPO8* (*RANBP8*) on 12p11.21, identified by whole exome sequencing, cosegregated with all eight clinical and EEG affected members. R638W variant encodes *Importin 8*, a member of the *Importin β* family. R638W is absent in 777 controls matched for ethnicity/ancestral origin. Five other rare missense mutations of *IPO8*, one rare 5'UTR SNP and one novel 3' UTR deletion of transcript variant2 of *IPO8* is present in nine out of 295 families with CAE/JME. *Importin 8* reportedly transports the transcription factor (TFEB) for autophagy genes to nuclei. Therefore, we have assessed expression levels of TFEB and autophagy related gene products using western blotting. HEK 293 cells were transiently transfected with plasmids carrying mutated human full length *IPO8*. R638W increased the expression of Calcineurin (P=0.006) in the cytoplasm, the nuclear translocation of TFEB, the total expression of TFEB, and the autophagy marker, LC3B (P=0.049) at 10 min after administration of hydrogen peroxide (160uM) and amino acid starvation by EBSS medium. We conclude *IPO8* regulates autophagy and the upregulated autophagy could be involved in photosensitivity and seizure induction during awakening seizures.

1104W

Novel variants causing inherited leukodystrophies in Sudanese families.

M. Amin¹, I. Dorboz², I. Mohamed¹, A. Hamed¹, M. Elaseed¹, A. Yahia¹, A. Babai¹, E. Badi¹, M. Hasan¹, E. Emad¹, S. Samaan², L. Elsayed¹, A. Ahmed¹, O. Boespflug-Tanguy². 1) Faculty of Medicine, University of Khartoum, Khartoum, Sudan; 2) Neuropediatrics and metabolic disorders unit, Hospital Robert Debre, Paris, France.

Background: Leukodystrophies are a group of heritable metabolic disorders resulting from defective myelination of the central nervous system. There are currently dozens known different types of leukodystrophies, all with distinct clinical and radiological characteristics. However, the genetics of leukodystrophies remain unknown in many parts of the world especially in Sub-Saharan Africa. **Methods:** In this study, we recruited three consanguineous families from Sudan: one with Metachromatic leukodystrophy (MLD) and two with Megalencephalic leukodystrophy with sub-cortical cysts (MLC) with typical clinical and radiological signs. Genomic DNA was extracted and screened for mutations in all leukodystrophy causing genes. **Results:** Three novel variants were discovered: one (c.380G>C, p.Arg127Pro) in *PSAP* gene causing MLD, and two variants (c.831_838DUPATATCTGT, p.Ser280Tyrfs*8 and c.971T>G, p.Ile324Ser) in *MLC1* gene in the two families with MLC. The segregation pattern was consistent with autosomal recessive inheritance. The pathogenicity of these variants was predicted using bioinformatics tools. **Conclusion:** This is the first study to underlie the genetics of leukodystrophy in Sudan. However, more families are needed to establish the whole spectrum of genetic variations causing inherited leukodystrophies in Sudanese families.

1105T

Effects of long-term eteplirsen treatment on upper limb function in patients with Duchenne muscular dystrophy: Findings of two phase 2 clinical trials. L.N. Alfano¹, K. Berry¹, J.R. Mendell¹, H. Eliopoulos², L. Han², L.P. Lowes². 1) Nationwide Children's Hospital, Columbus, OH; 2) Sarepta Therapeutics, Inc., Cambridge, MA.

Introduction: Duchenne muscular dystrophy (DMD) is a progressive, degenerative neuromuscular disease caused by gene mutations that result in the absence or deficiency of dystrophin protein, which causes muscle deterioration and the loss of upper limb function, negatively impacting quality of life. Eteplirsen (Exondys 51; Sarepta Therapeutics, Inc., Cambridge, MA), the first treatment approved for DMD, is indicated for treatment of patients with DMD amenable to exon 51 skipping. **Methods:** We evaluated long-term upper limb function outcomes in patients aged 7-13 years with DMD amenable to exon 51 skipping who participated in 2 consecutive clinical trials of eteplirsen. In Study 201, a 28-week trial, patients were randomized to receive once-weekly intravenous infusions of eteplirsen (30 mg/kg or 50 mg/kg) or placebo for 24 weeks. At Week 25, all patients received open-label eteplirsen through Week 28. In Study 202, an open-label extension trial, all patients received once-weekly eteplirsen at the same dose as in Study 201 through 240 weeks. Upper limb function tests included the 9-Hole Peg Test (seconds) and a maximum voluntary isometric contraction test with a quantitative movement assessment system to assess hand grip and elbow extension and flexion (kg). **Results:** Study 201 enrolled 12 patients who received once-weekly eteplirsen (30 mg/kg, n=4; 50 mg/kg, n=4) or placebo (n=4); all patients completed extension Study 202. Following the confirmation of dystrophin production, all patients showed relative stability or minimal decline of upper limb function on all measures assessed through week 240. Specifically, mean values at Week 240 for functional assessments of the dominant hand/arm were 22.0 seconds (9-Hole Peg Test), 7.68 kg (hand grip strength), 2.62 kg (elbow extension), and 1.53 kg (elbow flexion). **Conclusion:** Based on published data from healthy reference individuals, the mean 9-Hole Peg Test completion time of 19.9 seconds for boys aged 8-9 years decreases steadily with age to 18.0 seconds for boys aged 14-15 years. As expected, mean completion times are lower in natural history studies of patients with DMD and typically decrease with age and disease progression. In the current report, patients treated with eteplirsen for up to 240 weeks had a mean 9-Hole Peg Test completion time of 22.0 seconds. Based on these and other functional assessments, eteplirsen may slow the deterioration of upper arm function in patients with DMD amenable to exon 51 skipping.

1106F

CRISPR/Cas9 preclinical studies in patient-derived *FOXC1* mutated cells.

S. Croci¹, S. Daga¹, F.C. Lorenzetti¹, F.T. Papa¹, F. Donati², C. Lo Rizzo³, D. Lopercolo^{1,2}, M. Doria⁴, A. Auricchio⁴, S. Conticello³, A. Renieri^{1,2}, I. Meloni¹. 1) Medical Genetics Laboratory, University of Siena, Siena, Italy; 2) Genetica Medica, Azienda Ospedaliera Universitaria Senese, Siena, Italy; 3) Molecular Mechanisms of Oncogenesis, ITT Core Research Laboratory (CRL), Firenze, Italy; 4) TIGEM (Telethon Institute of Genetics and Medicine), Napoli, Italy.

We first identified *FOXC1* gene as associated to disease in humans (Ariani F et al Am J Hum Genet 2008): the congenital variant of Rett syndrome, characterized by early developmental delay, seizures, lack of speech and inability to control body movements. A therapy for this devastating disorder is still lacking. We present here a successful example of transformative gene-based medicines using CRISPR/Cas9 gene editing technology and our previously established cellular model (Patriarchi T et al 2016 Eur J Hum Genet). We have engineered a two-plasmid system to correct specific *FOXC1* mutations, namely c.688C>T - p.Arg230Cys and c.765G>A - p.Trp255*. Plasmid constructs have been designed and produced for both mutations. Specific sgRNAs and donor DNAs have been selected and cloned into an appropriate plasmid with an mCherry/GFP reporter system. Cas9 flanked by mutation-specific sgRNA recognition sequences for auto-cleaving has been cloned in a second plasmid. Fibroblasts transfected with both constructs showed that cells correctly acquire them and produce Cas9. Since intravenous administration of AAV9 particles has been recently used in a phase I clinical trial for gene replacement delivery in patients affected by Spinal Muscular Atrophy type 1, with no relevant adverse effects, we decided to use the same delivery system. Cell types relevant to the project (fibroblasts, iPSCs, iPSC-derived neuronal precursors and neurons) have been infected with AAVs to select the most appropriate serotype. AAV9 infection in terminally differentiated iPSC-derived neurons demonstrated that these cells can be efficiently co-infected. Finally, mCherry+/GFP+ cells isolated by Fluorescent Activated Cell Sorting and analyzed by Next Generation Sequencing demonstrated high efficiency of gene editing. Further steps will include the ongoing mouse model and patient recruitment. Since the identification of *FOXC1* we have been following 16 *FOXC1*-mutated patients. Five families have already expressed their willingness to participate to the proposed research study and are eager to be involved in clinical trials. They include 2 male and 3 female patients with an age range 3-30 yrs (mean age of 12 yrs). The grant "a glimpse of hope for *FOXC1*" from Blackswan Foundation is acknowledged.

1107W

Exploring the potential of CRISPR/Cas9 for the treatment of Pelizaeus-Merzbacher disease.

E. Maino^{1,2}, E. Hyatt¹, E.A. Ivakine¹, R.D. Cohn^{1,2,3}. 1) Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Canada; 2) Department of Molecular Genetics, University of Toronto, Canada; 3) Department of Pediatrics, University of Toronto and The Hospital for Sick Children, Toronto, Canada.

Pelizaeus Merzbacher Disease (PMD) is an X-linked pediatric leukodystrophy that leads to limited quality of life and premature death of affected children. Currently, there is no curative treatment and patient care is restricted to symptom management. PMD is caused by mutations in the *PLP1* gene, which codes for Proteolipid Protein 1, one of the main components of myelin. Mutations in *PLP1* result in impaired myelination of the nervous system which ultimately leads to the onset of the typical PMD symptoms, such as locomotor disabilities and cognitive impairment. Duplications of the X chromosome region containing the *PLP1* gene account for 70% of PMD cases. Here, we propose to develop a clinically relevant CRISPR/Cas9-based genome editing strategy to correct the duplication mutation and to ameliorate the disease phenotypes in a PMD mouse model which has a *Plp1* duplication. First, we characterized this model and we confirmed the presence of dysmyelination and neurological deficits. To remove the *Plp1* duplication we employed a single gRNA-CRISPR/Cas9 approach developed by our group. In proof-of-principle experiments, I validated the single gRNA strategy *in vitro*, using primary cells derived from the PMD mice and *in vivo*, through localized administration of the CRISPR/Cas9 components. Data obtained from these experiments suggest the successful removal of the duplicated region. The preliminary results deriving from this project emphasize the potential of CRISPR/Cas9-based therapies as a future treatment for PMD patients and open a potentially new avenue for the treatment of genetic disorders caused by genomic duplications.

1108T

Research genomics expands the genetic etiology and phenotypic spectrum of rare pediatric conditions. D.C. Koboldt¹, T. Mihalic Mosher¹, B.J. Kelly¹, P. Martin², S.E. Hickey², K.L. McBride², P. White¹, R.K. Wilson¹. 1) Institute for Genomic Medicine, Nationwide Children's Hospital, Columbus, Ohio, USA; 2) Division of Molecular & Human Genetics, Nationwide Children's Hospital, Columbus, Ohio, USA; 3) The Research Institute at Nationwide Children's Hospital, Columbus, Ohio, USA.

Children with rare inherited conditions are increasingly referred for clinical exome sequencing, which yields a positive finding in only ~25-35% of them. For the remaining undiagnosed cases, research sequencing of the proband and available family members has the potential to uncover new genetic etiologies of disease. Our institute has enrolled more than 40 families suffering rare inherited conditions into a research genomics protocol. Using predominantly whole genome sequencing of multiple family members, we have identified likely causal variants in 30% of cases and strong candidate variants in another 20%. Here, we describe three unpublished cases whose underlying etiology or phenotypic association challenges the current knowledge of genotype-phenotype relationships. First, we found that a 6-year-old male with congenital contractures and severe muscular atrophy harbored a *de novo* inframe indel in *BICD2*, a gene in which missense mutations cause a milder form of muscular atrophy restricted to the lower extremities. We established the pathogenicity after meeting the author of a poster at ASHG 2017 describing the same mutation in a 12-year-old female with nearly identical features. Second, whole-genome sequencing uncovered a *de novo* 182-kb inversion disrupting *SGCE* in a male with myoclonic dystonia who was negative for clinical testing. Orthogonal validation assays confirmed the variant's *de novo* status and its breakpoint in the third intron. This finding not only provides a molecular diagnosis to this patient, but also demonstrates the power of WGS to detect pathogenic variants missed by conventional testing. Third, we enrolled a family with multiple pregnancy losses and a child born with severe skeletal dysplasia who died on the first day of life. Sequencing the proband, a fetal demise, and both parents revealed compound heterozygous variants in the gene for galactosyltransferase-I (*B4GALT7*; associated with the spondylodysplastic form of Ehlers-Danlos syndrome). One (p.R270C) was a known pathogenic variant but the other (p.Q133R) had never been reported. Enzymatic activity assays demonstrated that while p.R270C reduces B4GALT7 activity, the novel p.Q133R variant abrogates it entirely, which likely explains the lethal disease manifestation in the family we studied. Taken together, these case studies demonstrate the power of research genomics to provide long-sought molecular diagnoses and further delineate the phenotypic spectrum of rare inherited conditions.

1109F

Mutations in *PIGS* encoding a GPI transamidase protein cause a neurological syndrome ranging from fetal akinesia to epileptic encephalopathy. N.V. Baratang¹, T.T.M. Nguyen¹, Y. Murakami², K.M. Wigby³, J. Rousseau¹, A. St-Denis¹, J.A. Rosenfield⁴, S.C. Laniewski⁵, J. Jones⁶, A.D. Iglesias⁷, M.C. Jones⁸, D. Masser-Frye⁹, A.E. Scheuerle⁹, D.L. Perry¹⁰, R.J. Taft¹⁰, F. Le Deist¹, M. Thompson³, T. Kinoshita², P.M. Campeau¹. 1) CHU Sainte-Justine, Montreal, Quebec, Canada; 2) Research Institute for Microbial Diseases, Osaka University, Osaka 565-0871, Japan; 3) Department of Pediatrics, University of California, San Diego, California, USA, 92093; 4) Department of Molecular & Human Genetics, Baylor College of Medicine, Houston, TX, USA, 77030; 5) University of Rochester Medical Center, NY, USA, 14642; 6) Greenwood Genetic Center, Greenwood, SC, USA 29646; 7) NewYork-Presbyterian Morgan Stanley Children's Hospital, New York, NY 10032; 8) Rady Children's Hospital-San Diego, San Diego, CA, USA 92123; 9) University of Texas, Southwestern Medical Center, Dallas, Texas; 10) Illumina Inc, San Diego, CA 92121, USA.

Inherited GPI deficiencies (IGD) are a subset of congenital disorders of glycosylation that are increasingly recognized due to advances in whole exome (WES)/whole genome (WGS) sequencing. IGDs cause a series of overlapping phenotypes consisting of seizures, dysmorphic features, multiple congenital malformations, and severe intellectual disability. We present a study of six individuals from three unrelated families in which WES or WGS identified bi-allelic phosphatidylinositol glycan class S (*PIGS*) biosynthesis gene mutations. Phenotypes included severe global developmental delay, seizures (partly responding to pyridoxine), hypotonia, weakness, ataxia, and dysmorphic facial features. Two of them have compound heterozygous variants (NM_033198.3) c.108G>A, p.Trp36* and c.101T>C, p.Leu34Pro, and two siblings of another family were homozygous for a deletion and insertion leading to p.Thr439_Lys-451delinsArgLeuLeu. The third family had two fetuses with multiple joint contractures consistent with fetal akinesia. They were compound heterozygous for c.923A>G, p.Glu308Gly, and c.468+1G>C, a splicing mutation. Flow cytometry analyses demonstrated that the individuals with *PIGS* mutations show a GPI-AP deficiency profile. Expression of the p.Trp36* mutation in *PIGS*-deficient HEK293 cells revealed only partial restoration of cell surface GPI-APs. In terms of both biochemistry and phenotype, loss of function of *PIGS* shares features with *PIGT* deficiency and other IGDs. This study contributes to the understanding of the GPI-AP biosynthesis pathway by describing the consequences of *PIGS* disruption in human and extending the family of IGD.

1110W

Clinical whole genome sequencing (WGS): A precision genomic approach to solving “diagnostic odysseys”. S.P. Shankar¹, V. Wheelock², V.T. Triano³, A. Moshiri⁴, M. Hegde⁴, K.A. Rauen¹. 1) Pediatrics, UC Davis, Sacramento, CA; 2) Ophthalmology, UC Davis, Sacramento, CA; 3) Neurology, UC Davis, Sacramento, CA; 4) Perkin Elmer, USA.

The cost of WGS is rapidly declining allowing its use in clinics. We report whole genome sequencing results in 2 male siblings from the UC Davis Precision Genomics Clinic that identified their underlying genetic disorder. The siblings aged 38 and 35 years, presented in early childhood with bilateral sensorineural hearing loss, normal motor and cognitive development until ~14 years of age when they developed progressive tremors, dystonic movements and ataxia. They also have dental enamel issues, and adult onset retinal dystrophy. Significant exam findings include normal cognition, dysarthria, amelogenesis imperfecta, gaze abnormalities, and severe ataxia. Brain MRI revealed pontocerebellar atrophy and signal changes in basal ganglia. Their symptoms are responsive to DOPA and deep brain stimulation (DBS). Extensive work up including muscle biopsy and extensive genetic testing including Fragile X, SCAs 1, 2, 3, 6, 17, Friedreich ataxia, and mitochondrial genes that were negative. Given the clinical presentation, our differential diagnosis was autosomal recessive, peroxisomal disorders caused by *PEX1* or *PEX6* genes respectively or progressive parkinsonian tremor, dystonia and ataxia and hearing loss caused by *ATP1A3* gene with other cause for retinal phenotype. WGS revealed an extremely rare missense variant R867W in cis with a 4-bp deletion (c.*442_445delTAAA) in the 3-prime UTR of the *PEX6* gene in both siblings, inherited from father, who was homozygous for the 4-bp deletion in the 3-prime UTR. A recent study by Falkenberg KD et al reported on allelic expression imbalance in *PEX6*, with overrepresentation of a pathogenic *PEX6* variant, than wild-type, when in cis with this same 3-prime UTR deletion. Mother had 2 wild type alleles and no 4-bp deletion in the 3-prime UTR. This finding confirms our clinical suspicion of a peroxisomal biogenesis disorder spectrum and is the underlying genetic etiology for the siblings. Further functional studies are planned to corroborate these findings further. As the cost WGS decreases further, the advantages of WGS with option of providing pharmaco-genomic data, copy number variations and covering noncoding regions along with better coverage of coding regions will make this a “one test” of the future allowing for personalizing treatments, providing better prognosis and counseling for individuals with genetic disorders.

1111T

Relaxation of selective constraints on mitochondrial genome correlates with progression of Huntington's disease. Y. Wang¹, X. Guo¹, M. Orth², Z. Gu¹. 1) Division of Nutritional Sciences, Cornell University, Ithaca, NY; 2) Department of Neurology, Ulm University, Ulm, Germany.

Mitochondrial dynamics and function may be impaired in the brain and peripheral tissues in Huntington's disease (HD). A genome-wide association study revealed an association of genetic variants in loci relevant for mitochondrial homeostasis with CAG independent motor age-at-onset. It suggests the effectiveness of mitochondrial quality control which ensures the maintenance of mitochondrial homeostasis such as mitochondrial DNA (mtDNA) quality may be compromised in HD. Expansion of mtDNA mutations in somatic tissues can cause mitochondrial dysfunction in age-related diseases including neurodegeneration. However, it has not been studied in the context of HD. Here, we applied a targeted sequencing approach to evaluate mtDNA mutation spectra in HD lymphoblast and longitudinal blood samples collected in the European Huntington's Disease Network's REGISTRY cohort. We achieved deep sequencing coverage on mtDNA (median depth: ~3600X in lymphoblast and 6100X in blood). Despite similar incidence and variant allele fraction (VAF) distribution for the overall mtDNA mutations in lymphoblasts between 1549 HD patients and 182 unaffected controls, we found increased VAFs for predicted pathogenic mtDNA mutations in HD lymphoblasts. Accordingly, the variant dosage as the sum of VAFs for predicted pathogenic mutations was elevated in HD patients compared to controls by > 70% (0.11 vs 0.06), which was also higher in HD patients in middle-late stages compared to that in patients in prodromal-early stages of the disease (0.13 vs 0.10; $P < 0.006$). We further found that the expansion of VAFs of pre-existing predicted pathogenic mutations in the baseline blood of 181 HD patients was correlated with the degree of disease stage change and the exacerbation of clinical symptoms during a median follow-up of 6 years. Our results point to the expansion of pathogenic mtDNA mutations in the hematopoietic system as a molecular feature of HD progression, supporting the role of defective mitochondrial quality control in HD pathogenesis. In sum, our targeted mtDNA sequencing approach provides insights into the molecular basis of the progressive decline of mitochondrial function in HD and a potential biomarker for HD progression using mtDNA in peripheral tissues.

1112F

A new case of an intermediate phenotype along the spectrum of

ATP1A3-related neurological disorders. L. Fernandez^{1,2}, J. Kohler^{1,2}, D. Zastrow^{1,2}, D. Bonner^{1,2}, C. McCormack^{1,2}, M. Majcherska^{1,2}, S. Marwaha^{1,2}, C. Curnin^{1,2}, C. Reuter^{1,2}, Members. Undiagnosed Diseases Network³, C. Eng⁴, MRZ. Ruzhnikov^{1,5,6}, PG. Fisher^{1,5,7}, EA. Ashley^{1,2,6}, JA. Bernstein^{1,7,8}, MT. Wheeler^{1,2}. 1) Stanford Center for Undiagnosed Diseases, Stanford University, Stanford, CA; 2) Division of Cardiovascular Medicine, Stanford University, Stanford, CA; 3) NIH Undiagnosed Diseases Network, Common Fund, Office of the Director and the National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 4) Baylor Miraca Genetics Laboratories, Houston, TX; 5) Department of Neurology, Stanford University School of Medicine, Stanford, CA; 6) Department of Genetics, Stanford School of Medicine, Stanford, CA; 7) Department of Pediatrics, Stanford University School of Medicine, Stanford, CA; 8) Lucille Packard Children's Hospital Stanford, Stanford, CA.

The *ATP1A3* gene encodes the alpha-3 catalytic subunit of the Na⁺/K⁺(⁺)-ATPase ion pump, critical for the maintenance of the electrochemical gradient of the plasma membrane. It is expressed in the central nervous system and is fundamental for neuronal homeostasis and synaptic transmission. *ATP1A3*-related neurologic disorders represent a clinical continuum in which at least three distinct phenotypes have been described: Rapid-onset Dystonia-Parkinsonism (RDP); Alternating Hemiplegia of Childhood (ACH); and, Cerebellar ataxia, areflexia, pes cavus, optic atrophy, and sensorineural hearing loss (CAPOS) syndrome. Some affected individuals have intermediate phenotypes that do not fit well into one of these major phenotypes. We report on an 8 year-old Hispanic male with a history of developmental delay, ataxia, chorea, epilepsy, hypotonia, and hyporeflexia. Trio whole exome sequencing performed through the Undiagnosed Diseases Network identified a *de novo*, pathogenic variant in the *ATP1A3* gene: heterozygous c.2267G>A (p.R756H). This mutation was first reported in severe early onset RDP phenotype, and has been also associated with a familial childhood-onset progressive cerebellar syndrome, and intermediate phenotypes. Six patients with *ATP1A3* mutations at residue 756, including four with R756H, were recently reported with a distinct phenotype characterized by fever-induced paroxysmal weakness and encephalopathy (FIPWE). Our patient's clinical features overlap those reported with *ATP1A3*-related conditions; shared features include ataxia, hyporeflexia and hypotonia. However, our patient lacks the characteristic features of AHC, RDP, CAPOS syndrome, as well as, FIPWE. Recent reports of atypical presentations related to *ATP1A3* associate residue 756 with diverse emerging phenotypes including ataxia as a predominant feature. This case report contributes to the understanding of the genotype-phenotype correlations of *ATP1A3*. Further identification and characterization of patients with intermediate phenotypes will be essential to understand the molecular mechanisms underlying the pathogenesis of *ATP1A3* mutations.

1113W

Predicting pathogenicity and functional effects of missense variants in

voltage-gated sodium and calcium channels. H.O. Heyne^{1,2,11}, S. Iqbal^{1,2,11}, T. Singh^{1,2,11}, E. Perez³, D. Palmer^{1,2,11}, H.-R. Wang², N. Whiffin⁴, K. Johannesen^{5,12}, R. Møller^{5,12}, M.K. Veerapen^{6,9,11}, A. Palotie^{6,9,11}, F. Wagner², P. May⁷, A.J. Campbell⁸, D. La^{2,3,8,10}, M.J. Daly^{1,2,6,11}. 1) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; 2) Broad Institute, Stanley Center for Psychiatric Research, Cambridge, MA, USA; 3) Cologne Center for Genomics (CCG), University of Cologne, Germany; 4) MRC London Institute of Medical Sciences, National Heart and Lung Institute, Imperial College London, London, UK; 5) The Danish Epilepsy Center, Dianalund, DK; 6) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; 7) Luxembourg Centre for Systems Biomedicine, Esch-sur-Alzette, University of Luxembourg, Luxembourg; 8) Genomic Medicine Institute, Lerner Research Institute Cleveland Clinic, OH, US; 9) Psychiatric & Neurodevelopmental Genetics Unit, Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA; 10) Epilepsy Center, Neurological Institute, Cleveland Clinic, Cleveland, US; 11) Program in Medical and Population Genetics, Broad Institute of MIT & Harvard, Cambridge, Massachusetts, USA; 12) Institute for Regional Health Services, University of Southern Denmark, Odense, Denmark.

Malfunctions of multiple voltage-gated sodium (Na.s encoded by *SCNxA* genes) and calcium channels (Ca.s encoded by *CACNA1x* genes) have been associated with severe neurological, psychiatric, cardiological and other diseases. Most of those diseases are a consequence of either a gain or loss of ion channel function (GOF or LOF, respectively). Individual variants' functional effects are currently assayed in costly, difficult-to-interpret electrophysiological experiments. We thus developed a computational method that predicts LOF/GOF/neutral effects of missense variants. From the scientific literature, we determine whether genetic variants contribute to a given disease via GOF or LOF of the respective ion channel. For example, in *SCN1A*, which is associated with Dravet syndrome, the majority of electrophysiologically tested patients variants are LOF, so we assume, that variants in *SCN1A* lead to Dravet syndrome via a LOF mechanism. In this fashion, we categorized missense variants in fourteen different diseases of three *CACNA1x* (*CACNA1A/E/S*) and six *SCNxA* (*SCN1/2/4/5/8/9A*) genes into putative LOF or GOF. Mapped onto the gene family protein sequence alignment, we find significant clusters of published and unpublished inferred 537 LOF or 349 GOF missense variants, suggesting similar pathomechanisms across these ion channels. We identify >20 sequence-based and structure-based features that are associated with either LOF or GOF variants. Harnessing these observed correlations between protein features and functional impact across the gene family, we develop a machine learning method to: 1) Predict the functional impact (LOF/GOF/neutral) of genetic variants. The most important features include GOF/LOF/neutral variant density in other protein family members, gene family specific conservation and protein secondary structure. Excitingly, our model can predict LOF/GOF consequence of 91 published electrophysiologically characterized variants (ROC=0.75, AUC=0.65). 2) Predict overall pathogenicity in > 2000 cases and 2000 controls better than standard variant prediction tools CADD, Polyphen or MPC. We thus introduce the first computational prediction method of functional LOF/GOF/neutral variant effects in an important protein family and a strategy that can be applied to other gene families. While gaining insight into shared pathomechanisms in this gene family, our work should improve gene discovery and enhance personalized medicine.

1114T

Mutations nature and sex modify hereditary spastic paraplegia type 4

age at onset. L. Parodi¹, S. Fenu¹, M. Barbier¹, G. Banneau¹, C. Duyckaerts^{1,2}, S. Tezenas du Montcel³, M.L. Monin¹, S. Ait Said¹, CME. Tallaksen^{1,4,5}, B. Sablonniere^{6,7}, A. Brice¹, G. Stevanin^{1,8}, C. Depienne^{1,9}, A. Durr¹. 1) Institut du Cerveau et de la Moelle Epinière, Paris, France, France1. Institut du Cerveau et de la Moelle épinière (ICM), Inserm, CNRS, Assistance Publique-Hôpitaux de Paris (AP-HP), Pitié-Salpêtrière University Hospital, F-75006 Paris, France; 2) Raymond Escourolle Department of Neuropathology, La Salpêtrière Hospital, Paris, France; 3) 3.Assistance Publique-Hôpitaux de Paris (AP-HP), Pitié-Salpêtrière University Hospital, Biostatistics and Medical Informatics Unit and Clinical Research Unit, Paris, France; Sorbonne Universités, Université Pierre et Marie Curie Université Paris 06, UMR S1136, Institut Pierre Louis d'Epidémiol; 4) Department of Neurology, Oslo University Hospital, Oslo, Norway; 5) Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway; 6) Univ. Lille, Inserm, CHU Lille, UMR-S 1172 - JPArc - Centre de Recherche Jean-Pierre AUBERT Neurosciences et Cancer, F-59000, Lille, France; CHU Lille, Institut de Biochimie et Biologie moléculaire, Centre de Biologie Pathologie et Génétique, F-59000, Lille, France; 7) CHU Lille, Institut de Biochimie et Biologie moléculaire, Centre de Biologie Pathologie et Génétique, F-59000, Lille, France; 8) Ecole Pratique des Hautes Etudes (EPHE), Paris Sciences et Lettres (PSL) Research University, Neurogenetics team, 75013 Paris, France; 9) Institut für Humangenetik, Universitätsklinikum Essen, Hufelandstraße 55, 45147 Essen, Germany.

Hereditary Spastic Paraplegias (HSPs) are rare (1-5 : 100 000) neurological disorders caused by the progressive distal degeneration of the corticospinal tracts. Among the 79 loci and 65 Spastic Paraplegia Genes (SPGs) involved in HSPs, mutations in *SPAST*, encoding spastin and responsible for Spastic paraplegia 4 (SPG4), are the most frequent cause of both familial and sporadic HSP. SPG4 is characterized by a clinically pure phenotype, predominantly affecting corticospinal tracts and posterior columns of the spinal cord, and rarely associated with additional neurological signs. We analysed a cohort comprising 842 SPG4-HSP patients in order to assess genotype-phenotype correlations and possibly explain the extreme age at onset variability, typical of SPG4-HSP affected patients. Age of onset (n=547) followed a bimodal distribution, characterized by a first peak extending from birth to the first decade and a second one comprised between the third and fifth decades. The underlying cause of the observed bimodal trend, missing in previous reports, was identified when considering the nature of *SPAST* pathogenic mutations. SPG4-HSP causative mutations resulted mostly located in spastin AAA cassette (75%, 201/268), therefore impairing its microtubules severing activity and leading predominantly to a loss of function mechanism. The comparison of age at onset between missense and truncating mutation carriers revealed a significant lower age at onset in the former group when compared to the latter (23 ± 19.7 , n = 180 versus 33.4 ± 16.7 , n = 245; Mann-Whitney test, $p < 0.0001$). This observation not only gives an explanation to SPG4-HSP age at onset bimodal trend, but allows to establish for the first time a genotype-phenotype correlation that links the mutation nature and the disorder age of onset. The sex of the individual appeared as a second factor influencing our cohort's age at onset. When focusing on 84 pedigrees, intrafamilial correlation of age at onset resulted highly significant in same-sex sib-pairs, especially among brothers ($r=0.58$, n = 34, $p = 0.0002$). Moreover, disorder penetrance resulted being lower in women (0.86 versus 0.92 in men, $p = 0.03$). Altogether, these results show that SPG4-HSP age at onset is strongly influenced by mutation nature and sex linked-factors. The presence of still unidentified genetic and environmental modifiers should definitely be investigated to fully understand the observed variability.

1115F

NEDDFL syndrome due to an apparently balanced translocation t(1,17)

(q24.3;q24.2) disrupting BPTF in a 35-year-old man. P. Stankiewicz^{1,2}, N. Tommerup³, B. Panasiuk⁴, L. Cooper⁵, S.E. Scherer⁶, J. Borys⁶, B. Kosztyla Chojna⁷, R. Skowronski⁸, A.T. Midro⁴. 1) Dept of Molecular & Human Genetics, Baylor College of Medicine, Houston, TX; 2) Baylor Genetics, Houston, TX; 3) Wilhelm Johannsen Centre for Functional Genome Research Department of Cellular and Molecular Medicine (ICMM), University of Copenhagen, Denmark; 4) Department of Clinical Genetics, Bialystok, Poland; 5) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 6) Clinic of Maxillo-Facial Surgery; 7) Department of Phonoaudiology and Logopedics; 8) Department of Orthopaedics and Traumatology, Medical University, Bialystok, Poland.

Neurodevelopmental disorder with dysmorphic facies and distal limb anomalies (NEDDFL, OMIM 617755) was recently shown to result from haploinsufficiency of *BPTF* encoding the largest subunit of a nucleosome remodeling factor (NURF), a member of ISWI chromatin remodeling complex. We observed similar phenotypic features in a 35-year-old male with a *de novo* apparently balanced chromosomal translocation t(1;17)(q24.3;q24.2), reported in 1993 with Russell-Silver syndrome (PMID:8403458). FISH with BAC and fosmid clones and WGS revealed that the 17q24.2 breakpoint maps at chr17:65,944,420-65,955,659 (hg19) and disrupts the dosage sensitive *BPTF* (pLI=1.00). The 1q24.3 breakpoint was mapped at chr1:172,228,441-17,234,934 (hg19) disrupting the brain specific gene *DNM3* predicted to tolerate loss-of-function (pLI=0.03). No non-polymorphic CNVs were identified by CMA. Phenotypic analyses performed according to the Munich Dysmorphology Database (MDDb) methodology revealed short stature, hemihypotrophy, microcephaly, triangular shape of asymmetric face, prominent forehead, hypertelorism, protruding eyeballs, broad palpebral fissures, long eye lashes, long nasal bridge, short philtrum, thin lips, and microretrognathia. Limb anomalies include bilateral 5th finger clinodactyly, short phalanges, and broad big first toes. Intellectual disability, speech delay, vision problems, muscle hypotonia, and the defect of phonemic audition have improved over the course of 24 years of observations. Our results expand the clinical spectrum of human disorders caused by ablation of chromatin remodeling complexes.

1116W

Predictive modeling implicates threonyl-tRNA synthetase (TARS) in axonal peripheral neuropathy. R. Meyer¹, Y. Park¹, M. Meisler¹, A. Beg², A. Antonellis¹. 1) Department of Human Genetics, University of Michigan, Ann Arbor, MI 48109, USA; 2) Department of Pharmacology, University of Michigan, Ann Arbor, MI 48109, USA.

Aminoacyl-tRNA synthetases (ARSs) charge tRNA molecules with cognate amino acids, a critical step in translating the genetic code. Mutations in five genes encoding an ARS cause autosomal dominant, axonal peripheral neuropathy, which is characterized by impaired motor and sensory function in the distal extremities. Neuropathy-associated ARS mutations decrease enzyme function in biochemical and yeast complementation assays, and are dominantly toxic when modeled in *C. elegans*. These functional studies provide important evidence for determining the pathogenicity of newly identified mutations in patient populations, suggesting that such models have considerable predictive power. Here, we use the above model systems to determine if mutations in any ARS gene can cause peripheral neuropathy, beginning with threonyl-tRNA synthetase (TARS), which has not been associated with any human disease. Like the five ARS enzymes previously implicated in peripheral neuropathy, TARS is homodimeric and charges tRNA in the cytoplasm, making it a strong candidate for testing our predictive models. We first performed a loss-of-function screen in yeast by testing 10 TARS missense mutations that affect highly conserved amino-acid residues in critical regions of the enzyme. Three mutations (N412Y, R433H, and G541R) reduce or ablate yeast cell growth, indicating impaired enzyme function. We then used CRISPR/Cas9 mutagenesis to model the loss-of-function TARS mutations in the *C. elegans* ortholog *tars-1* and tested each variant for dominant toxicity. From these assays, we identified one TARS variant (G541R) that is loss-of-function in yeast and dominantly toxic in worm. These findings make G541R TARS similar to HARS and AARS variants that are implicated in neuropathy. Toward assessing G541R TARS for a role in peripheral neuropathy, we generated and validated a knock-in mouse strain for this variant. Here, I will present our unpublished computational, yeast, *C. elegans*, and mouse data indicating that certain human TARS mutations can cause axonal peripheral neuropathy. I will also present our efforts to directly link TARS to human disease by testing six TARS variants found in patient populations. This work provides a framework for using functional assays to proactively identify novel ARS loci and alleles that contribute to axonal peripheral neuropathy.

1117T

Accurate molecular detection and functional validation of PI3K-AKT-MTOR pathway in focal malformations of cortical development: Novel genetic insights. G. Ruggeri¹, L.L. Gonzalez², A.E. Timms³, A. Piacitelli⁴, W.B. Dobyns^{1,2}, P. Filomena¹, J.G. Ojemann⁵, E.J. Novotny⁶, S. Poliachik⁷, R.F. Hevner⁸, R.P. Kapur⁹, G.M. Mirzaa^{1,2}. 1) Center for integrative Brain Research, Seattle Children's Research Institute, Seattle, WA; 2) Seattle Children's Hospital, Division of Genetic Medicine, Department of Pediatrics, University of Washington, Seattle, WA; 3) Center for Developmental Biology and Regenerative Medicine, Seattle Children's Research Institute, Seattle, WA; 4) Broad Institute of MIT and Harvard, Boston, MA, USA; 5) Department of Neurosurgery, Seattle Children's Hospital, Seattle, WA, USA; 6) Department of Neurology, Seattle Children's Hospital, Seattle, WA, USA; 7) Department of Radiology, Seattle Children's Hospital, Seattle, WA, USA; 8) Department of Pathology, University of California San Diego, CA, USA; 9) Department of Pathology, Seattle Children's Hospital, Seattle, WA, USA.

Focal malformations of cortical development (FMCD) constitute a spectrum of developmental brain disorders associated with brain overgrowth, cellular dysplasia, and co-morbidities such as intractable epilepsy, autism and intellectual disability. Constitutional and mosaic mutations of the PI3K-AKT-MTOR pathway genes have been identified in this spectrum. Identifying the underlying molecular etiology in these disorders has significant benefits for patients and their families, including informed genetic counseling and personalized therapeutic interventions. Widely used diagnostic methods, such as standard-depth whole exome sequencing, are not sensitive enough to detect mosaic mutations common among FMCD patients. These methods are also not suitable for screening limited amounts of surgically resected FMCD brain tissue due to their high DNA input requirements. Ultimately, more optimal molecular diagnostic methods and high throughput pre-clinical models are necessary to identify and study the effects of FMCD-related mutations on brain development. To address these challenges for characterizing FMCD, we utilized digital droplet PCR (ddPCR) as a new and sensitive tool for screening common mutations in the PI3K-AKT-MTOR pathway. We screened a cohort of 54 families (110 samples), including 50 singletons and 4 trios for 8 of the most common FMCD-associated mutations: *PIK3CA* (E545K, E542K, H1047R, H1047L), *AKT3* (E17K), and *MTOR* (S2215Y, S2215F). Of 54 patients, we identified mosaic mutations in 9 individuals (27 samples; 16.7%) in *PIK3CA* (N=7), *AKT3* (N=1), and *MTOR* (N=1). Mutation-positive samples ranged in alternate allele frequencies (AAF) from 0.3% to 25.6% across several tissues. To understand the functional consequences of these mutations on early fore-brain development, we generated 7 fibroblast patient-derived induced pluripotent stem cell (iPSC) lines by transcriptional reprogramming and differentiated them into 2D neuronal cultures. Our results suggest that targeted molecular analysis for the common PI3K-AKT-MTOR pathway mutations by ddPCR is an effective molecular diagnostic approach for FMCD. Given the high sensitivity and low DNA input requirements for this assay, ddPCR may also be an effective molecular tool for other mosaic disorders with a narrow mutational spectrum. Finally, our iPSC lines provide a valuable model for dissecting the functional consequences of these mutations and future drug screens using PI3K-AKT-MTOR pathway inhibitors.

1118F

De novo mutations in *TMEM63A* result in transient hypomyelination during infancy. C. Simons^{1,2}, H. Yan^{3,4,5}, G. Helman^{1,2}, H. Ji^{3,4}, J. Crawford², T. Kubisiak⁶, S.J. Bent⁷, J. Xiao⁸, R.J. Taft⁹, Y. Wu^{3,4}, A. Pop^{3,9,10}, D. Li^{3,4,11}, L.S. de Vries^{12,13}, Y. Jiang^{3,4,14}, G.S. Salomons^{8,9,10,15}, M.S. van der Knaap^{9,16}, N.I. Wolf⁶, M. Burmeister^{5,17}, J. Wang^{3,4,14}. 1) Murdoch Children's Research Institute, Melbourne, Vic, Australia; 2) Institute for Molecular Bioscience, The University of Queensland, Brisbane Australia; 3) Department of Pediatrics, Peking University First Hospital, Beijing, China; 4) Beijing Key Laboratory of Molecular Diagnosis and Study on Pediatric Genetic Diseases, Peking University First Hospital, Beijing, China; 5) Molecular & Behavioral Neuroscience Institute, University of Michigan, Ann Arbor, MI, USA; 6) Department of Radiology, Peking University First Hospital, Beijing, China; 7) Illumina, Inc., San Diego, CA USA; 8) Metabolic Unit, Department of Clinical Chemistry, VU University Medical Center; 9) Department of Child Neurology and Amsterdam Neuroscience, VU University Medical Center, Amsterdam, the Netherlands; 10) Amsterdam Gastroenterology & Metabolism, Amsterdam, the Netherlands; 11) Henan Provincial Key Laboratory of Children's Genetic and Metabolic Diseases, Children's Hospital Affiliated to Zhengzhou University, Zhengzhou, China; 12) Department of Neonatology, Wilhelmina Children's Hospital, University Medical Center Utrecht, Utrecht, the Netherlands; 13) Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, the Netherlands; 14) Key Laboratory for Neuroscience, Ministry of Education/National Health and Family Planning Commission, Peking University, Beijing, China; 15) Genetic Metabolic Diseases, Academic Medical Center, Amsterdam, The Netherlands; 16) Department of Functional Genomics, VU University, Amsterdam, the Netherlands; 17) Departments of Computational Medicine & Bioinformatics, Psychiatry and Human Genetics, University of Michigan, Ann Arbor, MI, USA.

Hypomyelinating leukodystrophies form a large and heterogeneous group among inherited white matter disorders. Clinical presentations include nystagmus, developmental delay, hypotonia, ataxia, spasticity and intellectual disability. Using trio genome and exome sequencing, we identified four unrelated individuals that each harbor one of three novel heterozygous missense variants in *TMEM63A*. None of the variants are present in GnomAD or dbSNP, all are predicted damaging and encode highly conserved amino acids that reside either within or directly adjacent to transmembrane helices. The variants were confirmed to have arisen *de novo* in three of the four individuals. Each of the affected individuals presented in infancy with congenital nystagmus, motor delay and deficient myelination, prompting the diagnosis of Pelizaeus-Merzbacher or Pelizaeus-Merzbacher-like disease. Unexpectedly, the clinical evolution of all four individuals has been surprisingly favorable, with resolution of nystagmus and near-normalization of neurological signs and developmental progression. In the three individuals with follow-up imaging after four years of age, the myelin deficit had almost completely resolved. The fourth individual also displayed a substantial improvement of myelination when last imaged at 31 months of age. *TMEM63A* encodes an osmosensitive calcium-permeable cation channel that has not previously been associated with disease. The function of *TMEM63A* has not been well characterized. It is highly expressed in oligodendrocytes and has been shown to be enriched in lysosomal membranes. These findings add *TMEM63A* to the list of genes related to myelin disorders and represent a novel phenotype of an infantile disorder with transient hypomyelination, providing important prognostic information for affected families.

1119W

A novel de novo frameshift mutation in *KAT6A* identified by whole exome sequencing. A. Alkhateeb. Jordan University of Science and Technology, Irbid, Jordan.

Intellectual disability is a common condition with multiple etiologies. The identified genes of monogenic forms have increased steadily in recent years due to the implementation of next generation sequencing. Here, we describe a two-year old happy boy with global developmental delay and intellectual disability. The child had feeding difficulties since birth. He had delayed motor skills and muscular hypotonia. Upon brain MRI, diffuse white matter loss and thinning of the corpus callosum were found. Chromosome karyotyping and array CGH were normal. Whole exome sequencing revealed a novel de novo frameshift mutation c.3390delA (p.Lys1130Asnfs*4) in *KAT6A* gene (NM_006766.4). The heterozygous mutation was confirmed by Sanger sequencing in patient and its absence in parents. *KAT6A* gene, coding a histone acetyltransferase has been recently found to be associated with a neurodevelopmental disorder (OMIM: #616268, autosomal dominant mental retardation 32). Features of this specific disorder are nonspecific which makes it difficult to characterize the condition based on the clinical symptoms alone, thus, our findings confirm the utility of whole exome sequencing to quickly and reliably identify the etiology of such conditions.

1120T

De novo variants affecting autoinhibition of Calcium/calmodulin-dependent protein kinase II, cause neurodevelopmental disorders. H. Saitsu¹, T. Akita², K. Aoto¹, M. Kato³, M. Shiina⁴, M. Nakashima¹, I. Kuki⁵, S. Okazaki⁶, S. Magara⁷, T. Shihara⁸, K. Yokochi^{8,9}, K. Aiba⁸, J. Tohyama⁸, C. Ohba¹⁰, S. Miyake¹⁰, N. Miyake¹⁰, K. Ogata⁴, A. Fukuda², N. Matsumoto¹⁰. 1) Dept Biochem, Hamamatsu Ika Daigaku, Hamamatsu, Shizuoka, Japan; 2) Dept Neurophysiology, Hamamatsu Ika Daigaku, Hamamatsu, Shizuoka, Japan; 3) Dept Pediatrics, Showa Univ Sch Med, Japan; 4) Dept Biochem, Grad Sch Med, Yokohama City Univ, Japan; 5) Dept Pediatric Neurol, Pediatric Medical Care Center, Osaka City General Hospital, Japan; 6) Dept Pediatrics, Epilepsy Center, Nishi-Niigata Chuo National Hospital, Japan; 7) Dept Neurology, Gunma Children's Medical Center, Japan; 8) Dept Pediatric Neurol, Seirei-Mikatahara General Hospital, Japan; 9) Dept Pediatrics, Toyohashi Municipal Hospital, Japan; 10) Dept Human Gen, Grad Sch Med, Yokohama City Univ, Japan.

α (CAMK2A) and β (CAMK2B) isoforms of Calcium/calmodulin-dependent protein kinase II (CaMKII) play a pivotal role in neuronal plasticity and learning and memory processes in the brain. By whole exome sequencing, we identified *de novo* CAMK2A and CAMK2B variants in three and two individuals with neurodevelopmental phenotypes, respectively. Intellectual disability and developmental delay were common to all individuals, and were profound in four individuals. Seizures were also accompanied in three individuals with CAMK2A variants. Five variants include one splice site variant and four missense variants at evolutionarily conserved amino acids. Using a minigene splicing assay, we demonstrated that the splice site variant caused skipping of exon 11 leading to an in-frame deletion of the regulatory segment of CaMKII α . By mapping the variant sites on the crystal structures of human CaMKII α , all the four missense variants are predicted to impair the interaction between the kinase domain and the regulatory segment responsible for the autoinhibition of its kinase activity. The Thr286/Thr287 phosphorylation as a result of release from autoinhibition was increased in three mutants when the mutants were stably expressed in Neuro-2a neuroblastoma cells. Expression of a CaMKII α mutant (P212Q) in primary hippocampal neurons significantly increased A-type K⁺ currents, which facilitated spike repolarization of single action potentials. Knock-in mice of *Camk2b*^{P213L/P213L} showed ataxic gait and beam walking test revealed impairment of fine motor coordination and balance. Our data highlight the importance of CaMKII α and CaMKII β and their autoinhibitory regulation in human brain function.

1121F

Treatment in Niemann Pick type C patients with a HSP amplifier, arimoclomol: Results from an observational study and follow-on randomized placebo-controlled double-blind interventional trial. E. Mengel¹, B. Bembel², M. Del Toro³, F. Deodato⁴, M. Patterson⁵, M. Gautschi⁶, P. Giraldo⁷, S. Grunewald⁸, S. Grønberg⁹, P. Harmatz¹⁰, B. Heron¹¹, E. Maier¹², A. Robertson¹³, S. Santra¹⁴, A. Tylki-Szymanska¹⁵, S. Day¹⁶, L. Ingemann¹⁷, N.H.T. Petersen¹⁷, T. Kirkegaard Jensen¹⁷, C. iDali¹⁷. 1) Villa Metabolica Mainz; 2) Azienda Ospedaliero-Universitaria "Santa Maria della Misericordia" di Udine; 3) Hospital Vall D'Hebron Barcelona; 4) Ospedale Pediatrico Bambino Gesù, Rome; 5) Mayo Clinic Children's Center, Rochester; 6) "INSELSPITAL University Hospital Bern; 7) Hospital QUIRÓN, Zaragoza; 8) Great Ormond Street Hospital, London; 9) University Hospital Copenhagen (Rigshospitalet); 10) Children's Hospital Oakland, CA; 11) Hôpital Trousseau, Paris; 12) Dr. von Haunersches Kinderspital der Universität München; 13) CHU de Montpellier; 14) Birmingham Children's Hospital; 15) The Children's Memorial Institute Warsaw; 16) Portway, North Marston, Buckinghamshire UK; 17) Orphazyme A/S Copenhagen.

Background: Niemann Pick Type C (NPC) is a rare genetic, lysosomal storage disorder caused by mutations of the NPC protein. Systemic and neurological symptoms are heterogeneous progressive and lead to death. A prospective **observational study** in NPC patients was initiated leading into a **randomized placebo-controlled double-blind interventional trial with a heat shock protein (HSP) amplifier; arimoclomol**. HSP is known to improve the folding of NPC1 protein and improve lysosomal function. **Methods:** Patients, 2 to 18 years of age, diagnosed with NPC, who had at least one neurological symptom and preserved ability to walk with assistance were eligible. Patients were maintained on standard therapy. Assessments were conducted at baseline and minimally at 6 months to observe the natural disease course before patients rolled into the 2:1 randomized placebo-controlled interventional trial receiving a daily dose of up to 600 mg arimoclomol or placebo orally for 12 months. The key assessments performed were: a 5-domain NPC-severity score (ambulation, cognition, fine motor skills, speech and swallowing), global clinical impression score (CGI-I), safety and samplings for biomarkers. **Results:** In the observational study, 36 patients with a mean age of 9.86 years (SD4.61) were enrolled across 12 sites in 7 European countries. 27 patients continued in the interventional trial and an additional 25 patients were enrolled from the same sites along with additional 2 sites in the US. Data on disease progression showed a mean increase of 1.4 (SD=2.9) on the 5-domain NPC-CSS corresponding to a mean rate of change of 0.75/6 months (SD=1.58). The 5-domain NPC-CSS is an abbreviated score of the NPC-CSS. A novel glycosphingolipid species was discovered as potential biomarker and a dramatically decreased level of HSP70 in white blood cells were observed. The trial is completed and currently, data is being processed. The results at 12-months on the safety and efficacy of arimoclomol will be reported for the rate of progression and for a responder analysis of the proportion of patients who stabilize or improve by CGI-I compared to placebo. **Conclusions:** Progression rates in the cohort from the observational study were shown to be comparable to previously reported studies. These clinical findings as well as the biomarker data will support the interpretation of this Phase II/III intervention trial results that will show if orally administered arimoclomol can safely alter the disease course.

1122W

Inherited peripheral neuropathies: Analysis of *PDXK* gene identifies a new treatable disorder. V. Chelban^{1,2}, J. Vandrovcova², M. Wilson³, S. Efthymiou², E. Zamba², Papanicolaou², N. Wood², P. Mills³, P. Clayton³, H. Houlden², SYNAPSE Study Group. 1) Department of Neurology and Neurosurgery, Institute of Emergency Medicine, Chisinau, Moldova; 2) 1Department of Molecular Neuroscience, Institute of Neurology, University College London, London, UK; 3) Genetics and Genomic Medicine, GOS Institute of Child Health, University College London, London, UK; 4) The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus.

Peripheral neuropathies are amongst the most common neurological conditions worldwide affecting over 20 million people. Until now, no disease-modifying treatment has been established in these conditions. This is unsurprising since 40% of patients with primary peripheral neuropathies have no disease-causing mutation identified. We performed a combination of whole genome sequencing, homozygosity mapping and segregation analysis for novel disease-causing gene discovery in patients with severe peripheral neuropathy. Pathogenicity was confirmed via enzymatic assays and mass spectroscopy on recombinant protein and patient-derived fibroblasts, plasma and erythrocytes. We used circular dichroism to show secondary structure changes and isothermal titration calorimetry to investigate the impact of mutation on the ATP binding. We report that biallelic mutations in human *PDXK* are associated with primary axonal peripheral neuropathy and optic atrophy, characterised by blindness and severe, muscle weakness and wasting, leading to patients becoming wheelchair-bound. Pyridoxal kinase (*PDXK*) is involved in converting vitamin B₆ to its active form, pyridoxal 5'-phosphate (PLP). We show that *PDXK* mutations lead to pathology via decreased plasma PLP concentrations. Our functional studies revealed conformational rearrangement in the mutant enzyme around the kinase ATP binding pocket with impaired *PDXK* ability to bind ATP and reduced erythrocyte *PDXK* activity. We show that both the human clinical picture and biochemical profile in *PDXK* mutations are rescued by PLP supplementation. Patients regained their ability to walk independently. Furthermore, treatment led to normalisation of plasma PLP levels, correlated with reduction of neurofilament light chain concentrations, a biomarker of axonal breakdown. In conclusion, biallelic mutations in human *PDXK* are associated with a novel disorder leading to a treatable primary axonal peripheral neuropathy and optic atrophy. As B₆ is a cofactor for more than 70 enzymes representing diverse, essential biological pathways it can be linked to diseases through a wide-range of processes involved in the vitamin B₆ pathway, these findings have direct implications for neuropathies of different aetiology characterised by reduction of PLP levels. Collectively, these data establish the importance of the vitamin B₆ pathway for normal neurological function in both central and peripheral nervous systems and thus identifies PLP as a therapeutic target.

1123T

Molecular analysis and functional characterization of mutations in Niemann-Pick disease type A and type B in Indian patients. D. Deshpande¹, S. Gupta², P. Ranganath^{1,2}, J. Seth³, M. Kabra⁴, N. Gupta⁴, KM. Girisha⁵, S. Phadke⁶, C. Datar⁷, A. Dalal^{1,2}. 1) Center for DNA fingerprinting and Diagnostics, Hyderabad, (TS), India; 2) Nizam's Institute of Medical Sciences, Hyderabad, (TS) India; 3) Foundation for Research in Genetics and Endocrinology (-FRIGE), Ahmedabad, Gujrat, India; 4) All India Institute of Medical Sciences, New Delhi, India; 5) Kasturba Medical College, Manipal, Karnataka, India; 6) Sanjay Gandhi Post Graduate Institute of Medical Sciences, Lucknow, (UP) India; 7) Sahyadri Medical Genetics and tissue culturing facility, Pune (MH), India.

Niemann-Pick disease (NPD) is a rare Mendelian - autosomal recessive - lysosomal storage disease, characterized by accumulation of sphingomyelin inside the lysosomes. Mutations in (Sphingomyelin-phosphodiesterase-1) *SMPD1* gene, encoding enzyme Acid sphingomyelinase (ASM), Cause Type A or Type B NPD. Inability of mutated ASM to cleave sphingomyelin, results in lysosomal accumulation causing cell death. The severity of NPD depends on type of mutation and its location in *SMPD1* domains. In this study, patients with NPD were recruited based on clinical evaluation by geneticist and enzyme assay results, after due informed consent from parents and approval from institutional ethical committee. After DNA isolation, all 6 exons with flanking intronic regions were sequenced. Identified variants were verified by checking in 1000G, ExAC, dbSNP databases for polymorphisms and in HGMD and Clinvar databases for known pathogenic variants. Variants were then subjected to various mutation prediction software e.g. Polyphen2, SIFT, and Hansa, CADD etc. Variants predicted as disease causing and absent in above databases (novel variants) were selected for functional characterization. We found total 81 mutations in 138 patients, including 46 Novel mutations out of which, 31 missense, 1 nonsense and 14 frameshift. The PDB structure of *SMPD1* enzyme shows a Saposin domain which plays crucial role in recruiting sphingomyelin to lysosomal membrane and the phosphodiesterase domain 3 (catalytic domain) which cleaves the lipid. We further characterized missense mutations in these regions e.g. c.314T>C, c.647T>G, c.1783C>G, c.1028G>C, c.1429C>A using Molecular Dynamic Simulations. Till now, 10 missense mutations have been analysed and *in vitro* functional characterisation for the same is being done. The change in an amino acid can alter or destabilize the enzyme structure or substrate interactions. Mutation at position c.1783C>G makes enzyme structure unstable whereas c.1048G>C makes the structure rigid. This may result in no substrate binding at all or too tight substrate binding, respectively. Knowledge regarding pattern of mutations in different populations helps to discern demography and pathophysiology of the disease. We have been successful in establishing mutation spectrum for Indian patients. The present study has aided in accurate prenatal diagnosis and proved helpful for genetic counselling.

1124F

Clinical features of arginase 1 deficiency: Review of literature case series. G.A. Diaz¹, N. Longo², A. Schulze³, G. Bubb⁴, S. Eckert⁴, K.C. Patki⁴, J.E. Wooldridge⁴, J.L. Merritt II⁵. 1) Icahn School of Medicine at Mt. Sinai, New York City, NY; 2) University of Utah, Salt Lake City UT, USA; 3) University of Toronto and The Hospital for Sick Children, Toronto, ON, Canada; 4) Aeglea BioTherapeutics, Austin, TX, USA; 5) University of Washington, Seattle, WA, USA.

Background: Arginase 1 Deficiency (ARG1-D) is a rare, progressive, autosomal recessive disease caused by deficiency of a key arginine metabolizing enzyme. The high plasma arginine level is believed to be the key driver of spasticity, developmental delay and seizures that develop in early childhood and progress over time. **Methods:** We identified more than 100 unique, historical ARG1-D cases in the literature to develop a more systematic understanding of the clinical presentation, disease progression, and impact of available management on both arginine levels and disease manifestations. **Results:** Initial review of 3 largest series comprised of 85 unique ARG1-D patients was completed. At publication, mean age of living patients was 16.1 years (range 1–37; median 13); 11 patients were deceased with mean age at death of 13.9 years (range 0.13 – 45; median 12). Mean age at presentation was 2.1 years (range 0–6; median 1.9), and mean age at diagnosis was 6.9 years (range 0–27; median 5). The primary disease manifestation was spasticity, with lower limb (LL) and upper limb (UL) spasticity reported in 66% and 31% of cases respectively. In all but 2 cases with UL spasticity LL spasticity was also seen. Abnormal gait/ataxia was reported in 21% and 10% were non-ambulatory. Cognitive impairment was reported in 74% and was moderate or severe in 40%. Seizures were reported in 66% of patients, 58% had periodic vomiting, and 22% had growth retardation. A protein-restricted diet was used in 89% of patients and 58% received nitrogen scavengers. Transaminases >ULN (upper limit of normal) were reported in 39% of cases and 35% had liver pathology. Mean plasma arginine was 634 μ M (range 96–1756), mean serum α -ketoglutarate was 5.72 μ M (ULN 0.25) and mean serum arginase activity was 3.37 μ M (ULN 0.115), mean plasma guanidinoacetate was 4.74 μ M (ref range: 0.8–2.6). Plasma ammonia > ULN was reported in 93% cases and 57% had a least one value >100 μ M. Percentages were based on publication report; therefore, actual incidence could be higher. **Discussion:** This analysis demonstrates the significant morbidity in patients with ARG1-D, including evidence of premature death. Despite symptoms in early childhood, diagnosis is often delayed. Plasma arginine remains markedly elevated despite standard disease management, including dietary protein restriction, with evidence of continued disease progression. This analysis supports the development of much needed new therapies in ARG1-D.

1125W

Oral plasmalogen precursor, PPI-1040, improves plasmalogen levels and hyperactivity in a Pex7 deficient mouse model of Rhizomelic Chondrodysplasia Punctata type 1 (RCDP1). W. Fallatah^{1,2}, T. Smith³, W. Cui¹, D. Jayasinghe³, E. Di Pietro¹, S. Ritchie³, N. Braverman¹. 1) Department of Human Genetics, McGill University, Research Institute of the MUHC, Montreal, Canada; 2) Department of Medical Genetics, King Abdul-Aziz University, Jeddah, Saudi Arabia; 3) Med-Life Discoveries LP, Saskatoon SK, Canada.

Background: RCDP1 is a peroxisome biogenesis disorder distinguished by congenital cataracts, a characteristic skeletal dysplasia and neurological impairment. This results in profound psychomotor and growth retardation, and early demise. RCDP1 is caused by defects in the peroxisome receptor, PEX7, resulting in negligible biosynthesis of plasmalogen phospholipids (PL). Disease severity directly correlates to residual PL levels. In fact, patients with milder forms of RCDP have residual erythrocyte PLs that are around 34% of controls. Currently, there is no disease modifying therapy for RCDP. We generated novel hypomorphic Pex7 deficient mice (B6;129S6-Pex7^{tm2.38rev}, referred here as Pex7^{HypoNull}) with 90% survival, resembling mild RCDP. The total PL levels were decreased to 20% of Pex7^{WT/WT} (WT) in plasma, erythrocytes. PL levels were 17% of WT in brain, and 26–40% in liver, small intestine and skeletal muscle. Our Pex7^{HypoNull} mice exhibit a hyperactive behavior in the open field environment. These data establish useful biochemical and behavioral endpoints for preclinical trials. **Aim:** To evaluate the therapeutic response of Pex7^{HypoNull} mice to oral PL precursor supplementations PPI-1011, sn1-1-0-alkyl[C16:0], sn2-DHA, sn3-lipoic acid and PPI-1040, sn1-1-0-alkenyl[C16:0], sn2-DHA, sn3-cyclic phosphoethanolamine (PE). **Method:** 3 cohorts of adult male and female Pex7^{HypoNull} mice received oral gavage with PPI-1011, PPI-1040 or vehicle control at 50 mg/kg for 4 weeks (5x/week). The study included a matched cohort of untreated WT. The open field test was performed at baseline and after treatment. Tissues were collected for PL analysis by LC/MSMS. **Results:** In PPI-1040 treated Pex7^{HypoNull} mice, total C16:0 PE and target C16:0/DHA PL levels increased in plasma to 81% and 72% of WT, in erythrocytes to 28% and 38% of WT, in liver, small intestine, and skeletal muscle to 52–60% and 60–65% of WT, respectively; there was no increase in PLs in brain. Interestingly, PPI-1040 treated Pex7^{HypoNull} mice showed significant improvement of the hyperactivity phenotype compared to vehicle treated Pex7^{HypoNull} controls. Although PPI-1011 did not improve PL levels or hyperactivity, PPI-1040 is a promising therapy for RCDP. Recovery of the behavioral phenotype could be secondary to changes in brain neurochemistry not apparent by analysis of PL levels in the whole brain, or secondary to the documented PL increase in peripheral tissues.

1126T

Expansions of intronic TTTCa and TTTTA repeats in three genes cause benign adult familial myoclonic epilepsy. H. Ishiura¹, K. Doi², J. Mitsui³, J. Yoshimura², M.K. Matsukawa¹, A. Fujiyama⁴, A. Kakita⁵, W. Qu², K. Ichikawa², S. Shibata¹, A. Mitsue¹, Y. Ichikawa⁶, Y. Takahashi⁷, M. Tanaka⁸, K. Abe⁹, R. Koike¹⁰, T. Yasuda¹¹, A. Sano¹², A. Ikeda¹³, J. Goto¹⁴, S. Morishita², S. Tsuji¹⁴.

1) Department of Neurology, The University of Tokyo, Tokyo, Tokyo, Japan; 2) Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Chiba, Japan; 3) Department of Molecular Neurology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; 4) Advanced Genomics Center, National Institute of Genetics, Shizuoka, Japan; 5) Department of Pathology, Brain Research Institute, Niigata University, Niigata, Japan; 6) Department of Neurology, Kyorin University, Tokyo, Japan; 7) Department of Neurology, National Center of Neurology and Psychiatry, Tokyo, Japan; 8) Institute of Medical Genomics, International University of Health and Welfare, Chiba, Japan; 9) Department of Neurology, Graduate School of Medicine, Dentistry, and Pharmaceutical Sciences, Okayama University, Okayama, Japan; 10) Department of Neurology, Nishi-Niigata Chuo National Hospital, Niigata, Japan; 11) Department of Neurology, Kurashikiken Hospital, Okayama, Japan; 12) Department of Psychiatry, Kagoshima University Graduate School of Medical and Dental Sciences, Kagoshima, Japan; 13) Department of Epilepsy, Movement Disorders and Physiology, Kyoto University Graduate School of Medicine, Kyoto, Japan; 14) Department of Neurology, International University of Health and Welfare Mita Hospital, Tokyo, Japan.

[Background] Benign adult familial myoclonic epilepsy (BAFME) is an autosomal dominant disorder characterized by infrequent epilepsy and myoclonic tremor. The causative genes for BAFME have not been identified. [Methods] Fifty three families including 93 patients were enrolled. We performed SNP typing-based linkage analysis and whole genome sequence and RNA-seq analyses employing HiSeq2500. To search for expanded tandem repeats, we applied TRhist (*Bioinformatics* 2014). [Results] Parametric linkage analysis of four families revealed a linkage to 8q24, which has been shown as a candidate region of BAFME in Japanese families. We found conserved haplotypes among the six families, which enabled us to narrow the candidate region to 134 kb. An intensive search of the whole genome sequence data revealed TTTCa and TTTTA repeat expansions in intron 4 of *SAMD12*, which were found exclusively in the patients in the 50 families. In a family, a different repeat configuration in *SAMD12* was revealed that also included TTTCa repeat expansions. RNA foci consisting of UUUA repeats, but not of UUUUA repeats, were observed in neurons of the autopsied brains. RNA-seq analysis of brain revealed increased coverages in introns upstream of the mutation, suggesting abortive transcription. Applying TRhist to RNA-seq data, we found short reads filled with [UUUA]_{exp} in autopsied brains with BAFME. In addition, reads filled with altered repeat motifs ([CUUA]_{exp} and [GUUA]_{exp}) were also found in autopsied brains, raising the possibility of modification of RNA sequences. No ubiquitinated inclusions were found in autopsied brains. In the two families with the clinical diagnosis of BAFME, we did not detect expanded TTTCa and TTTTA repeats in *SAMD12*. Hypothesizing that the same repeat motifs in other genes might be involved in the two families, we searched for accumulated short reads filled with TTTCa and TTTTA repeats employing TRhist, and identified repeat expansion mutations in introns of *TNRC6A* in a family and of *RAPGEF2* in the other family. [Discussion] The findings that the same expanded repeat motifs in the three independent genes lead to BAFME phenotypes emphasize the role of TTTCa repeat expansions in BAFME, presumably through RNA-mediated toxicity, which was further supported by the presence of RNA foci without ubiquitinated inclusions. The search for RNA binding proteins that specifically bind to the expanded repeat motifs is underway.

1127F

A novel disease gene *DTMPK* identified in 2 siblings with mitochondrial DNA depletion syndrome. C. Lam¹, W.L. Yeung². 1) Pathology, The University of Hong Kong, Hong Kong, China; 2) Department of Pediatrics and Adolescent Medicine, Alice Ho Miu Ling Nethersole Hospital, Hong Kong, China.

We described 2 brothers, aged 1 and 6 years, with microcephaly, hypotonia and severe intellectual disability (ID) and they shared common biochemical abnormalities including high serum lactate, pyruvate and alanine. The younger brother is now 12 months old (m.o.), who was a full term baby with low birth weight of 2.46 kg. The baby was noted to have congenital microcephaly, hypotonia and global delay at 6 m.o. He suffered from status epilepticus associated with febrile illness at 7 m.o. and subsequently had developmental regression, hypertonia, cortical blindness and dysphagia. CT brain showed no focal lesion while MRI brain revealed bilateral symmetric restricted diffusion at caudate heads, basal ganglia, thalami, corona radiata, cerebellar dentate nuclei and crus cerebri. Repeated MRI brain at 10 m.o. showed marked cerebral atrophy with bilateral subdural haemorrhagic effusion, and lactate peak in MR spectroscopy. At age of 12 m.o., he had no head control and was unable to roll over or reached out for objects. His 6-year-old elder brother also has congenital microcephaly (head circumference: 31 cm) and low birth weight of 2.1 kg. He also presented with hypotonia, global developmental delay and severe ID. He can now walk independently and speak few single words. His vision and hearing are normal and has no seizure attack all along. His serum lactate fluctuated from 2.0 - 3.5 mmol/L with slight elevation of alanine. MRI brain was unremarkable. His head circumference remained static in the past 3 years and is now 5 cm below the 3rd percentile. The 2 siblings were recruited to our Undiagnosed Diseases Program for whole exome sequencing (WES) analysis after informed consents. Two variants in the *DTYMK* gene were identified. The first variant is a frameshift mutation NM_012145.3:c.287_320del (p.Asp96Valfs*8) and the second one is a novel missense mutation c.295G>A (p.Ala99Thr). Segregation study showed the two variants were *in trans*. Deoxythymidylate kinase (thymidylate kinase), *DTYMK* is an essential gene in both *de novo* and salvage pyrimidine deoxyribonucleotide biosynthetic pathway. Knockdown of *DTYMK* gene would lead to reduced dTDP levels in the mitochondria. This suggests the vital role of *DTYMK* in mitochondria and explain the clinical and biochemical features seen in the siblings. Our novel findings suggested *DTYMK* is a new mitochondrial disease gene and should be included in the future gene panel analysis for mitochondrial DNA depletion syndrome.

1128W

Defining the natural history of cognition in type 3 Gaucher disease. A.M. Steward¹, E.A. Wiggs¹, T. Lindstrom^{1,2}, S. Ukwuan³, T. Roshan-Lali¹, E. Ryan¹, G. Lopez¹, R. Schiffman⁴, E. Sidransky¹. 1) NIH, Bethesda, MD; 2) University of Miami, Miami, FL; 3) Emory University, Atlanta, GA; 4) Baylor Research Institute, Dallas, TX.

Gaucher disease (GD) is an autosomal recessive disorder caused by mutations in the *GBA1* gene, leading to deficiency of the lysosomal enzyme glucocerebrosidase and subsequent buildup of substrate. Type 3 GD is characterized by varying degrees of chronic neurological involvement. The most common neurological feature is horizontal supranuclear gaze palsy, though manifestations range from intellectual delay to progressive myoclonic epilepsy, among others. No correlations exist between degree of neurologic involvement and *GBA1* genotype. Currently-approved therapies for GD do not affect neurological manifestations. We aimed to characterize the natural history of cognition in 33 individuals with type 3 GD who completed neuropsychological evaluations at the National Institutes of Health Clinical Center between 1982 and 2017. Individuals were evaluated between 1 and 18 times, using age-appropriate Wechsler Intelligence Scales over 1 to 20 years. Mean verbal, performance, and full-scale intelligence quotients (IQs) were 86.2, 79.3, and 81.0, respectively. Variation in all IQ domains was not linear with time, excluding regression and MANOVA as reliable statistical methods. The coefficient of variation (CV, standard deviation/mean) was used to characterize variation for each individual. Mean (sd) verbal, performance, and full-scale CVs were 0.08 (0.05), 0.09 (0.06), and 0.07 (0.05), respectively. Mixed effects regressions were used to determine if IQ was associated with clinical features, such as splenectomy and behavioral characteristics, among others. Models were optimized using backward step reduced fit with Satterthwaite degrees of freedom. Subject ID was included as a random effect to account for internal correlation of an individual's IQ scores. Splenectomy and autism were associated with verbal IQ; splenectomy, behavioral issues, ataxia, myoclonus, ADD/ADHD, and having an individualized education plan (IEP) were associated with performance IQ; splenectomy and behavioral issues were associated with full-scale IQ. These results indicate that IQ in type 3 GD varies over time, with no clear trajectory and average variation of roughly 2-13%. Changes outside this range should prompt evaluation of specific clinical features associated with that IQ domain. Finally, to use IQ as a clinical trial efficacy measure, multiple evaluations prior to intervention are needed to establish a reliable mean IQ and determine whether IQ changes can be attributed to that intervention.

1129T

Replication of MIR149 association with onset age and severity of CMT1A in a European population. F. Tao¹, G. Beecham¹, A. Rebelo¹, S. Blanton¹, L. Abreu², D. Rizzo², C. Kirk², X. Wu³, S. Feely², C. Verhamme⁴, M. Saporta², D. Herrmann⁵, J. Day⁶, C. Sumner^{8,9}, T. Lloyd^{8,9}, J. Li¹⁰, S. Yum¹¹, F. Baas¹², B. Choi¹³, D. Pareyson¹⁴, S. Scherer¹⁵, M. Reilly¹⁶, M. Shy³, S. Zuchner¹, *Inherited Neuropathy Consortium*. 1) Department for Human Genetics and Hussman Institute for Human Genomics, University of Miami, Miami, FL 33136, USA; 2) Data Management and Coordinating Center, Rare Diseases Clinical Research Network, Pediatrics Epidemiology Center, University of South Florida, Tampa, FL 33620, USA; 3) Department of Neurology, University of Iowa, Iowa City, IA 52242, USA; 4) Department of Neurology, Academic Medical Centre, 1105 AZ Amsterdam, The Netherlands; 5) Department of Neurology, University of Miami, Miami, FL 33136, USA; 6) Department of Neurology, University of Rochester, Rochester, NY 14642, USA; 7) Department of Neurology, Stanford University, Palo Alto, CA 94304, USA; 8) Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; 9) Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; 10) Department of Neurology, Wayne State University School of Medicine, Detroit, MI 48201, USA; 11) Division of Neurology, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; 12) Department of Clinical Genetics, Leiden University Medical Center, 2300 RC Leiden, The Netherlands; 13) Department of Neurology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul 06351, Korea; 14) Department of Clinical Neurosciences, IRCCS Foundation, C. Besta Neurological Institute, 20133 Milan, Italy; 15) Department of Neurology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA; 16) MRC Centre for Neuromuscular Diseases, UCL Institute of Neurology, Queen Square, London WC1N 3BG, UK.

Charcot-Marie-Tooth disease type 1A (CMT1A) is the most common form of inherited peripheral neuropathies. The genetic basis of CMT1A is a 1.5-Mb duplication on chromosome 17p11.2, which includes the causative gene *Peripheral Myelin Protein 22 (PMP22)*. In search for modifier genes and alleles, a common SNP rs2292832 in microRNA 149 (MIR149) was recently reported to be significantly associated with age at onset (AAO) and severity of CMT1A in an Asian population. However, the SNP has not been studied in other populations. In this study, we aimed to examine the association of rs2292832 with AAO and severity of CMT1A in a European population. The study included 644 CMT1A patients of European ancestry, which was confirmed by principal component analysis (PCA) using patients' genotyping data. Three clinical outcomes were tested: self-reported AAO of difficulty walking, CMT neuropathy score (CMTNS), and median motor nerve conduction velocity (MNCV). The analysis was performed using linear regression in PLINK. However, we did not observe any significant association of rs2292832 with the three phenotypes ($P = 0.1061$ for AAO, $P = 0.2444$ for CMTNS, $P = 0.3180$ for MNCV). We then divided the cases into early-onset group (AAO < 20 years) and late-onset group (AAO ≥ 20 years), and performed the association analysis using chi-square test. No significant difference was identified in allele frequencies between early and late-onset individuals ($P = 0.2580$, OR = 1.271). We also tested for disease severity as a binary outcome of severe cases (CMTNS > 10, MNCV ≤ 20) and mild cases (CMTNS ≤ 10, MNCV > 20). The analysis did not identify any significant association ($P = 0.6884$, OR = 0.9290 for CMTNS, $P = 0.7593$, OR = 0.9382 for MNCV). In conclusion, our study did not replicate the association of the MIR149 SNP in an independent European population. Since the minor allele frequency of rs2292832 differ among populations ($T = 0.637$ in East Asians and $T = 0.282$ in Europeans reported from 1000 Genomes), it is possible that the effect of rs2292832 is ethnic-specific. Additional studies in other populations are required to further validate the role of MIR149 in CMT1A.

1130F

Identification of pathogenic variants of *DMD* gene in Indian Backers Muscular Dystrophy: A comparison to the *DMD* spectrum. A. Anand¹, R. Tyagi¹, K. Sharma¹, C. Thakur¹, S. Kumar¹, A. Dalal², M. Faruq³, M. Mohanty⁴. 1) Neuroscience Research Lab, Department of Neurology, Post Graduate Institute of Medical Education and Research, Chandigarh, India; 2) Diagnostics Division, Center for DNA fingerprinting and Diagnostics, Hyderabad, India; 3) Genomics and Molecular Medicine Unit, Institute of Genomics and Integrative Biology, New Delhi, India; 4) Department of Neurosurgery, Postgraduate Institute of Medical Education and Research, Chandigarh, India.

Introduction: *DMD* is a widely studied gene responsible for a fatal inherited disorder Duchenne Muscular Dystrophy (DMD) and its milder variant Backers Muscular Dystrophy (BMD). DMD and BMD phenotypes are results of preservation of Open Reading Frame due to variation. Increased number of sporadic cases and the emergence of novel therapeutic approaches including exon skipping necessitates repurposing identification and characterization of pathogenic variants in DMD and BMD. **Methods:** A total of 100 DMD and 21 BMD subjects were recruited as per guidelines outlined by Institute Ethics Committee. Variations were detected by Multiplex Ligation-dependent Probe Amplification (MLPA) followed by capillary electrophoresis in the sequencing platform. Coafflyser.Net was used to obtain the relative ratios based on which type of variation is detected. Ratios between .70-1.30 were considered to show normal dosage quotient, 0.00 as deletion and 1.5-2.0 as duplication.

Results: Out of 21 BMD subjects 14/21 (66.66 %) showed deletion/duplication. Among these 11/14 (78.57 %) showed deletion between the hot spot region (Exon 45-55). 3/14 (21.42 %) showed deletion in the proximal region of the *DMD* gene. Deletion/Duplications were not detected in 33.33 % BMD cases. In contrast to BMD subjects, 80 % DMD subjects revealed deletion or duplication. 45/80 (56.25 %) variations in DMD cases, were amenable to Exon 51 skipping for the restoration of open reading frame and its conversion to BMD like phenotype. Skipping can potentially impact Central nervous system specific Dp160, Dp71, Dp40 isoforms leading to improved quality of life.

Conclusion: Spectrum of DMD pathogenic variants reveals an increased shift in the deletion/duplication profile. This data can be not only utilized to develop database but also for exon skipping interventions based on single exon or multi-exon administration of PMOs.

1131W

Wide range of phenotypic variability caused by missense variants in *NARS2* gene. L. Kalfon¹, S. Ben Harouch^{1,2}, H. Ladany^{1,2}, H. Mandel¹, T.C. Fallik-Zaccari^{1,2}. 1) Institute of Human Genetics, Galilee Medical Center, Nahariya, Israel; 2) The Azrieli Faculty of Medicine, Bar Ilan, Safed, Israel.

Combined oxidative phosphorylation deficiency-24 is an autosomal recessive mitochondrial disorder with a wide phenotypic variability. The disease is caused by bi allelic variants in the *NARS2* gene, encoding for the Mt-Asparaginyl-tRNA Synthetase. The protein belongs to a class of enzymes called aminoacyl tRNA synthetase (aaRS) that charge tRNAs with their cognate amino acids. Mt-aaRSs are synthesized in the cytosol and are imported to the mitochondria, taking a vital role in the mitochondrial translation machinery. In the current study we describe three patients originating from two unrelated families. Two affected siblings are from a consanguineous Muslim family. The older sister (18 years), presented with congenital sensorineural hearing loss, recurring seizures (from age 5 years) followed by profound developmental delay and regression. Her brother (age 3 years), has shown earlier and more severe clinical manifestation (onset of intractable seizures at 6 months) and developmental regression. Trio exome sequencing identified a homozygous missense variant in the *NARS2* gene (c.434T>G). Sanger sequencing and RFLP confirmed the variant segregation in the family in an autosomal recessive manner. The third proband is the youngest offspring in a consanguineous Druze family, who presented seizures (at age of 5 months), sensorineural hearing loss, truncal hypotonia and developmental regression. The proband deceased at 6 months of age. Trio exome sequencing revealed a novel homozygous missense variant in *NARS2* (c.500A>G) predicted to be deleterious. The two variants reported here are in proximity to a dimerization motif and the catalytic site, and therefore may affect the dimerization and thus the function of Mt-Asparaginyl-tRNA Synthetase. To the best of our knowledge, 12 missense variants in the *NARS2* gene were reported to date, associated with hearing loss, mitochondrial respiratory chain deficiency, microcephaly and Leigh syndrome. Our patients provide evidence for inter and intra familial phenotypic variability and expand the phenotypes associated with pathogenic genetic variants in *NARS2* gene.

1132T

Peripheral neuropathies: Application of targeted next generation sequencing for Lithuanian patients cohort. B. Burnyte^{1,2}, L. Ambrozaityte^{1,2}, K. Grigalioniene^{1,2}, A. Morkuniene^{1,2}, K. Baronas^{1,2}, V. Kucinskas^{1,2}, A. Utkus^{1,2}. 1) Inst of Biomed Sciences, Fac of Med, VU, Vilnius, Lithuania; 2) VUHSK, Vilnius, Lithuania.

Hereditary peripheral neuropathies (HPN) are an etiologically heterogeneous group of neurodegenerative disorders of the peripheral motor, sensory, and autonomic nerves. Molecular diagnosis of HPN is often very challenging. There is a considerable overlap of phenotypes, and variants in one gene can cause different entities. In this study, we have customised a panel of 150 HPN associated genes for next-generation sequencing (NGS). A cohort of 34 patients with a yet molecularly undiagnosed HPN with or without additional features were included in this study. The sequencing data was filtered for the variants meeting the set of quality requirements, having a population frequency less than 1% and coverage of at least 20x. Pathogenic or likely pathogenic variants in genes known to cause HPN were identified in 29% (10/34) of the patients. Five patients (15%) harbored novel pathogenic variants in HPN associated genes *AARS*, *MARS*, *DCTN1*, *GBN4* or *LITAF*. One patient (3%) with intellectual disability and axonal neuropathy had novel variant in *ATAD3A* gene. Additionally, two patients (6%) had variants in a novel candidate HPN gene *ARHGEF10*. Previously reported biallelic pathogenic variants were identified for four patients (12%); three of those in *HINT1* gene. Digenic inheritance was identified in one patient with heterozygous variants in *GNB4* and *DNM2* genes, where *GNB4* gene variant was novel. In a few patients variants of unknown significance were detected highlighting the dilemma in prioritizing them due to insufficient *in silico* predictions and other prognostic scores. In 56% of the cases further molecular studies were recommended (WES or WGS) for identifying new genes implicated in the pathogenesis of HPN. Therefore, application of the targeted gene panel for HPN can significantly enhance the effectiveness of genetic diagnosis. Identification of new pathogenic variation and characterization of associated phenotypes advances the knowledge of the pathogenesis of peripheral neuropathies.

1133F

Parkinson's disease age of onset GWAS: Defining heritability, identifying genetic loci, implicating synuclein mechanisms. C. Blauwendraat^{1,2}, K. Heilbron³, C.L. Valleria⁴, S. Bandres-Ciga¹, S.W. Scholz², The. 23andMe Research Team⁵, P. Heutink^{6,7}, J.M. Shulman^{7,8}, H.R. Morris^{9,10}, D.A. Hinds³, J. Gratten^{4,11}, P.M. Visscher^{4,11}, Z. Gan-Or^{12,13}, M.A. Nalls^{1,14}, A.B. Singleton¹, International Parkinson's Disease Genomics Consortium (IPDGC). 1) Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA; 2) Neurodegenerative Diseases Research Unit, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, USA; 3) 23andMe, Inc., 899 W Evelyn Avenue, Mountain View, CA, USA; 4) Institute for Molecular Bioscience, The University of Queensland, Brisbane, Australia; 5) Department for Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research, University of Tübingen, Tübingen, Germany; 6) German Center for Neurodegenerative Diseases (DZNE), Tübingen, Germany; 7) Departments of Molecular & Human Genetics and Neuroscience, Baylor College of Medicine, Houston, USA; 8) Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, USA; 9) Department of Clinical Neuroscience, UCL Institute of Neurology, London UK; 10) UCL Movement Disorders Centre, UCL Institute of Neurology, London, UK; 11) Queensland Brain Institute, The University of Queensland, Brisbane, Australia; 12) Department of Human Genetics, McGill University, Montreal, Quebec, Canada; 13) Montreal Neurological Institute, McGill University, Montreal, Quebec, Canada; 14) Data Tecnica International, Glen Echo, MD, USA.

Parkinson disease (PD) is the most common neurodegenerative movement disorder. PD is pathologically characterized by the loss of dopaminergic neurons in the substantia nigra, spreading to other brain areas, and α -synuclein (encoded by *SNCA*) protein aggregates. Increasing evidence supports an extensive and complex genetic contribution to PD. Previous genome-wide association studies (GWAS) have shed light on the genetic basis of risk for this disease. However, the genetic determinants of PD age of onset are largely unknown. Here we performed an age of onset GWAS based on 28,568 PD cases. We estimated that the heritability of PD age of onset due to common genetic variation was ~0.11, lower than the overall heritability of risk for PD (~0.27). We found two genome-wide significant association signals, *SNCA* and *TMEM175*, both of which are known PD risk loci. In addition, we identified that *GBA* coding variant carriers have a lower age of onset compared to non-carriers. Notably, *SNCA*, *TMEM175*, and *GBA* have all been shown to directly influence alpha-synuclein aggregation. Similar trends were seen for other loci implicated in alpha-synuclein aggregation pathways (e.g. *SCARB2*, *BAG3*) although not reaching genome-wide significance. Interestingly, other well-established PD risk loci such as *STK39*, *MAPT* and *RAB7L1* (PARK16) did not show a significant effect on age of onset of PD, implying different mechanisms for disease. Overall, the significantly associated variability is centered on the gene encoding alpha-synuclein and on variability in several other lysosomal proteins that have been shown to directly influence alpha-synuclein aggregation or clearance. These results highlight that therapies that target alpha-synuclein aggregation are more likely to be disease-modifying than therapies targeting other pathways.

1134W

Leveraging consanguinity in essential tremor families from Turkey to identify candidate genes. L. Clark^{1,2}, D. Robakis³, A. Tekinay⁴, C. Akbostancı⁵, F.N. Durmaz-Celik⁶, T. Ozcelik⁷. 1) Department of Pathology and Cell Biology, College of Physicians and Surgeons, Columbia University Irving Medical Center, New York, NY, 10032, USA; 2) Taub Institute for Research on Alzheimer's Disease and the Aging Brain, College of Physicians and Surgeons, Columbia University, New York, NY, 10032, USA; 3) Department of Neurology, Yale School of Medicine, Yale University, New Haven, CT, USA; 4) Institute of Materials Science and Nanotechnology, Bilkent University, Ankara, Turkey; 5) Faculty of Medicine, Department of Neurology, Ankara University, Ankara, Turkey; 6) Faculty of Medicine, Department of Neurology, Osmangazi University, Eskisehir, Turkey; 7) Department of Molecular Biology and Genetics, Bilkent University, Ankara, Turkey.

Background: Genetic analyses of consanguineous families or populations with endogamy is a powerful approach for gene identification and can help to unravel the underlying genetic defects in diseases that are genetically heterogeneous. Essential Tremor (ET), a disease whose hallmark feature is action tremor (i.e. tremor occurring during voluntary movements), is the most common cause of tremor in humans. Prevalence increases with age reaching ~20% during the ninth decade of life in the US. In Turkey, prevalence estimates are very similar to US estimates with the prevalence of ET in Mersin province, Turkey estimated to be 4% in individuals aged >40 years. **Objectives:** Here we have leveraged consanguinity and population endogamy in a unique cohort of ET families collected throughout Turkey at Bilkent University to identify candidate genes. **Methods:** As of 2018, the Turkish ET family study at Bilkent University consists of 160 ET families (443 collected samples), of which 51% are consanguineous or endogamous. Detailed clinical assessment of these families was performed and six families were selected for whole genome sequencing and genetic analyses. An integrated analyses including homozygosity mapping and WGS analyses (SNV/indel and CNV/SV) was performed. **Results:** We have identified a number of interesting candidate genes in these ET families which are currently being validated by functional studies. In one family with ET and parkinsonism, we identified compound heterozygosity for two heterozygous missense variants, predicted to be pathogenic by *in silico* tools and that are rare in populations databases (GnomAD), in the vacuolar protein sorting 13A gene (*VPS13A*) that co-segregated with the ET and parkinsonism phenotype. Homozygous or compound heterozygous mutations in *VPS13A* cause Chorea-acanthocytosis (ChAc, OMIM#200150), a rare autosomal recessive inherited syndrome characterized by a hyperkinetic movement disorder, progressive neurodegeneration and red cell acanthocytosis. ChAc can also manifest clinically as chorea, tremor, parkinsonism, dystonia, cognitive impairment, seizures and psychosis. **Conclusion:** Here, we show the utility of genetic studies in consanguineous pedigrees for genetically heterogeneous diseases such as ET. The candidate genes that we have identified may represent novel genetic causes of ET and represent possible gene targets for therapeutic intervention. This study is funded by the National Institutes of Health, R21 NS098930 (Clark).

1135T

Whole genome sequencing of a large Lewy Body dementia and Frontotemporal dementia cohort for genetic discovery: A public resource for the research community. C.L. Dalgard¹, C. Viollet², E. McGrath-Martinez³, C. Alba⁴, M. Sabir⁵, S. Ahmed⁶, Y. Abramzon⁷, R. Ferrari⁸, J. Hardy⁹, A. Chio¹⁰, O.A. Ross¹¹, M.D. Wilkerson¹², R. Gibbs¹³, A. Singleton¹⁴, B.J. Traynor¹⁵, S.W. Scholz¹⁶. 1) Anatomy, Physiology & Genetics, Uniformed Services University of the Health Sciences, Bethesda, MD, USA; 2) The American Genome Center, Uniformed Services University of the Health Sciences, Bethesda, MD, USA; 3) Neurodegenerative Diseases Research Unit, National Institute of Neurological Disorders and Stroke, National Institute of Health, Bethesda, MD, USA; 4) Laboratory of Neurogenetics, National Institute of Aging, National Institute of Health, Bethesda, MD, USA; 5) Department of Molecular Neuroscience, Institute of Neurology, University College London, London, UK; 6) Department of Neuroscience, University of Turin, Turin, Italy; 7) Department of Neuroscience, Mayo Clinic, Jacksonville, FL, USA.

Lewy body dementia (LBD) and frontotemporal dementia (FTD) are common complex neurodegenerative diseases with multifactorial etiologies. Familial cases of LBD and FTD suggest genetic predisposition for these diseases. Advances in whole genome sequencing technologies and computational resources have lead us to initiate a genomic profiling study for large cohorts of LBD and FTD patients to identify new genetic risk loci associated with dementia and provide raw data to publicly accessible repositories upon production. Here, we report a summary for whole genome sequencing of 6,761 samples at a mean depth of 35.5x [range: 23.8-42.5x], including 2,112 LBD cases, 2,592 FTD cases and 2,057 matched controls. We also describe detailed demographics of control subjects with matched age greater than 60 years, neurological assessments, and/or histopathological negative features after autopsy examination within this resource for multiple potential study designs and analyses by the research community. Whole genome sequencing was completed by PCR-free DNA library preparation from mechanically sheared 1 microgram gDNA input, followed by single library, single lane clustering on the Illumina cBot2 and paired-end 150bp sequencing on the Illumina HiSeq X platform. Sample-level data was analyzed on the seqEngine pipeline using the HAS2 package with Isaac Aligner and Strelka Germline Variant Caller for alignment and variant calling. Analysis metrics for within- and cross-sample contamination, heterozygous/homozygous SNV ratio, sex chromosome SNV ratio and counts, identity by descent replicate assessment were applied for quality control sample-level filter, with a dropout rate of 1.67%. Sequencing results were clustered using 54 quality metrics to identify and correct technical bias in relationship with sample assembly site, library preparation date, sequencing data and additional processing factors, yielding negligible observed batch effects. Rare variation (MAF < 0.05) burden across cohorts from this workflow is ~9,000 mean variants per 0.5mb window, including ~120 mean exonic rare variants per 0.5mb window. Raw sequencing data will be assembled on a cloud storage instance for replicate primary and secondary analysis in advance of rapid sharing to dbGaP and other public repositories. We anticipate that access and analysis of this resource, in combination with additional data resources will accelerate the pace of genetic discovery within the dementia research field.

1136F

Two Japanese familial cases of C9ALS/FTD complex in Fukuoka prefecture, Northern Kyushu Island, Japan. H. Furuya¹, A. Watanabe², N. Fujii². 1) Department of Neurology, Kochi Medical School, Kochi, Japan, Kochi, Japan; 2) Department of Neurology, Neuro-Muscular Center, NHO Omuta Hospital, Omuta, Fukuoka, Japan.

Amyotrophic Lateral Sclerosis (ALS) caused by expansion of (GGGGCC) repeats in *C9orf72(C9)* gene is extremely rare in Asian ethnic group including Japanese and its prevalence is almost 1/20 to 1/40 compared with caucasian. We had two familial Japanese C9ALS/FTD cases in Fukuoka prefecture, Northern Kyushu, Japan. Case 1. 56 years old male. He suffered character change and troubles in carrying his work three months before. Cerebral MRI image showed mild frontal lobe atrophy and cerebral blood flow SPECT revealed decreased cerebral blood flow in frontal area. After 1.5 years of onset, he showed dysphagia, muscle atrophy, weakness in extremities and bilateral pyramidal tract signs. He had under the control of mechanical ventilator due to severe respiratory failure. He died with empyema after 8 years from onset. Fronto-temporal cerebral atrophy progressed constant and he had diagnosed as bvFTD/ALS complexes. Gene analysis revealed expansion of repeats in *C9*. Case 2. 63 years old right-handed male noticed progressive motor aphasia and dementia. He was diagnosed as transcortical motor aphasia at age 64 and showed no neurological abnormality such as muscle weakness, parkinsonism or pyramidal tract sign except configuration apraxia. He had three sisters and one brother. His second sister died with early onset dementia and third sister suffered primary progressive dementia and epilepsy, his elderly brother suffered FTD with parkinsonism. All these members possessed expansion of repeats in *C9*. In Japan, C9ALS is rare, but it exists at a certain rate. Clinical phenotype differs in each cases.

1137W

Reduced penetrance of *SOD1* p.Ala90Val mutation in Finnish ALS patients. L. Kuuluvainen^{1,2}, K. Kaivola³, S. Mönkä^{2,4}, H. Laaksovirta^{3,5}, M. Jokela^{6,7}, S. Penttilä⁸, B. Udd⁶, M. Valori³, P. Pasanen^{4,8}, A. Paetau⁹, J. Schleutker^{4,8}, M. Pöyhönen^{1,2}, L. Myllykangas², P.J. Tienari^{3,5}. 1) Department of Clinical Genetics, Helsinki University Central Hospital, Helsinki, Finland; 2) Department of Medical Genetics, University of Helsinki, Helsinki, Finland; 3) Molecular Neurology, Research Programs Unit, University of Helsinki, Helsinki, Finland; 4) Turku University Hospital, Laboratory Division, Genetics and Saske, Department of Medical Genetics, Turku, Finland; 5) Department of Neurology, Helsinki University Hospital, Helsinki, Finland; 6) Neuromuscular Research Center, Tampere University Hospital and University of Tampere, Tampere, Finland; 7) Division of Clinical Neurosciences, Turku University Hospital and University of Turku, Turku, Finland; 8) Institute of Biomedicine, University of Turku, Turku, Finland; 9) Department of Pathology, University of Helsinki, Helsinki, Finland.

Superoxide dismutase 1 (*SOD1*) mutations are the second most common cause of familial amyotrophic lateral sclerosis (FALS) accounting for approximately 15-20% of FALS and approximately 1-2% of sporadic ALS. We examined the clinical and neuropathological phenotype of an ALS patient with the *SOD1* mutation c.269C>T, p.Ala90Val and analyzed its frequency and carriers' clinical characteristics in two ALS cohorts from Finland. The index patient was screened for mutations using whole exome sequencing. Two ALS cohorts were studied, Helsinki cohort (n=300) with whole genome sequencing data and Turku cohort (n=153) from samples sent to a diagnostic laboratory for *SOD1* sequencing. The mutation carriers' clinical data was examined from patient records. Allele-sharing analysis was conducted using single nucleotide polymorphism markers. Eight patients (1.8%) were found to carry the *SOD1* p.Ala90Val heterozygous mutation. Common clinical features included a long disease course and sensory neuropathy. Neuropathologically changes consistent with motor neuron disease without TDP43 pathology were observed. Allele-sharing analysis suggested a common founder. None of the patients had family history of ALS. All patients had additional potentially contributing variants in other genes. Lack of family history and finding other potentially contributing variants suggests the importance of oligogenic mechanisms for disease penetrance.

1138T

Sequencing of the *GFAP* canonical transcript may not be suitable for the diagnosis of adult-onset Alexander disease. F. Vairo¹, N. Bertsch², E. Klee³, R. Gavrilova⁴. 1) Center for Individualized Medicine, Mayo Clinic, Rochester, MN., USA; Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA; 2) Department of Clinical Genomics, Mayo Clinic, Rochester, MN, USA; 3) Center for Individualized Medicine, Mayo Clinic, Rochester, MN, USA; Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA; Department of Clinical Genomics, Mayo Clinic, Rochester, MN, USA; Department of Biomedical Informatics, Mayo Clinic, Rochester, MN, USA; 4) Center for Individualized Medicine, Mayo Clinic, Rochester, MN, USA; Department of Clinical Genomics, Mayo Clinic, Rochester, MN, USA; Department of Neurology, Mayo Clinic, Rochester, MN, USA.

Alexander disease (AD) is an autosomal dominant progressive leukoencephalopathy caused by heterozygous mutation in the glial fibrillary acidic protein (*GFAP*) gene. Based on the age at presentation, onset of symptoms can range from infancy to adulthood, with the infantile form being most common and the adult onset form accounting for one third of cases. The adult form is characterized by variable symptoms, most commonly bulbar, but also may present with pyramidal, cerebellar, dysautonomia, sleep disturbance, and seizures. The diagnosis is typically confirmed via Sanger sequencing of the *GFAP* gene, as 98% of the individuals have detectable single nucleotide or small indels pathogenic variants. Here, we report on an adult-onset case with radiographic features and positive pathological findings in an affected family member suggestive of Alexander disease, who had negative *GFAP* Sanger sequencing in a clinical/CLIA laboratory.

1139F

Whole exome analyses discover risk factors for familial dystonia. R. Yadav^{1,2,3}, C.T. Russo², S. Bressman⁴, R. Rachel Saunders-Pullman¹, K. Simon-yan⁵, M. Hutchinson⁶, A. Lang⁷, N. Sharma², M.E. Talkowski^{1,2,3}, L.J. Ozelius². 1)) Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA; 2) Department of Neurology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA; 3) Program in Medical and Population Genetics and Stanley Center for Psychiatric Research, Broad Institute, Cambridge, MA, USA; 4) Department of Neurology, Mount Sinai Beth Israel Medical Center, New York, NY, USA; 5) Department of Otolaryngology, Massachusetts Eye and Ear Infirmary, Boston, MA, USA; 6) Department of Neurology, St Vincent's University Hospital Dublin, Dublin, Ireland; 7) Division of Neurology, Toronto Western Hospital, Toronto, ON, Canada.

Dystonia is a heterogeneous group of neurological movement disorders characterized by sustained or intermittent muscle contractions causing abnormal, often repetitive, movements and/or postures. Both body regions affected and age of onset vary among the dystonia types. There is a significant genetic contribution to many forms of dystonia, though incomplete penetrance is a prominent feature that often confounds gene discovery efforts, particularly when using population-based references. Here, we sought to explore differences in coding variant burden between familial and sporadic forms of dystonia. We sequenced 554 individuals affected with various forms of dystonia of unknown etiology, as well as 49 unaffected family members. Sporadic cases represented 335 samples while 219 cases had a family history (FH) of dystonia. The FH cohort consisted of individuals from 39 multiplex pedigrees with two or more affected individuals sequenced. All variants were called using GATK haplotype caller and Scalpel, followed by variant annotation and comparison of rare variants filtered by minor allele frequency < 0.001 in gnomAD, CADD score >20 and missense z-score >3.09 or if the variation is found in a gene with pLI score >0.9. Here we focus on the variant burden in FH positive (FH+) dystonia (219 cases, inclusive of 86 cases from 39 multiplex pedigrees). We observed pathogenic variants that segregated with dystonia in two genes, *ATP1A3* and *TUBB4A*, thus explaining 5% of FH+ dystonia in 39 multiplex families; no other known pathogenic variants were detected. In the remaining FH+ cases (n=133), multiple rare and predicted deleterious variants were observed in previously defined dystonia genes: *ATP1A3*, *ANO3*, *COL6A3*, *SGCE*, *TH*, *THAP1*, *CIZ1*, *PRRT2*, *PRKRA*, *TOR1* and *KMT2B*, accounting for an additional 17.7% of genetic risk of dystonia. In *ATP1A3*, R756H and T613M have both been previously identified as pathogenic and segregated with disease in the families. The *TUBB4A* variant (Q424H) segregated in a family with 5 affected members with isolated dystonia. *ANO3* variants S95L and V568D both segregated with disease. No other family members were available for testing the variants in *SGCE*, *THAP1*, *CIZ1* and *TOR1A* while segregation analysis in not-sequenced family members is in progress for variants in *PRRT2* and *KMT2B*. Genome wide association analysis for predicted deleterious variants in sporadic dystonia are ongoing with ancestry-matched controls from large population-based cohorts.

1140W

Exome first: An efficient approach to diagnosing inherited white matter disorders.

P. Kaur Bhullar¹, M. Hebbar¹, A. Shrikiran², R. Bhat Y², S. Nam-pootheri³, S. J Patil⁴, S. Sharma⁵, S. Bijarnia Mahay⁶, S. Pai⁷, A. Kaur⁸, K. C Rakshith⁹, N. Kamath¹⁰, S. Bajaj¹⁰, A. Kumble¹¹, R. Shetty¹², R. Shenoy¹³, B. Banerjee¹⁴, M. Kamte¹⁵, B. Shantharam Baliga¹, L. Edward S Lewis¹⁶, A. Moirangthem¹, K. M Girisha¹, A. Shukla¹. 1) Department of Medical Genetics, Kasturba Medical College, Manipal Academy of Higher Education, Manipal, India; 2) Department of Paediatrics, Kasturba Medical College, Manipal Academy of Higher Education, Manipal, India; 3) Department of Paediatric Genetics, Amrita Institute of Medical Sciences, Kochi, India; 4) Division of Genetics, Mazumdar Shaw Medical Center, Narayana Health City, Bangalore, India; 5) Department of Paediatrics, Lady Hardinge Medical College, New Delhi, India; 6) Institute of Medical Genetics and Genomics, Sir Ganga Ram Hospital, New Delhi, India; 7) Department of Neurology, Kasturba Medical College, Mangalore, Manipal Academy of Higher Education, Manipal, India; 8) Department of Paediatrics, Postgraduate Institute of Medical Education and Research, Chandigarh, India; 9) Department of Paediatrics, Kasturba Medical College, Mangalore, Manipal Academy of Higher Education, Manipal, India; 10) Jaslok Hospital and Research Centre, Mumbai, India; 11) Department of Paediatrics, Indiana Hospital and Heart Institute, Mangalore, India; 12) First neuro hospital, Mangalore, India; 13) Department of Paediatrics, K.S.Hegde Hospital, NITTE University, Mangalore, India; 14) Department of Paediatrics, Manipal Hospitals, Bangalore, India; 15) Department of Paediatrics, Jawaharlal Nehru Medical College, Belgaum, India; 16) Department of Neonatology, Kasturba Medical College, Manipal Academy of Higher Education, Manipal, India.

Inherited white matter disorders can be subdivided into leukodystrophies and genetic leukoencephalopathies. Leukodystrophies involve primary defect in myelination whereas genetic leukoencephalopathies are caused due to secondary involvement of myelin. At present, variations in approximately 150 genes are known to cause inherited white matter disorders. Owing to the phenotypic and genetic heterogeneity of these conditions whole exome sequencing (WES) was employed as first line genetic testing. We recruited sixty-one individuals from fifty-six families with clinical and radiological profile suggestive of an inherited white matter disorder. Of these, thirty-eight were males and twenty-four were females. The age range was from new born to twenty-five years. Nine families (16.07%) presented with hypomyelination while other white matter pathologies were noted in the forty-seven remaining families (83.93%). WES was performed for index patients of fifty-two families and trio (index patient and parents) was performed for four families. Overall, final diagnosis was established for forty families (71.42%). A total of 38 pathogenic variants were identified in this study, of which 25 (65.78%) variants are novel. Four families (10%) were diagnosed with leukodystrophies and thirty-six families (90%) were diagnosed with genetic leukoencephalopathies. Among the genetic leukoencephalopathies were inborn errors of metabolism (n=17, 47.23%), mitochondrial encephalopathies (n=7, 19.45%), congenital muscular dystrophies (n=3, 8.34%) and other neurodegenerative disorders (n=9, 25%). Three novel leukoencephalopathies were identified along with the causative genes, *ISCA1*, *AIMP2* and *TFIP11*. Extremely rare conditions namely ribose 5-phosphate isomerase deficiency (MIM #608611), leukodystrophy, progressive, early childhood-onset (MIM #617762) and epileptic encephalopathy, early infantile, 35 (MIM #616647) were identified. Founder effect was noted on three instances including Multiple mitochondrial dysfunctions syndrome 5 (MIM# 617613), Aicardi-Goutieres syndrome 3 (MIM# 610329) and Maple syrup urine disease, type 1A (MIM# 248600). This is an ongoing study with the aim of further delineation of phenotypes and molecular mechanisms of inherited white matter abnormalities in the underrepresented Indian population. The study would help in definitive genetic counseling and prenatal diagnosis as a direct benefit for families with inherited white matter disorders.

1141T

Proband-only exome sequencing in small families with late onset hereditary neuropathy.

M.K. Ndugga-Kabuye¹, D.H. Chen², K.G. Gwathmey³, M.O. Dorschner⁴, J. Wolff¹, T.D. Bird^{1,2}, W.H. Raskind¹. 1) Division of Medical Genetics, University of Washington, Seattle, WA; 2) Department of Neurology, University of Washington, Seattle, WA; 3) Department of Neurology, University of Virginia, Charlottesville, VA; 4) Center for Precision Diagnostics, University of Washington, Seattle, WA.

Introduction: The primary hereditary neuropathies demonstrate wide clinical and genetic heterogeneity with more than 100 genes implicated thus far. They include Charcot Marie Tooth (CMT) and the hereditary sensory and autonomic neuropathies (HSAN/HSN). Many patients remain undiagnosed even after molecular diagnostic testing for the most common etiologies. **Methods & Results:** We performed proband-only exome sequencing in three adults from three unrelated families with a clinical diagnosis of axonal neuropathy with symptom onset after the age of 50yo. All three individuals had prior negative multi-gene testing for neuropathy. We identified three novel variants in genes that are rare causes of autosomal dominant forms of hereditary neuropathy; *MARS* (p.Gly570Glu), *SPTLC1* (p.Gln47fs) and *ATL3* (c.510+1G>T). The *MARS* (p.Gly570Glu) variant had a CADD score of 27.3, a GERP score of 5.66 and was not present in gnomAD or ClinVar. The *SPTLC1* frameshift variant (p.Gln47fs) had an allele frequency of 1.792E-5 in gnomAD and is not in ClinVar. Through ClinVar we identified an additional affected patient with a *SPTLC1* (p.Leu344fs) frameshift variant and an axonal neuropathy with symptom onset after 50yo. The *ATL3* (c.510+1G>T) variant had a CADD score of 24.2, a GERP score of 5.14 and is not present in gnomAD or ClinVar. We confirmed all three candidate variants via Sanger sequencing. **Discussion:** *SPTLC1* encodes serine palmitoyltransferase 1, and is implicated in HSN1A (MIM # 162400). Our finding of two frameshift variants in *SPTLC1* lead us to propose that loss of function/haploinsufficiency can be a mechanism of disease in HSN1A. *ATL3* encodes GTPase Atlantin-3 and is implicated in HSN1F (MIM # 61563). The two previously implicated variants in HSN1F were missense variants. Splice site variants have not been reported in HSN1F. *MARS* encodes methionyl-tRNA synthetase and is implicated in CMT2U, (MIM # 616280). The three previously described pathogenic variants in *MARS* have also been missense variants. **Conclusions:** In four families with primary hereditary neuropathy with symptom onset after the age of 50 we report the first splice site variant in *ATL3* (c.510+1G>T), the first two frameshift variants in *SPTLC1* (p.Gln47fs & p.Leu344fs), and the fourth pathogenic variant in *MARS* (p.Gly570Glu). Proband-only exome sequencing can be an effective molecular diagnostic modality in late onset primary hereditary neuropathy.

1142F

6-year-old girl with autism and *de novo* missense variant in *NECTIN3*. J. Pappas¹, R. Rabin¹, R. Person², J. Juusola². 1) NYU School of Medicine, New York, NY; 2) GeneDx, Gaithersburg, MD.

The association of autism with *de novo* pathogenic variants has been established in hundreds of genes (Basu SN et al., 2009). We present a 6-year-old girl with autism and *de novo* missense variant in *NECTIN3* (previously known as *PVRL3*), a gene that has never been associated with autism in the medical literature. Our case was born full term 3 kg weight and 48 cm length to 21 year old primigravida without any complications. She was developing typically until about 30 months when her speech and social development arrested. Neurodevelopmental assessment diagnosed autism. Brain MRI, EEG and ophthalmology evaluations were normal. The parents were not consanguineous and they had four typically developing children after our patient. The clinical examination of our case at age 6 revealed no apparent malformations, growth at about the 20th centile for all anthropometrics and impulsive behaviors with loss of personal space and inappropriate laughter. Chromosome analysis, SNP microarray analysis using the Affymetrix Cytoscan HD platform and DNA methylation test for Angelman was reported normal. Trio exome sequencing analysis revealed the p.Arg503Gly (AGG>GGG): c.1507 A>G *de novo* novel variant in exon 6 in the *NECTIN3* (NM_015480.2). This variant is not observed in gnomAD (Lek et al., 2016). It is a non-conservative amino acid substitution, likely to impact secondary protein structure as these residues differ in polarity, charge and size. In-silico analyses, including protein predictors and evolutionary conservation, support that this variant does not alter protein structure/function (MutationTaster, Schwarz JM et al., 2014). Our case introduces a possible association of *NECTIN3* and autism. The possible association is supported by the localization of *NECTIN3* in synapses. *NECTIN3* is an immunoglobulin-like protein that connects with afadin to actin cytoskeleton and organizes adherens junctions. The nectin-afadin system co-localizes with the cadherin-catenin system at synapses (Mizoguchi A et al., 2002). The possible association of *NECTIN3* and autism is also supported by the family reported by Sundaram SK et al., 2011. This 3-generation family studied by exome and included 7 individuals with Tourette syndrome. These 7 individuals had 4 novel nonsynonymous variants, one of the variants in *NECTIN3*; none of the unaffected in the family had all four variants. Additional cases may prompt functional studies and possible delineation of the role of *NECTIN3* in neurodevelopment.

1143W

Evaluation of a 12 lysosomal storage disorder gene panel in a Parkinson's disease cohort. L.J. Racacho^{1,2}, D. Durie^{1,2}, D.A. Grimes^{3,4}, M. Schlossmacher^{3,4,5}, D.E. Bulman^{1,2,5,6}. 1) Children's Hospital of Eastern Ontario Research Institute, Ottawa, ON, Canada; 2) Newborn Screening Ontario, Ottawa, ON, Canada; 3) Division of Neurology, Ottawa Hospital, Ottawa, ON, Canada; 4) Neuroscience Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada; 5) Department of Cellular and Molecular Medicine, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada; 6) Department of Pediatrics, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada.

A genetic link between Parkinson's disease (PD) and Gaucher's disease (GD) across various ethnic populations have provided valuable insight on lysosome pathology in PD. GD is an autosomal recessive condition caused by mutations in the beta-glucocerebrosidase gene (*GBA*) which lead to either the reduction or elimination of enzymatic activity. Carriers of *GBA* mutations have an increased risk of developing PD. The extent of rare variant burden in other lysosomal storage disorder genes for PD is largely unknown. As part of an evaluation process for a diagnostic test of lysosomal storage disorders, we implemented the Accel-Amplicon NGS Panel for 12 classical lysosomal storage disorder genes, *ABCD1*, *ARSB*, *GAA*, *GALC*, *GALNS*, *GBA*, *GLA*, *GLB1*, *GUSB*, *IDS*, *IDUA* and *SMPD1* to assess rare variant burden in 223 individuals diagnosed with idiopathic PD. The one tube solution which incorporates both PCR-based enrichment and dual-barcoded library construction provided a quick and adaptable solution to work within our clinical setting. In our preliminary assessment of *GBA*, we found that missense variants p.(Asp365Lys) and p.(Thr408Met) were detected at allele frequencies of 4.26% and 1.12%, respectively. Previous reports have shown that these variants reduce beta-glucocerebrosidase activity in carriers and are classified as risk alleles for PD.

1144T

DMD gene specific shorter dystrophin isoform's (Dp140) association to the working memory: A novel IBLT strategy based validation. R. Tyagi¹, P. Aggarwal², M. Mohanty², V. Dutt², A. Anand³. 1) Neuroscience Research Lab, Department of Neurology, Post Graduate Institute of Medical Education and Research, Chandigarh, India; 2) Indian Institute of Technology, Mandi, Himachal Pradesh, India; 3) Department of Neurosurgery, Postgraduate Institute of Medical Education and Research, Chandigarh, India.

Introduction: DMD, the largest gene in the human genome; encoding the dystrophin protein, is associated with a fatal neuromuscular disorder, Duchenne Muscular Dystrophy. Majorly, deletions and duplications are involved in the pathogenesis of DMD. Along with progressive muscle wasting, DMD has also been associated with cognitive and neuropsychological abnormalities specifically due to brain-specific shorter dystrophin isoforms. **Methodology:** A total of 100 DMD subjects were recruited for identification of the spectrum of pathogenic variants in DMD gene. PCR based Multiplex Ligation-dependent probe amplification was performed to detect deletions and duplications. The absence of short dystrophin isoform, Dp140 termed as Dp140 -ve, was predicted based on variant location. General Intelligence was assessed by Malin's Intelligence Scale for Indian Children. Verbal, working memory, and executive functioning were assessed by RAVLT, Digit Span, and Stroop color and word task; respectively. A novel computational strategy based on Instance Based Learning Theory (IBLT) was adopted on Adaptive Control of Thought-Rational cognitive architecture to model human performance. **Result:** Eighty out of 100 DMD subjects harbored deletion (72%) or duplication (8%), out of which 64 were predicted for the absence of Dp140 isoform. Five cases were found to have mild (4/5) and moderate (1/5) mental retardation; and DMD variants in these cases were predicted to affect Dp140 isoform. Assessment of cognitive and neuropsychological profiles of the Dp140-ve group revealed alterations in general intelligence, verbal and visual memory, executive functioning. When Dp140 +ve and Dp140 -ve group were compared with control group, Dp140 -ve group performed worse in RAVLT Trial-1 ($p=0.048$), Trial-5 ($p=0.024$), learning capacity ($p=0.046$), immediate recall ($p=0.031$), delayed recall ($p=0.012$), commission ($p=0.018$), LTPR ($p=0.016$), RAVLT-Memory Efficiency Index ($p=0.007$), all measures of stroop task and digit span task. IBLT based modeling revealed DMD groups' non-reliance to recency indicating poor working memory capacity, and lower activation of instance retrieval resulted in increased response time. **Conclusion:** Working memory was found to be affected in the DMD and its association with Dp140 isoform was observed. Modeling strategies can be devised to understand dynamic decision processes. Interventions to improve the working memory may be used as a management of cognitive alterations.

1145F

Depletion of the mRNA export factor GANP underlies axonal neuropathy and intellectual disability. R. Woldegebriel¹, E. Ylikallio^{1,2}, M. Sainio¹, L. Mäenpää³, C. Bonnemann³, S. Donkervoort³, D. Bharucha-Goebel³, M. Walsh⁴, Z. Stark⁴, M.-J. van den Boogaard⁵, P. Isohanni^{1,6}, T. Lönnqvist⁶, H. Tyynismaa^{1,7}. 1) Research Programs Unit, Molecular Neurology, University of Helsinki, Helsinki, Finland; 2) Clinical Neurosciences, Neurology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland; 3) Neuromuscular and Neurogenetic Disorders of Childhood Section, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, USA; 4) Murdoch Children's Research Institute, Victoria, Australia; 5) Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands; 6) Department of Child Neurology, Children's Hospital and Pediatric Research Center, University of Helsinki and Helsinki University Hospital, Helsinki, Finland; 7) Department of Medical and Clinical Genetics, University of Helsinki, Helsinki, Finland.

Nuclear mRNA export defects have been recognized as an important cause of human diseases from cancer to neurological disease. We recently identified a new recessive disease gene *MCM3AP* to underlie early-onset axonal neuropathy and mild intellectual disability in patients from five families. Others reported three more families with a similar phenotype and biallelic *MCM3AP* variants. *MCM3AP*-related neuropathy is axonal and predominantly motor, with less pronounced sensory findings. The disease is progressive leading to loss of ambulation in childhood, or in milder cases progression is slower and ambulation is preserved in early adulthood. The most consistent additional finding is intellectual disability. *MCM3AP* encodes for the nuclear mRNA export factor GANP (germinal associated nuclear protein), which forms a scaffold for proteins of the TRanscription-EXport-2 (TREX-2) complex in the nuclear envelope, with a proposed role in exporting selective mature mRNAs from the nucleus or in regulating export of stress-related transcripts. The identified homozygous *MCM3AP* variants all cluster in the Sac3 (suppressor of actin) domain, which is implicated in mRNA binding, whereas the compound heterozygous mutations are mostly located outside the Sac3 domain. Early onset and severe disease course is more typical for the compound heterozygous mutations, whereas a more benign disease associates with the homozygous Sac3 mutations. We studied skin fibroblasts of patients who had *MCM3AP* mutations outside the Sac3 domain, and found a marked reduction in GANP immunostaining in the nuclear envelope and by Western blotting. Ultra-deep mRNA sequencing of fibroblasts from five patients confirmed *MCM3AP* depletion on transcriptional level and identified common differentially expressed genes. Our results suggest that defective mRNA binding as well as an overall severe reduction in GANP protein may compromise mRNA export as the pathogenic mechanism. Using cultured motor neurons differentiated from human induced pluripotent stem cells, we showed that GANP is a nuclear envelope protein also in neurons. Our current efforts are focused on understanding the role of GANP in nuclear-cytoplasmic mRNA transport in patient-specific motor neuron models.

1146W

A novel *ABCD1* gene mutation unveil a Colombian female carrier with adrenomyeloneuropathy and a family history of adrenoleukodystrophy. F. Suarez, M. Manotas, A. Marroquin, R. Quero, A. Paredes. Pontificia Universidad Javeriana, Bogotá D.C., Colombia.

X-linked Adrenoleukodystrophy (X-ALD [MIM 300100]) is the most prevalent peroxisomal disorder in hemizygous males occasionally causes neurological symptoms in female carriers. This disorder is caused by mutations of the *ABCD1* gene located at Xq28 which encodes for the peroxisomal membrane protein ALDP, an essential transmembrane transporter of very long-chain fatty acids (VLCFA>22). Thus, when defective leads to elevated VLCFA levels in plasma and tissues, causing a broad clinical spectrum. In males, it manifests as adrenocortical insufficiency, myelopathy and cerebral demyelination, whereas in female carriers it presents as adrenomyeloneuropathy with heterogeneous severity, ranging from mild hyperreflexia and impaired vibration perception to severe spastic paraparesis. Although elevated VLCFA concentrations in plasma suggests the disease, up to 15% of female carriers can have normal VLCFA levels, and therefore *ABCD1* gene mutational analysis should be executed for confirmation. Our aim is to report a novel mutation associated with the disease in a female carrier with a familial history of X-ALD. We present the case of a 51-year-old female with a six-year history of weakness, spasticity, frequent falls and urinary incontinence who presented to genetics clinic with worsening symptoms. She was previously assessed by neurology under the suspicion of tropical spastic paraparesis, but all infectious etiologies were discarded as well as optic neuromyelitis. Given the negative work up, a demyelinating disease was suspected. Biochemical parameters were evaluated, showing abnormal VLCFA quantification, with increased C26:0 levels and C24:C22 ratio and suggesting a carrier state for X-ALD. In a second interview, she reported a positive family history of her father, who was diagnosed with multiple sclerosis for similar symptoms. Moreover, a cousin on her father's side had amyotrophic lateral sclerosis (ELA) and her nephew died at the age of 22, with a clinical diagnosis of X-ALD and a *ABCD1* mutational analysis showing an uncertain significance variant (VUS), c.2002A>G (p.Thr668Ala). Thereafter we evaluated our patient and found she was heterozygous for the same mutation and with current mutational analyzes, this variant is predicted to disease causing. This case exhibit the first time this variant is found in a patient with the disease phenotype and underlines the relevance of evaluating the family history in women presenting with neurological symptoms.

1147T

***CYP2C9* alleles are associated with patient-centered prescribing outcomes in a real-world patient population undergoing phenytoin treatment.** A.E. Fohner¹, K.K. Thai¹, D. Ranatunga¹, A.T. Jelanian², B.L. Lawson¹, V.X. Liu¹, C.A. Schaefer¹. 1) Division of Research, Kaiser Permanente Northern California, Oakland, CA; 2) Department of Neurology, Kaiser Permanente Northern California, Walnut Creek, CA.

Phenytoin is an anticonvulsant medication prescribed > 4million times each year. The Clinical Pharmacogenetics Implementation Consortium recommends lowering phenytoin dose in patients with reduced function *cytochrome P450 2C9* (*CYP2C9*) alleles due to increased risk for side effects. However, the presumed benefits of this recommendation have not been tested in a large real-world healthcare setting, where we hypothesized that reduced function *CYP2C9* alleles would be associated with increased side effects and decreased medication adherence. We identified 1070 members of a large, integrated healthcare delivery system in Northern California who filled an outpatient phenytoin prescription between 1/1/1996 and 8/31/2017 and who had *CYP2C9* genotyping. From the electronic health record, we extracted patient demographics, phenytoin utilization, and phenytoin blood concentration test results. We categorized patients into whether or not they had at least one reduced function *CYP2C9* allele (*2, *3, *5, *8, *11, and *12). We calculated medication adherence using an algorithm validated in a diverse patient population. We performed chart reviews on 413 patients to identify side effects within the first 100 days of treatment that were attributed to phenytoin. Our cohort was 48% male, 83% White, 4% Asian, 5% Black, and 7% Hispanic. It included 22,122 phenytoin prescriptions and 10,245 blood test results. We identified 340 (32%) patients with, and 730 (68%) patients without, reduced function *CYP2C9* alleles. Patients with reduced function alleles had higher phenytoin blood concentrations at their first test (mean 12.7 +/- 7.3) compared to those without (9.22 +/- 5.3; p-value < 0.001). Adjusting for age, gender, and race/ethnicity, patients with reduced function alleles were more likely to have poor medication adherence [odds ratio (OR) 1.66; 95% confidence interval (CI) 1.14–2.40; p-value 0.008], and a shorter time to discontinuing phenytoin (hazard ratio 1.28; 95% CI 1.07–1.52; p-value 0.007), compared to patients without reduced function alleles. Among those with chart reviews, patients with reduced function alleles were not more likely to have recorded side effects (p=0.16). Patients with reduced function *CYP2C9* alleles may have more, non-severe side effects due to higher blood concentrations, which could decrease medication adherence. Using genetic information to guide medication dose and choice could improve patient-centered prescribing outcomes and treatment efficacy.

1148F

Major intra-familial phenotypic heterogeneity and incomplete penetrance due to a CACNA1A pathogenic variant. J. Van Gils¹, C. Angelini¹, P.S. Jouk¹, D. Lacombe¹, S. Moutton², F. Riant³, G. Sole⁴, E. Tourmier-Lasserve⁵, A. Trimouille¹, M. Vincent⁶, C. Goizet^{1,6}. 1) Service de Génétique Médicale, CHU Bordeaux and Laboratoire MRGM, INSERM U1211, Univ. Bordeaux, Bordeaux, France; 2) Service de génétique médicale, Dijon, France; 3) Service de génétique moléculaire, Hôpital Lariboisière, Paris, France; 4) Service de neurologie médicale, Bordeaux, France; 5) Service de génétique moléculaire - Lariboisière, Paris, France; 6) Centre de référence neurogénétique, Service de génétique médicale, Bordeaux, France; 7) Service de génétique médicale, CHU de Grenoble; 8) Service de génétique médicale, CHU de Nantes.

The CACNA1A gene encodes a calcium-dependent voltage channel, localized in neuronal cells. Pathogenic variants in this gene are known to lead to a broad clinical spectrum including episodic ataxia type 2, spinocerebellar ataxia type 6, familial hemiplegic migraine, and more recently epileptic encephalopathy. We report a large family revealing dramatic variability of neurological manifestations associated with a CACNA1A missense pathogenic variant. The index case had early-onset epileptic encephalopathy with progressive cerebellar atrophy, although his mother and his great-grandmother suffered from paroxysmic episodic ataxia. His grandfather and great-grand-aunt reported no symptoms, but two of her sons displayed early-onset ataxia with intellectual disability. Two of her little daughters suffered from gait disorders, and also from epilepsy for one of them. All these relatives were carriers of the previously described heterozygous variant c.835C>T, p.(Arg279Cys) in CACNA1A (NM_023035.2). Pathogenic variants in CACNA1A may lead to several different phenotypes, including severe epileptic encephalopathy which was only recently described. We report here the first family leading to major clinical variability and incomplete penetrance. Our family highlights the difficulties to provide accurate genetic counselling concerning prenatal diagnosis regarding highly variable severity of the clinical presentation.

1149W

Usefulness of whole exome sequencing: De novo PURA gene pathogenic variant. E.M. Albino¹, J.F. Munoz-Bibiloni², I.M. Gracia-Vega³, F. Velez^{4,5}, S. Carlo^{6,7}, A. Santiago-Corrier^{8,9}. 1) School of Health Professions, Medical Sciences Campus, University of Puerto Rico, San Juan, PR; 2) Pediatric Neurology Section, San Jorge Children's Hospital, San Juan, PR; 3) Medical Technology Program, School of Health Professions, Medical Sciences Campus, University of Puerto Rico, San Juan, PR; 4) Genetics Division, Stanford University, Palo Alto, CA; 5) Department of Pediatrics, San Juan City Hospital, San Juan, PR; 6) Genetic Section, San Jorge Children's Hospital, San Juan, PR; 7) Department of Biomedical Sciences, Ponce Health Sciences University, Ponce, PR; 8) Department of Public Health, Ponce Health Sciences University, Ponce, PR; 9) Department of Pediatrics, Universidad Central del Caribe School of Medicine, Bayamon, PR.

GAL was referred to Genetics at 8 ½ months of age due to hypotonia and undescended testes. The infant had significant hypotonia and gross/fine motor delay at the time. Complete metabolic evaluations with a chromosomal microarray were normal. At 10 months of age the infant developed infantile spasms. Brain MRI showed non-specific bilateral white matter hypomyelination. Infantile spasms were improved with antiepileptics by 18 months of age. Whole exome sequencing was requested and a novel pathogenic variant was identified at the PURA gene, c.615_625 del (p.Ile206Argfs*84). Myoclonic and focal onset seizures have persisted but other seizure types, such as generalized tonic-clonic, tonic or atonic, can also occur. The PURA gene encodes Pur-alpha, a 322 amino acid protein with repeated nucleic acid binding domains that are highly conserved from bacteria through humans. Pur-alpha is a protein that binds specific DNA and RNA sequence elements. Human PURA, located at chromosome band 5q31, is under complex control of three promoters. The entire protein coding sequence of PURA is contiguous within a single exon. The C.615_625 del is a frameshift mutation located in exon 1 of the PURA gene. Other frameshift variants, which occur before and after this patient's variant, have been described in the medical literature as pathogenic and some of these results in similar frameshifts with early termination of the protein at the same residue. In neurons, Pura accompanies mRNA transcripts to sites of translation in dendrites. At 4 years of age, GAL has improved truncal stability, is able to roll over and sit on his own but without speech. The frameshift mutation described is consistent with a severe neuro-developmental phenotype that features hypotonia, and gross psychomotor delay with intractable seizures. The Pura protein is especially important for normal brain development. Mutations in PURA gene should be part of the genetic evaluation of infants with profound hypotonia, epileptic encephalopathy and intellectual disability. Study this de novo mutation is an important step toward the clinical and molecular characterization of this disease that has been associated with significant neurological outcome. While targeted gene panels are indicated for analyzing specific mutations, whole exome sequencing was an excellent approach in this case due to the spectrum of neurological manifestations.

1150T

***MAPT* p.V363I mutation is a rare cause of corticobasal degeneration.**

S. Ahmed^{1,2}, A. Pantelyat⁶, A. Butala³, M.S. Sabir⁶, L. Isperto⁵, E. Gelpi⁴, P. Pastor⁶, S.W. Scholz^{2,3}. 1) Laboratory of Neurogenetics, National Institutes on Aging, National Institutes of Health, Bethesda, USA; 2) Neurodegenerative Disease Research Unit, National Institute of Neurological Disorders and Strokes, National Institutes of Health, Bethesda, USA; 3) Department of Neurology, Johns Hopkins University Medical Center, Baltimore, Maryland, USA; 4) Neurological Tissue Bank, University of Barcelona-Hospital Clinic, IDIBAPS, 08036 Barcelona, Spain; 5) Neurology Service, Hospital Universitari Germans Trias, Pujol, Badalona, Spain; 6) Neurogenetics Laboratory, Division of Neurosciences, Center for Applied Medical Research, University of Navarra School of Medicine, Pamplona, Spain.

Background: Corticobasal degeneration (CBD) is a neurodegenerative tauopathy that presents with heterogeneous clinical features, including asymmetric parkinsonism, dyspraxia, aphasia and cognitive impairment. The genetics of CBD is poorly understood, but few reports exist linking mutations and risk variants in *MAPT* to this rare disease. **Methods:** We performed a genetic evaluation of patients with corticobasal syndrome (CBS) using the NeuroChip Illumina genotyping array followed by validation using Sanger sequencing. **Results:** We identified two patients who were heterozygous for the rare c.1087G>A (p.V363I) mutation in the highly conserved C-terminal region of *MAPT*. *In silico* pathogenicity predictions indicated that this variant is disease-causing, but literature review of previously reported cases was inconclusive. One of the two patients in our cohort was pathologically confirmed and demonstrated 4-repeat tau inclusions consistent with CBD. **Conclusion:** Our report describes two new CBS cases carrying the rare p.V363I *MAPT* mutation, one of which was pathologically confirmed. Taken together, our findings support the notion that this rare missense mutation is pathogenic.

1151F

Chemical suppressors of nonsense mutations as potential treatment for NGLY1-related disorder. A. Baradaran-Heravi¹, A. Balgi¹, F. Shidmoossavee², J. Tan², J. Jaquith², M. Roberge¹. 1) University of British Columbia, Vancouver, British Columbia, Canada; 2) The Centre for Drug Research and Development, 2405 Wesbrook Mall, Vancouver, British Columbia, Canada.

NGLY1-related disorder is a rare autosomal recessive disease characterized by developmental delay and neurological, hepatic and ophthalmological findings. It is associated with biallelic loss-of-function mutations in *NGLY1*, which encodes an enzyme, N-glycanase 1, involved in deglycosylation of glycoproteins, an essential step in the endoplasmic reticulum associated degradation (ERAD) pathway. Currently, there is no treatment for this disease. Nonsense mutations in *NGLY1* are the most frequent type of mutations in this disease. These mutations change an amino acid codon to a premature termination codon (PTC), resulting in truncated and non-functional protein. High concentrations of aminoglycosides can restore gene function by eliciting PTC readthrough but with low efficiency. We have recently identified an aminoglycoside that is a more potent PTC readthrough inducer than other currently known aminoglycosides. Also, using a yeast high-throughput screen, we have identified a class of compounds that potentiate PTC readthrough by aminoglycosides in mammalian cells. To assess PTC readthrough of *NGLY1* nonsense mutations, we generated HEK-293 cells stably expressing various *NGLY1* nonsense mutations. The aminoglycoside alone at high concentration induced barely detectable N-glycanase 1 production. However, combination of a low concentration of this aminoglycoside with an enhancer significantly increased PTC readthrough at most *NGLY1* nonsense mutations. This combination could also induce PTC readthrough at a *NGLY1* nonsense mutation in fibroblasts derived from a patient with *NGLY1* deficiency. Whether the newly synthesized N-glycanase 1 is enzymatically active remains to be elucidated. These findings open up the possibility of treating a subset of patients with *NGLY1* deficiency caused by nonsense mutations.

1152W

Genetic study of patients with lipoid proteinosis, a disease of medial temporal lobe calcification and skin manifestations. S. Geiler¹, L. Yousefian¹, S. Pajouhanfar¹, A.H. Saeidian¹, M. Daneshpazhooh², H. Mahmoudi², S. Sotoudeh², P. Mansouri², P. Petramfar¹, S. Zeinali³, J. Uitto¹, H. Vahidnezhad^{1,3}. 1) Thomas Jefferson University, Philadelphia, PA, USA; 2) Tehran University of Medical Sciences, Tehran, Iran; 3) Pasteur Institute of Iran, Tehran, Iran.

Lipoid proteinosis (LP, [OMIM #247100]), also known as Urbach-Wiethe disease, is a rare monogenic multisystemic disorder. Approximately, 250 LP cases exist worldwide. LP manifestations are a consequence of pathogenic mutations in the *ECM1* gene. A primary phenotypic presentation of this disorder is deposition of hyaline-like material in various tissues. In addition, there are medial temporal lobe calcifications which typically manifest within the: amygdalae, hippocampi, hippocampal gyrus and basal ganglia. We analyzed DNA from 18 unrelated families including 27 patients, which were referred to our study through their primary care physician or dermatologist. Prior to genomic analysis, we obtained informed consent, patient and family history, and a blood sample from each patient. Neurological imaging (e.g. CT scan) was utilized to determine if medial temporal lobe calcifications were present. The DNA was extracted from the blood samples and subjected to mutation analysis with PCR and Sanger sequencing. Within our cohort, we identified four previously reported mutations and three novel mutations in *ECM1*. The two primary recurrent mutations were *ECM1*: NM_004425: c.735_736delTG, p.C245* and c.507delT, p.R171Gfs*7 which were present in 10 unrelated families with a total of 13 patients. Within our cohort we observed a vast spectrum of neuropsychiatric symptoms and disorders. Specifically, we noted the following neuropsychiatric symptoms: epilepsy, hallucinations, memory deficit, mental retardation, social isolation, depression and others. With the continuous acquisition of LP patients we will monitor and assess for intracranial calcifications through imaging techniques. Identifying calcifications and neuropsychiatric symptoms may improve diagnostics and treatment of neuropsychiatric disorders within the LP community. In a holistic sense, addressing the neuropsychiatric phenotype of LP patients will provide a deeper comprehension of neuropsychiatric function that may be applied to the greater population.

1153T

The Adult NCL Gene Discovery Consortium (The Kufs Consortium). I. Jedličková¹, V. Stránecký¹, A. Přistoupilová¹, H. Hartmannová¹, K. Hodaňová¹, L. Nosková¹, H. Hůlková¹, J. Sikora¹, V. Barešová¹, M. Živná¹, L. Piherová¹, B. Dermaut², J. van den Aemeele², M. Cadieux-Dion³, P. Cossette³, K.L. Oliver⁴, S. Cotman⁵, S.E. Mole⁶, S.F. Berkovic¹, S. Knoch¹, *The Adult NCL Gene Discovery Consortium*. 1) Research Unit for Rare Diseases, Department of Pediatrics and Adolescent Medicine, First Faculty of Medicine and General University Hospital in Prague, Prague, Czech Republic; 2) Department of Neurology, Ghent University Hospital, Ghent B-9000, Belgium; 3) Centre de Recherche du Centre Hospitalier de l'Université de Montréal, University of Montreal, Canada; 4) Epilepsy Research Centre, Department of Medicine, University of Melbourne, Austin Health, Heidelberg, Australia; 5) Center for Human Genetic Research and Department of Neurology, Harvard Medical School, Massachusetts General Hospital, Boston; 6) MRC Laboratory for Cell Biology, Department of Genetics, Evolution & Environment and UCL Institute of Child Health, University College London, UK.

The adult forms of neuronal ceroid lipofuscinoses (ANCLs) remain the least understood among NCLs. Clinically, they are characterized by progressive dementia and/or epilepsy with cognitive decline and behavioural changes. The hallmark of ANCLs is accumulation of ceroid in brain tissue leading to loss of neurons and neurodegeneration. Genetically, ANCLs can be caused by mutations in following genes: *CLN6*, *CTSF*, *GRN*, *CLN1*, *CLN5*, *ATP13A2* and *DNAJC5*, which remains the only known dominant gene for ANCL. Within the frame of genetic studies of **the Adult NCL Gene Discovery Consortium** we performed exome sequencing in 22 families and sporadic cases evaluated as suspect to have ANCL based on clinico-pathological criteria defined by The Consortium. Up to now we confirmed ANCL diagnosis using exome sequencing in combination with Sanger and targeted resequencing in one Canadian family with autosomal dominant (AD) form of ANCL (*DNAJC5*, p.Y120YCCCLCCFN). We further defined genetic lesions in two cases resulting in an alternate diagnosis, namely one sporadic Italian case (**C19ORF12**, compound heterozygous mutations – neurodegeneration with brain iron accumulation, NBIA-4) and one large Belgian family with AD inheritance and heterogeneous neuropsychiatric phenotype (*C9ORF72* hexanucleotide repeat expansion – fronto-temporal dementia). With this respect we analysed and excluded the *C9ORF72* expansion in additional dominant cases from our study cohort. The annotated exome data set of our study cohort provided a list of highly suggestive candidate variants of unknown significance in functionally relevant genes – *CLN6*, *NPC1*, *GRN*, *CYP27A1*, *SPG11*. Our results demonstrate a poor correlation of primary diagnosis with defined genetic lesions in our probands. This is due to the rarity and heterogeneity in phenotype of ANCLs and neurodegenerative diseases in young adults in general. The correct evaluation of ceroid in tissue samples by pathologists is challenging and needs extensive experience.

1154F

Pathogenic missense variants in the histone acetyltransferase complex component gene *TRRAP* cause autism and syndromic intellectual disability. S. Ehresmann¹, B. Cogné^{2,3}, E. Beauregard-Lacroix¹, J. Rousseau¹, T. Besnard^{2,3}, T. Garcia¹, S. Petroski^{4,5}, S. Küry^{2,3}, P.M. Campeau^{1,6}, *TRRAP Consortium, CAUSES Study, Deciphering Developmental Disorder Study*. 1) CHU Sainte Justine, Université de Montréal, Montreal, Québec, Canada; 2) CHU Nantes, Service de Génétique Médicale, 9 quai Moncousu, Nantes, France; 3) l'Institut du Thorax, INSERM, CNRS, UNIV Nantes, 44007 Nantes, France; 4) Department of Medicine, The University of Melbourne, Melbourne, Australia; 5) AstraZeneca Centre for Genomics Research, Precision Medicine and Genomics, IMED Biotech Unit, AstraZeneca, Cambridge, UK; 6) Department of Pediatrics, University of Montreal, Montreal, QC H3T1J4, Canada.

Acetylation of the lysine residues in histones and other DNA-binding proteins play a major role in regulation of eukaryotic gene expression. Regulation of this process is controlled by histone acetyltransferases (HATs/KATs) found in multiprotein complexes that are recruited to chromatin by the scaffolding subunit TRansformation/tRanscription domain Associated Protein (*TRRAP*). *TRRAP* is evolutionarily conserved and is among the top five genes intolerant to missense variation. Through an international collaboration, 17 distinct de novo or apparently de novo variants were identified in *TRRAP* in 24 individuals. A strong genotype-phenotype correlation was observed with two distinct clinical spectra. The first is a complex, multi-systemic syndrome characterized by a wide range of intellectual functioning, including a number of individuals with intellectual disability (ID) and markedly impaired basic life functions and associated with various malformations of the brain, heart, kidneys, and genitourinary system. Individuals with this phenotype had missense variants clustering around the c.3127G>A p.(Ala1043Thr) variant identified in five individuals. The second spectrum manifested with autism spectrum disorder and/or ID and epilepsy. RNA sequencing analysis on affected individual-derived skin fibroblasts showed significant changes in expression of several genes implicated in neuronal function and ion transport. Importantly, we confirmed by quantitative PCR the upregulation expression of the potassium ion transporters *KCNJ8* and *ABCC9*, known to be involved in Cantú syndrome, with which there is some minimal phenotypic overlap. Thus, we describe here the clinical spectrum associated with *TRRAP* variants and suggest dysregulated ion transporters expression as part of the pathophysiology driving the disease.

1155W

Pilarowski-Bjornsson syndrome: A novel cause of speech apraxia and neurodevelopmental delay. G. Pilarowski^{1,2}, K. Lindstrom³, N. Robin⁴, N. Rudy⁵, K. Chandler^{5,6}, S. Banka^{5,6}, R. Bernier⁷, G. Bowman^{8,9}, H. Bjornsson^{2,10,11}. 1) Predoctoral Program in Human Genetics, Johns Hopkins University School of Medicine, Baltimore, MD; 2) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD; 3) Division of Genetics and Metabolism, Phoenix Children's Hospital, Phoenix, AZ; 4) Department of Genetics, University of Alabama at Birmingham, Birmingham, AL; 5) Manchester Centre for Genomic Medicine, St Mary's Hospital, Manchester University NHS Foundation Trust, Health Innovation Manchester, Manchester, UK; 6) Division of Evolution & Genomic Sciences, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK; 7) Department of Psychiatry & Behavioral Sciences, University of Washington, Seattle, WA; 8) T.C. Jenkins Department of Biophysics, Johns Hopkins University, Baltimore, MD; 9) Department of Biology, Johns Hopkins University, Baltimore, MD; 10) Faculty of Medicine, University of Iceland, Reykjavik, Iceland; 11) Landspítali University Hospital, Reykjavik, Iceland.

Variants in components of the epigenetic machinery have been implicated in many types of intellectual disability. Our group recently identified the chromatin remodeler, *CHD1*, as a cause of autism, speech apraxia, and developmental delay. *CHD1* is an ATP-dependent chromatin remodeler protein, which reorganizes DNA structure by sliding nucleosomes along the DNA. In previous work, we described six female patients with autism, speech apraxia, developmental delay, and facial dysmorphism harboring heterozygous missense variants in *CHD1*. We previously hypothesized a potential dominant negative mechanism of disease, considering the absence of neurological features in a patient with a copy number variant at the *CHD1* locus, although another possible mechanism could be haploinsufficiency with decreased penetrance. Here we report on four additional patients with variants in *CHD1*, two of which do not harbor missense changes. Patient 1 is a 3-year-old male who has limited expressive speech, daily seizures and autistic type behaviors. He harbors a missense variant (p.R827C) in a possible protein-protein interaction region of *CHD1*. Patient 2 is a 17-year-old male with a diagnosis of autism spectrum disorder and speech apraxia. Interestingly, this patient harbors a *de novo* frameshift variant in *CHD1* which is predicted to be haploinsufficient due to a premature stop codon. Alternatively, there may be incomplete activation of the nonsense mediated decay cascade, resulting in a truncated protein. Patient 3 is a 2-year-9-month old male who displays gross motor and expressive speech delay, macrocephaly, stereotypic behaviors, and minor dysmorphic features. This patient harbors a *de novo* single base substitution at the beginning of intron 16. Although this is not a coding variant, it is predicted to affect splicing, which could lead to an altered protein. Patient 4 is a female presenting with developmental delay, intellectual disability, and macrocephaly. She carries a missense variant (p.R568K) in a region of *CHD1* which may be an indirect stabilizer for DNA binding. In conclusion, we describe four additional patients with Pilarowski-Bjornsson syndrome, all of which demonstrate speech apraxia or delay and autism or autistic type behaviors. These data support the hypothesis that Pilarowski-Bjornsson syndrome leads to disease phenotypes through a haploinsufficiency type mechanism and is a genetically determined cause of expressive speech apraxia and autism.

1156T

Expanding the phenotype and recommendations for management in *KMT5B*-associated neurodevelopmental disorder. S. Sheppard¹, A. Vitobello², D. Li¹, H. Hakonarson¹, E. Bhoji¹, *KMT5B Consortium*. 1) Children's Hospital of Philadelphia, Philadelphia, PA; 2) Université de Bourgogne-Franche Comté, France.

Heterozygous mutations in *KMT5B*, a histone methyltransferase, cause an autosomal dominant neurodevelopmental disorder that has been described in eleven patients (Stessman et al Nature Genetics 2017, Faundes et al Am J of Hum Genet 2018). The initial publications have focused on the neurodevelopmental phenotype of these patients. We present a case series of 12 additional patients to further characterize non-neurodevelopmental features of this new syndrome and make recommendations for management. Similar to what was published, the most common feature in our cohort was developmental delay, ranging from mild speech delay with near normal cognitive function to severe global developmental delay. In terms of growth parameters, the majority of the patients have small stature and macrocephaly. There does not seem that there is an easily recognizable facial gestalt associated with this syndrome. About half carry an autism spectrum disorder diagnosis, have hypotonia, and seizures. Multiple patients had GU abnormalities (hypospadias, nephrolithiasis, hydronephrosis with ureterocele, and small kidneys) and GI issues (mostly feeding difficulties ranging from TPN dependence; difficulty swallowing; gastroparesis). Ophthalmologic findings were present in the majority of patients and included hypermetropia, astigmatism, strabismus, and megalocornea. Craniofacial anomalies included craniosynostosis and cleft palate. Dermatologic findings included café au lait spots, hemangioma, and acanthosis nigricans. Cardiac issues included congenital heart disease (VSD, PDA, ASD) and dysautonomia. Finally, there also is musculoskeletal involvement including slight body asymmetry, slight pectus excavatum, genu valgum, bilateral thumb adductus and hypermobile joints. As with all patients with a neurodevelopmental disorder, a child development specialist is important for helping the child maximize his or her potential. Parents and frequent caregivers should be advised of the seizure risk. If there are signs of seizures, there should be evaluation by EEG and neurologist. We also recommend a screening echocardiogram and renal bladder ultrasound as some patients have had congenital heart defects and renal anomalies. Parents should also be aware of feeding issues, as we suspect this may be unrecognized in some patients due to the average short stature. Referrals to other specialists should be made as needed.

1157F

***MAP1B* mutations cause intellectual disability and extensive white matter deficit.** G.B. Walters^{1,2}, O. Gustafsson¹, V.K. Eiriksdottir¹, A.B. Agustsdottir¹, G.A. Jonsdottir¹, S. Steinberg¹, A.F. Gunnarsson¹, M.I. Magnusson¹, U. Unnsteinsdottir¹, A.L. Lee¹, A. Jonasdottir¹, A. Sigurdsson¹, A. Jonasdottir¹, A. Skuladottir¹, L. Jonsson^{1,3}, M.S. Nawaz^{1,2}, P. Sulem¹, M. Frigge¹, A. Ingason¹, A. Love^{2,4}, G.L. Norddahl¹, M. Zervas¹, D.F. Gudbjartsson^{1,5}, M.O. Ulfarsson^{1,6}, E. Saemundsen^{2,7}, H. Stefansson¹, K. Stefansson^{1,2}. 1) deCODE genetics (Amgen), Reykjavik, Reykjavik, Iceland; 2) Faculty of Medicine, University of Iceland, Reykjavik, Iceland; 3) Department of Pharmacology, Institute of Neuroscience and Physiology, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden; 4) Department of Radiology, Landspítali University Hospital, Fossvogur, Reykjavik, Iceland; 5) School of Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland; 6) Faculty of Electrical and Computer Engineering, University of Iceland, Reykjavik, Iceland; 7) The State Diagnostic and Counselling Centre, Kopavogur, Iceland.

Discovery of variants in the coding sequence of genes that confer risk of intellectual disability (ID) and Autism spectrum disorder (ASD), is an important step towards a better understanding of the pathophysiology of these disorders. Whole genome sequencing (WGS) of 31,463 Icelanders uncovered a frame-shift variant (E712KfsTer10) in *microtubule-associated protein 1B* (*MAP1B*) that associates with ID/low IQ in a large pedigree (genome-wide corrected $P = 0.022$). Additional stop-gain variants in *MAP1B* (E1032Ter and R1664Ter) validated the association with ID ($P = 4.9 \times 10^{-4}$). Carriers have less white matter (WM) volume ($\beta = -2.1$ SD, $P = 5.1 \times 10^{-4}$), particularly in the corpus callosum (CC) ($\beta = -2.4$ SD, $P = 5.5 \times 10^{-10}$), corresponding to a WM and CC volume deficit of 24% and 47%, respectively. Furthermore, lower brain-wide fractional anisotropy ($P = 6.7 \times 10^{-4}$), in carriers, may reflect disordered or compromised axons. Here we describe a syndrome involving *MAP1B* loss of function (LoF) variants. *MAP1B* LoF variants are found at a frequency of 1 in 10,000 in public databases which provides an opportunity for further investigations of the role of *MAP1B* dysfunction in neurodevelopmental disorders.

1158W

Case report of a Colombian patient with Wiedemann-Steiner Syndrome (WDSTS). A. Paredes¹, J. Prieto^{1,2}. 1) Instituto de Genética Humana, Pontificia Universidad Javeriana, Bogotá D.C., Colombia; 2) Hospital de la Victoria.

Wiedemann-Steiner syndrome (WDSTS [MIM:605130]) is a very rare autosomal dominant genetic disorder characterized by pre- and post-natal growth deficiency, mild to moderate intellectual disability, hipertrichosis cubiti and distinctive facial appearance. Is caused by heterozygous mutation in the gen *KMT2A* (prevpreviously known as MLL) mapped on chromosome 11q23.3. The *KMT2A* gene is involved in histone modification, it encodes a transcriptional coactivator that plays an essential role in regulating gene expression during early development and hematopoiesis. Currently there are very few confirmed cases worldwide and in Colombia there is no official report of this syndrome. We report a case of an 12-year-old male referred for study facial dysmorphism, delayed psychomotor development and short stature; he was diagnosed at 2 years old with hypothyroidism, treated with levothyroxine until 5 years. On physical examination the patient presented short stature, intellectual disability, facial dysmorphism, narrow palpebral fissures, downslanting palpebral fissures, long eyelashes, bilateral ptosis, long philtrum, thin upper lip, high-arched palate, dysplastic ears and hairy elbows. Was requested G-banded karyotype 46,XY, CGH array normal, finally he was diagnosed with WDSTS based on identification a pathogenic heterozygote variant c.10837C>T (p.Q3613X) in the *KMT2A* gene by whole exome sequencing study. The mutation found in the patient has not been previously reported, is predicted to result in premature termination of the protein product, with the molecular diagnosis we compared the phenotype of the patient to anteriorly reported patients in the literature. The ability of clinicians to recognize distinctive facial features has played an important role in the diagnosis of genetic syndromes, nevertheless its low prevalence and variable clinical phenotype makes it difficult to recognize. As more and more patients with developmental disorders undergo whole exome sequencing to achieve a diagnosis, there will be increasing numbers of individuals identified with Wiedemann Steiner Syndrome.

1159T

An 8-year-old girl with Pitt-Hopkins syndrome: New features and differential diagnosis of developmental delay in Colombia. J. Prieto¹, M. Mantas¹, A. Paredes¹, R. Quero¹, J. Acosta². 1) Pontificia Universidad Javeriana, Bogotá D.C., Colombia; 2) Instituto de Ortopedia Infantil Roosevelt.

Pitt-Hopkins syndrome (PHS [MIM 610954]) is an emerging neurodevelopmental disorder falling within syndromic forms of intellectual disability and developmental delay, was originally described in 1978 in two unrelated patients with mental retardation, recurrent episodes of hyperventilation. Currently there are less than 500 confirmed cases worldwide, in Colombia there is no official report of this syndrome. The syndrome is caused by haploinsufficiency of the *TCF4* gene on chromosome 18q21, *TCF4* is involved in embryologic neuronal differentiation and phenotypically is similar to other syndromic such as Angelman syndrome, Rett syndrome, Mowat-Wilson syndrome and ATR-X syndrome. Studies have revealed a molecular variant in *TCF4* in 2% of patients clinically diagnosed with Angelman syndrome and in 1 of 81 children clinically diagnosed with Rett syndrome. We report a case of an 8-year-old girl with intellectual disability, global developmental delay, motor incoordination, hypotonia, stereotypical movements, absence seizures, dysmorphic features such as café au lait spots on scalp, blue scleras, flared nostrils, large mouth, widely spaced teeth with absence of hyperventilation or other breathing abnormalities. Differential diagnoses were discarded such as Rett syndrome and inborn error of the metabolism. She came under Genetic clinical observation at 4 years of age. The girl underwent genetic (karyotype, Comparative genomic hybridization) and neuroradiological (cranial CT and MRI) tests, which proved to be normal. Due to severe neurodevelopmental delay and facial dysmorphisms we completed the genetic investigations with whole-exome sequencing (WES), showed c.1191_1194delTCAT variant in the *TCF4* gene. This variant has not been reported previously as a pathogenic variant nor as a benign variant. This variant causes a frameshift starting with codon isoleucine 397, changes this amino acid to a Methionine residue, and creates a premature Stop codon at position 64 of the new reading frame, denoted p.I397Mfs*64. This variant is predicted to cause loss of normal protein function through protein truncation. This variant is a strong candidate for a pathogenic variant. At our knowledge, our patient is the first Colombia case of PHS diagnosed at molecular level. This syndrome should be considered for patients with neurodevelopmental delay and searching for *TCF4* mutations is highly recommended when others overlapping syndromes was excluded.

1160F

Evidence that biallelic loss-of-function variants in *FBXL3* cause developmental delay/intellectual disability, facial dysmorphism and short stature.

P. Makrythanasis^{1,2}, S.A. Paracha³, A. Serretti⁴, M.T. Sarwar⁵, J. Khan³, E. Falconnet⁶, E. Ranza⁶, M. Guipponi⁶, F.A. Santoni^{7,8}, A. Megarbane⁷, J. Ahmed⁸, R. Colombo⁹, M. Ansari⁹, S.E. Antonarakis^{1,10}. 1) Department of Genetic Medicine and Development, University of Geneva, Geneva, Switzerland; 2) Greek Genome Center, Biomedical Research Foundation of the Academy of Athens, Athens, Greece; 3) Institute of Basic Medical Sciences, Khyber Medical University, Peshawar, Pakistan; 4) Department of Biomedical and Neuromotor Sciences, University of Bologna, Bologna, Italy; 5) Service of Genetic Medicine, University Hospitals of Geneva, Geneva, Switzerland; 6) Department of endocrinology diabetes and metabolism, university hospital of Lausanne, Switzerland; 7) Institut Jerome Lejeune, Paris, France; 8) Institute of Clinical Biochemistry, Faculty of Medicine, Catholic University and IRCCS Policlinico Gemelli, Rome, Italy; 9) Center for the Study of Rare Hereditary Diseases, Niguarda Ca' Granda Metropolitan Hospital, Milan, Italy; 10) iGE3 Institute of Genetics and Genomics of Geneva, Geneva, Switzerland.

FBXL3 (F-Box and Leucine Rich Repeat Protein 3) encodes a protein that contains F-box and several tandem leucine-rich repeats (LRR) domains. *FBXL3* is part of the SCF (Skp1-Cullin-F box protein) ubiquitin ligase complex that functions as phosphorylation dependent degradation of the central clock protein cryptochromes (CRY1 and CRY2) by the proteasome and its absence causes circadian phenotypes in mice and behavioral problems. No *FBXL3* related phenotypes have been described in humans. By a combination of exome sequencing and homozygosity mapping, we analyzed two consanguineous families with intellectual disability and identified homozygous LoF (loss-of-function) variants in *FBXL3*. In the first family, from Pakistan, an *FBXL3* frameshift variant (NM_012158.2:c.885del:p.(Leu295Phefs*25)) was the only segregating variant in five affected individuals in two family loops (LOD score: 3.12). In the second family, from Lebanon, we identified a nonsense variant (NM_012158.2:c.445C>T:p.(Arg149*)). In a third patient from Italy, a likely pathogenic missense variant (NM_012158.2:c.1072T>C:p.(Cys358Arg)) was identified in homozygosity. Protein 3D modeling showed that Cys358 is positioned on 10th LRR motif forming a solenoid around CRY2, and that the Cys358Arg change is predicted to influence the binding with CRY2 by destabilizing the structure of the *FBXL3* protein suggesting that this variant may also be LoF. The eight affected individuals from the three families presented with a similar phenotype. They have intellectual disability, developmental delay, short stature and mild facial dysmorphism, mainly large nose with bulbous tip. The phenotypic similarity and the segregation analysis suggest a link of *FBXL3* with syndromic autosomal recessive developmental delay/intellectual disability.

1161W

Expanding the phenotype of MED13L-associated mental retardation and distinctive facial features with or without cardiac defects; MRFACD. MJ. Hajianpour¹, K. Hunt², A. Wattad¹, K. Xixis¹. 1) East Tennessee State University, Johnson City, TN; 2) Holston Medical Group, Kingsport, TN.

Proband is a 4-year-old male with severe developmental delay and no speech, intellectual disability (ID), autism, seizures, ASD, VSD, PDA, renal cyst with hydronephrosis, and dysmorphic features: plagiocephaly, asymmetric face, downslanted palpebral fissures, left exotropia, small/ low-set ears, high nasal bridge, thin upper vermillion, tented upper lip, high arched palate, retrognathia, moderate pectus excavatum, clinodactyly of 5th fingers, simian crease, syndactyly between the 2nd & 3rd toes, and 4th & 5th toes (to middle of the toes), and sacral dimple. He has had G-tube feeding due to swallowing incoordination. Proband has no sibling. Mother is 24 yo; with a history of depression, Asperger syndrome, OCD, and obesity, but normal intelligence. Father is 22 yo; with bipolar disorder. Chromosome microarray, metabolic disorders screen, molecular studies for Angelman/ Rett syndrome panel (22 genes), and neurodevelopmental disorders panel (196 genes) were inconclusive. The Whole Exome Sequencing showed only a heterozygous missense mutation (p.P1526T; c.4576C>A) of MED13L gene, which was interpreted as a VUS, but likely the cause of the patient's features. Mother tested negative for MED13L alteration. The MED13L gene on chromosome 12q24.21 is involved in early embryonic development of heart and brain. Pathogenic mutations are associated with autosomal dominant mental retardation and distinctive facial features with or without cardiac defects; and with AD transposition of the great arteries (TGA); dextro-looped 1. Patients may show global developmental delay, moderate to severe ID, and absent speech. Most cases are de novo. Penetrance is reduced, but low-level parental mosaicism has been reported. Reported Features (proband's features are marked with *): Micrognathia*, Bulbous nasal tip*, Abn philtrum*, Low-set ears*, Strabismus*, Abn palmar creases*, Upslanted palpebral fissures, Truncal hypotonia*, CHD (dextro-looped TGA, conotruncal defects, and VSD*). Some patients: Low birth weight*, talipes equinovarus*, Seizures*, Motor delay*, ataxia, behavioral problems*. Less common: Autism*, Limb contractures, camptodactyly*, ventriculomegaly, global cerebral atrophy, Feeding difficulties*, sleep problem*, macroglossia, macrocephaly. Proband's features Not reported: Down-slanted PF (upslanted PF in reports), high nasal bridge, thin upper vermillion, retrognathia, high-arched palate, pectus excavatum, ASD, PDA, renal cyst, clinodactyly of 5th fingers, and 2-3 & 4-5 toes syndactyly.

1162T

Novel mutation in *KDM6A* gene associated with Kabuki syndrome type 2: Expanding literature on a rare condition. *E.M.G. Gershon¹, M.B. Byler¹, N.D. Dosa¹, W.B. B², R.L.L. Lebel¹.* 1) Upstate Medical University, Syracuse, NY; 2) Baylor College of Medicine, Atlanta, GA.

Kabuki Syndrome, also known as Kabuki make-up Syndrome (OMIM #147920), is associated with dwarfism, long palpebral fissures with eversion of the lower eyelid, skeletal anomalies, depressed nasal tip, intellectual disability, and a wide variety of other dysmorphic features. Most cases (56-76%) are associated with dominant mutations in the *KMT2D* gene (OMIM #602113), which maps to chromosome 12q13.12. *KMT2D* codes for a lysine specific methyltransferase. The remaining cases are referred to as Kabuki 2 (OMIM #300867), and are instead associated with dominant mutations in the *KDM6A* gene (OMIM #300128) which maps to Xp11.3, or they are not found to have mutation in either gene. These mutations can present with either familial inheritance or (more often) as a *de novo* event. *KDM6A* encodes a histone demethylase that interacts with the *MLL2* gene which in turn is also associated with the syndrome. Our patient, now 7 years old, presented at 3 years of age with global developmental delay, hearing deficits, poor vision, long palpebral fissures and hypotonia. He was born at 36 weeks via Cesarean Section due to breach presentation. The pregnancy was only complicated by diet controlled gestational diabetes. Whole Exome Sequencing revealed a novel mutation in the *KDM6A* gene: c.3109C>T (p.Q1037X). The mutation is thought to be *de novo* since it is absent in maternal lymphocytes, although gonadal mosaicism cannot be excluded. This case serves to expand the sparse literature available on Kabuki 2.

1163F

Causative novel *USP9X* variants in two Japanese patients with X-linked intellectual disability. *Y. Tsurusaki¹, K. Kuroda², H. Murakami³, Y. Enomoto¹, Y. Kimura¹, Y. Yamanouchi³, E. Kondoh⁴, K. Ouchi², M. Masuno^{3,4}, K. Kurosawa².* 1) Clinical Research Institute, Kanagawa Children's Medical Center, Yokohama, Japan; 2) Division of Medical Genetics, Kanagawa Children's Medical Center, Yokohama, Japan; 3) Genetic Counseling Program, Graduate School of Health and Welfare, Kawasaki University of Medical Welfare, Kurashiki, Japan; 4) Department of Pediatrics, Kawasaki Medical School, Kurashiki, Japan. Intellectual disability (ID) affects about 1% of the worldwide population. An IQ score of less than 70 is characteristic. Mutations in more than 100 genes have been found associated with X-linked ID. In 2014, hemizygous mutations in the *USP9X* were described as a cause of X-linked ID in males. *USP9X* encodes an Ubiquitin Specific Peptidase 9. *Usp9x* knockout mouse show reduction in both axonal growth and neuronal cell migration. To identify the pathogenic mutations for two unrelated Japanese patients with ID, we performed targeted resequencing using a TruSight One Sequencing Panel. Genomic DNA libraries were sequenced using MiSeq with 151-bp paired-end reads. Mapping to human genome hg19 was performed using Burrows-Wheeler Alignment. Aligned reads were processed by Picard to remove PCR duplicates. Variants were called using the Genome Analysis Toolkit, and annotated by ANNOVAR. Patient 1 is 4 year-old girl who is the second child of healthy and nonconsanguineous parents. She was delivered at 37 weeks gestation with a birth weight of 2,278 g (-1.1 SD), length 47cm (-0.1 SD), and occipital frontal circumference 30.5 cm (-1.5 SD). Her dysmorphic features consisted of upswept and curly hair, facial asymmetry, prominent forehead, bitemporal narrowing, short palpebral fissures, prominent nose with flared ala nasi, smooth philtrum, thin upper lip, full cheeks, dysplastic ears, tapering fingers, and Blaschko lines. Patient 2 is 1 year-old girl who is the first child of nonconsanguineous parents. The proposita was delivered at 41 weeks and 1 day gestation. Birth weight was 2,968 g (-0.5 SD), length 48.3 cm (-0.9 SD), and head circumference 34.2 cm (0.4 SD). The main clinical manifestations included epicanthus, telecanthus, short columella, depressed nasal tip, bilateral low-set and posteriorly rotated ears, overfolded helices, bifid uvula, umbilical hernia, and bilateral overlapping toes (T23, T45). Targeted resequencing identified two heterozygous mutations in *USP9X* [c.4068_4072del (p.Leu1357Tyrfs*12) and c.1201C>T (p.Arg401*)], respectively. Our study demonstrated that *USP9X* mutations can cause ID and the diagnostic utility of next generation sequencing technologies is obvious.

1164W

Triplet repeat primed PCR (TP-PCR) test for fragile X syndrome: Accuracy and efficiency evaluation as an alternative to conventional diagnostic methods. H. Gu, M.J. Kim, D. Yang, J.Y. Song, S.I. Cho, S.S. Park, M.W. Seong. Department of Laboratory Medicine, Seoul National University Hospital, Seoul, Korea, Republic of.

Backgrounds: The diagnosis of Fragile X syndrome (FXS) was based on fragment analysis (FA) and Southern blot (SB). Recently, large CGG expansion was confirmed by triplet repeat primed PCR (TP-PCR) to facilitate full mutation (FM) screening. We compared FA, SB and TP-PCR results of our hospital to evaluate the accuracy among the tests and efficiency of diagnostic workflow. **Methods:** From November 2013 to March 2018, a total of 462 patients were examined using FA, TP-PCR test as screening test and if necessary, using SB to be confirm the diagnosis. **Results:** Among the total patients, there were 9 FM, 12 premature (PM), 8 intermediate (INT) and 433 normal alleles (NA) patients diagnosed with previous workflow. The TP-PCR results in all the diagnosed groups were in good agreement with the diagnostic criteria according to RN. Of the FM group, 7 were unable to amplify by FA, so FA was confirmed through SB. 2 females with NA in FA were identified as FM carriers by detecting FM allele in SB. TP-PCR detected 200 or more RNs in all cases of FM, particularly both NA and FM in 2 FM carriers. In the PM group, TP-PCR results were consistent with the results from performing both FA/SB and FM alone. It was 8 cases in which observed only homozygous allele in FA, but heterozygous alleles in TP-PCR. In all INT cases, TP-FA and TP-PCR results were consistent in both homozygous and heterozygous alleles. In NA cases, FA and TP-PCR results showed about 94.5% agreement (409/433). **Conclusion:** The analyzed results showed that the agreement between FA and TP-PCR was very high, even TP-PCR showed higher sensitivity to heterozygous females or high RN. Hence it can be the first choice screening method to substitute for FA. Since FM results of TP-PCR are consistent with those of SB, it is possible to perform a reasonable workflow by reducing unnecessary SB. It is expected that the expanded use of TP-PCR will bring more positive and effective workflow improvement to the diagnosis of FXS in the future.

1165T

Paralog studies for disease gene identification: DDX and DHX genes. I.S. Paine¹, J. Posey¹, S.N. Jhangian², E. Karaca³, G. Lyon^{4,5}, M. Pinelli⁶, S. Rosenheck⁴, R. Kleyner⁴, T. Marmorale⁴, M. Yoon⁴, K. Wang⁷, R. Robinson⁸, G. Cappuccio⁹, A. Magli¹⁰, N. Brunetti-Pierri¹¹, Z. Coban-Akdemir¹², J. Hui¹³, W.L. Yeung¹⁰, B.K.Y. Wong¹¹, I. Glass¹², D. Lessel¹³, I.B. Van den Veyver¹¹, R. Sutton¹, D.M. Muzny^{1,2}, R.A. Gibbs^{1,2}, J.R. Lupski^{1,2,14}. 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas, USA; 3) Department of Genetics, University of Alabama at Birmingham, Birmingham, Alabama, USA; 4) Stanley Institute for Cognitive Genomics, Cold Spring Harbor Laboratory, NY, USA; 5) Utah Foundation for Biomedical Research, Salt Lake City, UT, USA; 6) Department of Translational Medicine, University of Naples "Federico II" Italy & Telethon Institute of Genetics and Medicine (TIGEM), Pozzuoli, Italy; 7) Raymond G. Perelman Center for Cellular and Molecular Therapeutics, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; 8) Department of Pediatric Ophthalmology, University of Salerno, Italy; 9) Department of Pediatrics, Prince of Wales Hospital, The Chinese University of Hong Kong; 10) Department of Pediatrics and Adolescent Medicine, Alice Ho Miu Ling Nethersole Hospital, Hong Kong; 11) Department of Obstetrics and Gynecology, Baylor College of Medicine and Duncan Neurological Research Institute, Texas Children's Hospital, Houston, Texas, USA; 12) Division of Genetic Medicine, University of Washington, Seattle, Washington, USA; 13) Institute of Human Genetics, University Medical Center Hamburg-Eppendorf, Hamburg, Germany; 14) Department of Pediatrics, Baylor College of Medicine, Houston, Texas, USA.

Intellectual disability (ID) affects 0.5-3% of the general population with 60% of cases remaining without a known cause. Many cases diagnosed by molecular testing reveal variants that have arisen *de novo*. Recent genetic screens in drosophila have identified 165 genes important for nervous system development (Yamamoto *et al.* 2014). Importantly, Yamamoto *et al.* noted that homozygous lethal drosophila genes with more than one homolog in humans were eight times more likely to be associated with human disease. It follows that gene families with a large number of paralogs in humans are more likely to be associated with disease if one member has been linked to disease already. The DEXD/H-box RNA helicase family represents one such gene family, with 58 members, all of which share 8-9 conserved amino acid motifs that make up the helicase core domain. *De novo* missense variants in the conserved motifs of *DHX30* were recently linked to a neurodevelopmental disorder characterized by intellectual disability (ID), developmental delay (DD), and gait abnormalities. We have identified a paralog of *DHX30*, *DHX37* as a candidate disease gene for a similar neurodevelopmental disorder. Five patients with homozygous (2), compound heterozygous (1), and *de novo* (1 confirmed, 1 suspected) missense variants in *DHX37* share a phenotype of ID/DD, brain anomalies, seizures, hypotonia, and vertebral anomalies. All the variants are rare, not present in the Thousand Genomes Project Database (TGP) or ExAC with the MAF for two variants in gnomAD calculated to be 8.3e⁻⁶, and have CADD scores >29. The variants also occur at residues that are predicted to be conserved by PhyloP and result in deleterious changes by four different prediction algorithms (PolyPhen2, MutationTaster, SIFT and LRT). Additionally, two other paralogs, *DHX16* and *DDX54*, represent candidate disease genes associated with neurodevelopmental phenotypes characterized by ID/DD, hypotonia, and brain anomalies (4 cases). The variants in *DHX16* and *DDX54* occur as *de novo* (3) and compound heterozygous (1) and are also rare, not present in TGP or ExAC, and have CADD scores >23. Together these cases implicate the DEXD/H-box gene family in both dominantly and recessively inherited ID/DD phenotypes and highlight the potential for more than one disease mechanism underlying *DHX30*-related disorders.

1166F

Whole genome sequencing analysis of rare variants in 475 individuals with neurodevelopmental disorders. A. Sanchis-Juan^{1,2}, K. J. Carss^{1,2}, C. Armirola^{1,2}, K. Megy^{1,2}, K. Low^{1,2}, C. E. French³, D. Grozeva⁴, E. Dewhurst^{1,2}, J. Stephens^{1,2}, N. Gleadall^{1,2}, L. Stefanucci^{1,5,6}, K. Stirrups^{1,2}, C. Penkett^{1,2}, G. Ambegaonkar⁷, M. Chitre³, D. Josifova⁸, M. Kurian^{9,10}, A. Parker¹¹, J. Rankin¹², E. Reid¹³, E. Wakeling¹⁴, E. Wassmer¹⁵, G. Woods^{4,11}, C. NIHR-BioResource Rare Diseases², W. H. Ouwehand^{1,2}, FL. Raymond^{2,4}. 1) Department of Haematology, University of Cambridge, Cambridge, Cambridgeshire, United Kingdom; 2) NIHR BioResource, Cambridge University Hospitals NHS Foundation Trust, Cambridge Biomedical Campus, Cambridge, United Kingdom; 3) Department of Paediatrics, University of Cambridge, Cambridge, United Kingdom; 4) Department of Medical Genetics, Cambridge Institute for Medical Research, University of Cambridge, Cambridge, United Kingdom; 5) National Health Service Blood and Transplant (NHSBT), Cambridge Biomedical Campus, Cambridge, United Kingdom; 6) BHF Centre of Excellence, Division of Cardiovascular Medicine, Addenbrooke's Hospital, Cambridge Biomedical Campus, Cambridge, United Kingdom; 7) Child Development Centre, Addenbrookes Hospital, Cambridge University Hospitals NHS Foundation Trust, Cambridge, United Kingdom; 8) Guy's and St Thomas' Hospital, Guy's and St Thomas' NHS Foundation Trust, London, United Kingdom; 9) Molecular Neurosciences, Developmental Neurosciences, UCL Great Ormond Street Institute of Child Health, London, United Kingdom; 10) Department of Neurology, Great Ormond Street Hospital for Children NHS Foundation Trust, London, United Kingdom; 11) Addenbrookes Hospital, Cambridge University Hospitals NHS Foundation Trust, Cambridge, United Kingdom; 12) Royal Devon and Exeter NHS Foundation Trust, Exeter, United Kingdom; 13) Department of Clinical Genetics, Addenbrookes Hospital, Cambridge University Hospitals NHS Foundation Trust, Cambridge, United Kingdom; 14) North West Thames Regional Genetics Service, London North West Healthcare NHS Trust, Harrow, United Kingdom; 15) Birmingham Children's Hospital, Birmingham Women's and Children's NHS Foundation Trust, Birmingham, United Kingdom.

Despite the significant progress in unravelling the origins of neurodevelopmental disorders (NDD), up to 50% of affected individuals remain without a genetic diagnosis. Here we demonstrate the power of using short-read whole-genome sequencing (WGS) to identify pathogenic variants in 475 affected individuals with a NDD, as part of the NIHR-BioResource Rare Diseases research study. We identified pathogenic variants in 150/475 (31%) of the individuals, and an additional 101/475 (21%) had a variant of uncertain significance. 90% of them were single-nucleotide variant or indels in coding regions of known NDD-associated genes. However, WGS provided power to detect three additional categories of variants: structural, intronic and mitochondrial variants. Structural variants comprised 19 deletions, 3 duplications, 2 inversions and 1 partial uniparental isodisomy. We also found 2 complex structural variants: a *de novo* deletion-inversion-duplication overlapping *HNRNPU* gene, and a *de novo* duplication-inversion-duplication overlapping *CDKL5* gene. Nanopore long-read WGS was used as complementary tool to resolve the genetic architecture of the latter. Pathogenic intronic variants were also identified, 5 of them being within the splice region of NDD-associated genes. We also identified a deep intronic variant in *TSC2* that caused Tuberous Sclerosis type 2 in a patient who had no molecular diagnosis for many years. Lastly, the mitochondrial genome was also considered and three likely pathogenic variants were identified. One was a missense in *MT-ATP6* gene that had not been previously reported, that was heteroplasmic (83%, with depth = 5681x) in blood from a patient with retinitis pigmentosa and ataxia. Our findings demonstrate the value of using WGS to investigate the genetic causes of NDD by identifying pathogenic single-nucleotide variants, indels and structural variants in coding and intronic regions of the human genome.

1167W

Recurrent *de novo* missense variation in the GTP/GDP-binding region of RALA is associated with a novel neurodevelopmental disorder.

M.B. Neu^{1,2}, S.M. Hiatt¹, R.C. Ramaker^{1,2}, J.W. Prokop³, M. Hancarova⁴, D. Prchalova⁴, M. Havlovicova⁴, J. Prchal⁵, V. Stranecky⁶, D.K.C. Yim⁷, Z. Powis⁸, B. Keren⁹, C. Nava⁹, C. Mignot^{9,10,11}, M. Rio^{10,12}, A. Revah-Politi¹³, P. Hemati¹³, M. Stong¹³, A.D. Iglesias¹⁴, S. Suchy¹⁵, R. Willaert¹⁵, I.M. Wentzensen¹⁵, P.G. Wheeler¹⁶, L. Brick¹⁷, M. Kozenko¹⁷, A.C.E. Hurst¹⁸, J.W. Wheless^{18,19}, Y. Lacasie^{20,21}, Z. Sedlacek⁴, G.M. Cooper⁴. 1) HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA; 2) Department of Genetics, University of Alabama at Birmingham, Birmingham, AL, USA; 3) Department of Pediatrics and Human Development, Michigan State University, East Lansing, MI, USA; 4) Department of Biology and Medical Genetics, Charles University 2nd Faculty of Medicine and University Hospital Motol, Prague, Czech Republic; 5) Laboratory of NMR Spectroscopy, University of Chemistry and Technology, Prague, Czech Republic; 6) Department of Pediatrics and Adolescent Medicine, Diagnostic and Research Unit for Rare Diseases, Charles University 1st Faculty of Medicine and General University Hospital, Prague, Czech Republic; 7) Kaiser Permanente-Hawaii, Honolulu, HI, USA; 8) Department of Emerging Genetic Medicine, Ambray Genetics, Aliso Viejo, CA, USA; 9) Department of Genetics, La Pitié-Salpêtrière Hospital, Assistance Publique-Hôpitaux de Paris, Paris, France; 10) Centre de Référence Déficiences Intellectuelles de Causes Rares, Paris, France; 11) Groupe de Recherche Clinique UPMC "Déficience Intellectuelle et Autisme", Paris, France; 12) Assistance Publique-Hôpitaux de Paris, service de Génétique, Hôpital Necker-Enfants-Malades, Paris, France; 13) Institute for Genomic Medicine, Columbia University Medical Center, New York, NY, USA; 14) Division of Clinical Genetics, Department of Pediatrics, Columbia University Medical Center, New York, NY, USA; 15) GeneDx, Gaithersburg, MD, USA; 16) Arnold Palmer Hospital, Division of Genetics, Orlando, FL, USA; 17) Department of Genetics, McMaster Children's Hospital, Hamilton, Ontario, Canada; 18) Division of Pediatric Neurology, University of Tennessee Health Science Center, Neuroscience Institute & Le Bonheur Comprehensive Epilepsy Program, Memphis, TN, USA; 19) Le Bonheur Children's Hospital, Memphis, TN, USA; 20) Division of Clinical Genetics, Louisiana State University Health Sciences Center, New Orleans, LA, USA; 21) Department of Genetics, Children's Hospital, New Orleans, LA, USA.

Genetic etiologies of developmental delay and intellectual disability (DD/ID) are still being discovered as a result of large-scale genomic sequencing and data sharing. Here we describe a GeneMatcher-facilitated cohort of eleven individuals with variation in the *RALA* gene, which encodes a RAS-like GTPase. All probands present with DD/ID, including significant speech and motor delays. Variable features in at least half of the cohort include facial dysmorphism and seizures. We found nine independent, *de novo*, variants in ten individuals, two of whom are monozygotic twins; the observed rate of *de novo* variants in affected individuals in large cohorts (8 in 16,086 probands) is significantly higher than that expected given estimated mutation rates ($p=4.93e-11$). These nine variants, including eight missense variants and one in-frame deletion, all affect deeply evolutionarily conserved residues that form the GTP/GDP binding region of *RALA*; in fact, six variants affect only two residues, Val25 and Lys128. This three-dimensional localization strongly suggests that these mutations impact the GTPase activity of *RALA* and thereby alter *RALA*-mediated signaling. Supporting this, both cancer- and DD/ID-associated mutations have been observed in spatially similar locations within *RALA* homologs. Further, although there are 34 *RALA* missense variants in large population databases (gnomAD and BRAVO), none occur at any of the 24 GTP/GDP-binding region residues of *RALA* (206 total amino acids), an observation that is likely non-random ($p=0.017$) and suggests that the binding region is highly intolerant to variation in the general population. We also identified one proband with a nonsense variant of unknown inheritance at position 176 of 206. While this variant is not predicted to lead to nonsense-mediated decay and may lead to production of a *RALA* protein missing residues required for membrane localization, its disease relevance is unclear. Characterization of the phenotypes and variation observed here lead us to suggest *RALA* as a new cause for DD/ID, adding to the group of previously established DD/ID genes in the RAS family.

1168T

36 novel recessive candidate genes for intellectual disability and visual impairment in 269 consanguineous families. S.E. Antonarakis^{1,2,3}, S.A. Paracha⁴, S. Imtiaz⁵, A. Nazir⁶, Y.M. Waryah⁶, P. Makrythanasis^{1,7}, S. Qureshi⁴, S. Saeed⁸, J. Khan⁴, E. Falconnet⁴, M. Guipponi⁸, C. Borel⁴, M.A. Ansari⁸, E. Frengen⁹, E. Ranza^{1,2}, F.A. Santoni^{1,9}, I. Shah⁴, K. Gul¹⁰, J. Ahmed⁴, A.M. Waryah⁶, M.T. Sarwar⁴, M. Ansari⁴. 1) Department of Genetic Medicine and Development, University of Geneva, Geneva, Switzerland; 2) Service of Genetic Medicine, University Hospitals of Geneva, Geneva, Switzerland; 3) iGE3 Institute of Genetics and Genomics of Geneva, Geneva, Switzerland; 4) Institute of Basic Medical Sciences, Khyber Medical University, Peshawar, Pakistan; 5) Department of Genetics, University of Karachi, Karachi, Pakistan; 6) Molecular Biology and Genetics Department, Medical Research Center, Liaquat University of Medical and Health Sciences, Jamshoro, Pakistan; 7) Biomedical Research Foundation of the Academy of Athens, Athens, Greece; 8) Department of Medical Genetics, Oslo University Hospital and University of Oslo, Oslo, Norway; 9) Department of Endocrinology Diabetes and Metabolism, University Hospital of Lausanne, Switzerland; 10) Department of Biosciences, Faculty of Life Sciences, Muhammad Ali Jinnah University, Karachi, Pakistan.

Consanguinity, practiced in a substantial fraction of human populations, reveals numerous rare recessive phenotypes because of the extensive regions of homozygosity by descent. The average total genomic homozygosity is 253Mb in the offspring of consanguineous parents; this is 10fold higher than that of 25Mb in outbred individuals. Similarly, the rare homozygous variants (<2% frequency) in the coding regions are on average 57 in the offspring of first-cousins compared to 18 in outbred individuals. In Pakistan, the frequency of consanguineous marriages approaches 70%. We have initiated a Swiss-Pakistani project to identify novel recessive gene candidates for two phenotypes: Intellectual disability (ID) and visual impairment (VI). We have collected samples from 145 ID and 247 VI families of first cousin marriages with at least 2 affecteds. Exome sequence of one affected and genotyping of the whole family (parents, all affected and unaffected siblings) has been completed in 122 ID and 147 VI families to date. The likely causative gene/variant in known genes was found in 67% of the VI and 34% in the ID families. Thus, there are more unknown recessive genes for ID. We have identified 21 novel candidate genes in VI and 15 in ID (to be presented in the conference). International matchmaking identified additional families in 20% of the candidate genes. Careful evaluation of the phenotypes is mandatory to assess the possibility of 2 or more causative genes in certain families, and to minimize false negative results. International databases from consanguineous individuals are needed to facilitate the assignment of pathogenicity to homozygous variants.

1169F

Phenotypic expansion of *POGZ*-related disorders (White-Sutton Syndrome). N. Assia Batzir^{1,2}, V.R. Sutton^{1,2}. 1) Baylor College of Medicine, Houston, TX; 2) Texas Children's Hospital, Houston, TX.

White-Sutton Syndrome (WSS, OMIM #616364) is a recently-identified genetic disorder resulting from *de novo* heterozygous pathogenic variants in the gene *POGZ*. Pathogenicity of *de novo* variants in this gene was suggested from a large cohort of individuals with autism spectrum disorders, schizophrenia and intellectual disability, and it became recognized as a syndrome following the description of 5 individuals with developmental delays, behavioral abnormalities and shared physical features including microcephaly, strabismus, short stature, skeletal abnormalities and typical facial characteristics. This illustrates the "genotype-first" concept where grouping of patients based on similar genetic etiologies leads to the characterization of genetics syndromes. Thus far, over 30 patients have been reported worldwide, however phenotypic characterization and data regarding the natural history and associated features are still incomplete. We undertook a study to better delineate the clinical phenotype of WSS. Information was obtained from medical records of 17 individuals with WSS identified through self-reporting and physician referrals. Gastrointestinal maladies and sleep abnormalities emerged as common features in patients with WSS, and so we further gathered information through validated patient/parent report tools. Results reveal a broad spectrum of intellectual disability/developmental delay, and suggest a wider phenotypic range than previously described. Moreover, a variety of gastrointestinal issues were identified ranging from congenital diaphragmatic hernia to poor feeding in infancy and early-onset obesity. Sleep apnea and sleep disordered breathing were also more common than in the general population. This study expands the phenotypic spectrum of this new syndromic form of intellectual disability and autism.

1170W

De novo mutations in the RNA helicase *DDX6* cause intellectual disability and dysmorphic features by disrupting p-body assembly and RNA regulation. C. Balak¹, M. Benard², E. Schaefer³, K. Ramsey⁴, M. Ernault-Lange⁵, F. Mattioli⁶, N. Belnap⁷, V. Geoffroy⁸, M. Courel⁹, J. Muller¹⁰, M. Naymik¹¹, K. Bachman⁶, M. Cho⁷, R. Richholt¹, A. Le Behec⁸, W.M. Jepsen¹, M. De Both¹, S. Rangasamy¹, J-F. Deleuze⁶, A. Boland⁹, S. Szelinger¹⁰, H. Dollfus³, I. Piras¹, J. Chelley⁴, D.W. Craig¹¹, J-L. Mandel¹, V. Narayanan¹, M. Huettelmann¹, D. Weil¹, A. Piton¹. 1) Neurogenomics / Ctr. for Rare Childhood Disorders, Translational Genomics Res. Inst., Phoenix, AZ; 2) UPMC Univ. Paris 06, Inst. de Biologie Paris-Seine (IBPS), CNRS UMR 7622, Paris, France; 3) Service de Génétique Médicale, Hôpitaux Universitaires de Strasbourg, Inst. de Génétique Médicale d'Alsace, Strasbourg, France; 4) Inst. de Génétique et de Biologie Moléculaire et Cellulaire, Illkirch, France; 5) Lab. de Génétique médicale, UMR_S INSERM U1112, IGMA, Faculté de Médecine FMTS, Univ. de Strasbourg, Strasbourg, France; 6) Genetics, Geisinger Med. Ctr., Dansville, PA; 7) GeneDx, Gaithersburg, MD; 8) Luxembourg Ctr. for Systems Biomedicine, Luxembourg, Luxembourg; 9) Ctr. Natl. de Recherche en Génomique Humaine, Inst. de Biologie François Jacob, CEA, Evry, France; 10) Pathology and Lab. Med., UCLA Ctr. for the Hlth. Sci., Los Angeles, CA; 11) Translational Genomics, Keck Sch. of Med. of USC, Los Angeles, CA.

The RNA helicase *DDX6* is a member of the DEAD-box protein family and plays vital roles in RNA metabolism processes such as translational repression, decapping and RNA decay. Previous *in vitro* and lower *in vivo* models have implicated altered *DDX6* in cellular dysfunction but clinical consequences and pathogenesis in humans have not yet been described. Here we report the identification of multiple unrelated patients carrying private, *de novo* missense mutations in *DDX6*, all presenting with non-epileptic intellectual disability, developmental delay, microcephaly and overlapping dysmorphic features of unknown etiology. All missense variants reside in exon 11 which encodes the highly-conserved C-terminal RecA domain of *DDX6* involved in RNA binding, helicase activity and protein partner binding. Comprehensive *in vitro* functional studies performed on *DDX6* p.Arg373Gln and p.Cys390Arg variants demonstrated that both mutations lead to significant defects in P-body assembly and interactions with protein partners involved in translation repression (and mRNA degradation). Study of patient-derived fibroblasts carrying the p.Cys390Arg mutation demonstrated a significant reduction in total P-body number when compared to an unrelated age-matched control individual. Additional transcriptomic studies in available patient cells revealed among the genes up-regulated an enrichment of mRNA excluded from P-bodies, targets of *DDX6*, and in mRNA encoding protein involved in regulation of translation. Additional patients with *de novo* mutations affecting the same or adjacent amino acid to p.Cys390 were also identified. Collectively, our clinical and molecular data define a novel role for *DDX6* mutations in human neurodevelopmental dysfunction and confirm the implication of this gene in an autosomal dominant form of intellectual disability with developmental delay and dysmorphic features.

1171T

De novo truncating mutations *HNRNPR* gene cause a new syndrome with intellectual disability and microcephaly. A. Bruel¹, A. MacInnes², S. Mouton^{1,3}, F. Tran Mau-Them^{1,4}, F.A.M. Duijkers⁵, M. Cho⁶, R.L. Ladda⁷, Z. Powis⁸, K.G. Monaghan⁹, Y. Duffourd¹, C. Philippe^{1,3}, L. Faivre^{1,3}, C. Thauvin-Robinet¹⁰. 1) UMR1231 GAD, Inserm - Université Bourgogne-Franche Comté, Dijon, France; 2) Laboratory Genetic Metabolic Diseases, Academic Medical Center, Amsterdam, The Netherlands; 3) Centre de Référence Anomalies du Développement et syndromes malformatifs, centre de génétique, FHU-TRANSLAD, CHU Dijon Bourgogne, Dijon, France; 4) Unité Fonctionnelle d'Innovation diagnostique en maladies rares, FHU-TRANSLAD, CHU Dijon Bourgogne, Dijon, France; 5) Department of Clinical Genetics, Academic Medical Center, Amsterdam; 6) GeneDx, Gaithersburg, Maryland, USA; 7) Department of Pediatrics, Penn State Hershey Children's Hospital, Hershey, USA; 8) Ambry Genetics, Department of Clinical Genomics, Aliso Viejo, California, USA; 9) Centre de Référence Déficience Intellectuelle, centre de génétique, FHU-TRANSLAD, CHU Dijon Bourgogne, Dijon, France.

Approximately 1% of individuals present with an intellectual disability (ID) that can be caused by genetic and/or exogenous factors. Genetic causes are multiple with more than 500 genes associated to isolated ID and more than 2,000 genes involved in syndromic ID reported in OMIM database. The clinical heterogeneity of ID is reflected with the wide genetic heterogeneity. During these last years, NGS has proven to be useful for the identification of the molecular bases of heterogeneous conditions such as ID. However, a large number of results remain non-conclusive because the diagnosis of ultra-rare conditions limits the genotype-phenotype correlations. Broad international data-sharing catalyzes the identification of additional cases and allows the delineation of new syndromes. In this study, we report a girl with ID associated with microcephaly, facial dysmorphism, brachydactyly, short feet and a thin corpus callosum. A trio-based WES strategy identified a *de novo* frameshift variant in the *HNRNPR* gene. To assess the relevance of this finding we asked international collaborators for patients with *HNRNPR* truncating variants. Three additional patients were recruited with *de novo* truncating variant by GeneMatcher. All the 3 patients present with a similar syndromic phenotype including ID, microcephaly, generalized hypotonia, brachydactyly, feeding difficulties and corpus callosum anomalies. The *HNRNPR* gene encodes a nuclear ribonucleo-protein of the large superfamily of the RNP consensus RNA-binding proteins, involved in the nucleic acid metabolism which functions in pre-mRNA processing, in transcriptional and translational regulation. *HNRNPR* is a member of the spliceosome C complex and seems more specifically implicated in the stress granule metabolism. Stress granules are mRNA-protein assemblies formed from nontranslating mRNAs, and are essential in the stress response and may contribute to some degenerative diseases and ID. Functional studies in patient's fibroblasts revealed stress granule defects including mislocalization and associated with a cell death. An ongoing RNAseq in the stress granule transcriptome will enable to identify targets of *HNRNPR* to understand the function of *HNRNPR*. In conclusion, we report *HNRNPR* as a new gene associated with syndromic ID and microcephaly, involved in the stress granule metabolism.

1172F

The GeJo-ARID project: Identification of new genes in individuals with intellectual disability. R. Buchert¹, O. Nafie², L. Laugwitz^{1,3}, S. Al Hait⁴, M. Sturm¹, J. Sommerfeld¹, A. Baraghiti⁵, P.A. Koch¹, J. Hanselmann¹, B. Kootz¹, J. Gohlke¹, P. Bauer¹, M. Hamdallah⁶, W. Al Ameri⁷, R. Abou Jamra⁸, A.J. Alfrook⁹, A. Rieß¹, T.B. Haack¹, O. Rieß¹, T. Froukh⁵. 1) Institute of Medical Genetics and Applied Genomics, University of Tübingen, Tübingen, Germany; 2) Mutah University, Jordan; 3) Department of Neuropaediatrics, Developmental Neurology and Social Pediatrics, University of Tübingen, Tübingen, Germany; 4) Private Clinic, Jordan; 5) Dept. of Biotechnology and Genetic Engineering, Philadelphia University, Amman, Jordan; 6) Farah Hospital, Amman, Jordan; 7) Ibn Alhytham Hospital, Amman, Jordan; 8) Institute of Human Genetics, University Medical Center Leipzig, Leipzig, Germany; 9) private practice, Amman, Jordan.

In a bilateral project between Germany and Jordan, we examined 103 Jordanian families with intellectual disability (ID). 80% of the families were consanguineous and 50% had more than one affected child. We conducted whole exome sequencing (WES) for one affected individual per family and prioritized the data according to putative causative variants. In total, we solved 32% of the cases finding mutations in known disease genes, and identified good candidate variants in another 10% of the patients. 63% of the causative variants were autosomal recessive, 30% were *de novo* and 7% were X-chromosomal. In case of identified splice variants we used cDNA to verify aberrant splicing. Using the WES coverage data we also identified 3 large deletions with sizes of 600 kb to 2 Mb which occurred *de novo*, as confirmed with microarray analysis. While data analysis is still in progress we could already identify a new candidate gene, *DENND5A*, and prove the causative effect of bi-allelic *DENND5A* mutations in epileptic encephalopathy. Further, we identified new pathogenic variants in recently identified candidate genes for ID. In one family we identified a homozygous truncating variant in *PUS3*, a gene which was previously implied in another family with similar symptoms. In another family we identified a *de novo* truncating mutation in *YWHAQ*, which was recently linked to epileptic encephalopathy. Currently, we are investigating various other candidate genes which were identified in families in this project. This project demonstrates a beneficial cooperation between various specialties such as pediatricians and human geneticists from Jordan and Germany and helps to improve genetic diagnostics and counselling of children with neurodevelopmental disorders in Jordan. This project is funded through a grant by the DAAD.

1173W

***De novo* variants in the F-Box protein *FBXO11* in 20 individuals with a variable neurodevelopmental disorder.** A. Gregori¹, L.G. Sadleir², R. Asadolah³, L. Bomme Ousager⁴, A.L. Bruehl⁵, R. Buchert⁶, E. Calpena⁷, B. Cogné^{8,9}, F. Elmslie¹⁰, T.B. Haack⁶, A. Henderson¹¹, D. Hunt¹², B. Isidor⁶, P. Joset⁶, A.M.A. Lachmeijer¹³, M. Lees¹⁴, S.A. Lynch¹⁵, H.C. Mefford¹⁶, S. Moutton⁶, A. Novelli¹⁷, A. Rauch⁶, M. Rosello¹⁸, K. Saida¹⁹, A.B. Santani^{20,21}, A. Sarkar²², M. Shinawi²³, D.D.D. study²⁴, A. Reis¹, H. Sticht²⁵, C. Zweier¹. 1) Institute of Human Genetics, Erlangen, Germany; 2) Department of Paediatrics and Child Health, University of Otago, Wellington 6242, New Zealand; 3) Institute of Medical Genetics, University of Zurich, 8952 Schlieren-Zurich, Switzerland and Radix – “Rare Disease Initiative Zurich, Clinical Research Priority Program for Rare Diseases University of Zurich”, 8032 Zurich, Switzerland; 4) Department of Clinical Genetics, Odense University Hospital, 5000 Odense, Denmark; 5) INSERM U1231, LNC UMR1231 GAD, Burgundy University, 21079 Dijon, France; 6) Institute of Medical Genetics and Applied Genomics, University of Tübingen, 72076 Tübingen, Germany; 7) Clinical Genetics Group, MRC Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DS, UK; 8) Department of Medical Genetics, CHU Nantes, 44093 Nantes, France; 9) l'Institut du Thorax, INSERM, CNRS, UNIV Nantes, 44007 Nantes, France; 10) South West Thames Regional Genetics Service, St. George's, University of London, London SW17 0RE, UK; 11) Northern Genetics Service, Newcastle upon Tyne Hospitals, NE1 3BZ, UK; 12) Wessex Clinical Genetics Service, Princess Anne Hospital, Southampton, SO16 5YA, UK; 13) Department of Genetics, University Medical Center Utrecht, P.O. box 85090, 3508 AB Utrecht, The Netherlands; 14) North East Thames Regional Genetics Service, Great Ormond Street Hospital for Children NHS Foundation Trust, Great Ormond Street Hospital, London WC1N 3JH, UK; 15) Dept of Clinical Genetics, Temple Street Children's Hospital Dublin 1, D12 V004 Dublin, Ireland; 16) Division of Genetic Medicine, Department of Pediatrics, University of Washington, Seattle, WA 98195, USA; 17) Genetics and Rare Diseases Research Division, Ospedale Pediatrico Bambino Gesù, IRCCS, Rome 00146, Italy; 18) Unidad de Genética, Hospital Universitario y Politécnico La Fe, Avda Fernando Abril Martorell 106, 46026 Valencia, Spain; 19) Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama 236-0004, Japan; 20) Division of Genomic Diagnostics, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; 21) Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA; 22) Nottingham Regional Genetics Service, City Hospital Campus, Nottingham University Hospitals NHS Trust, The Gables, Hucknall Road, Nottingham NG5 1PB, UK; 23) Division of Genetics and Genomic Medicine, Department of Pediatrics, Washington University School of Medicine, St. Louis, MO 63110, USA; 24) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK; 25) Institute of Biochemistry, Emil-Fischer Center, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91054 Erlangen, Germany.

Next-Generation sequencing combined with international data sharing has enormously facilitated identification of new disease associated genes and mutations. This is particularly true for genetically extremely heterogeneous entities such as neurodevelopmental disorders (NDDs). Through exome sequencing and world-wide collaborations we identified and assembled 20 individuals with *de novo* variants in *FBXO11*. They present with mild to severe developmental delay associated with a range of features including short (4/20) or tall (2/20) stature, obesity (5/20), microcephaly (4/19) or macrocephaly (2/19), behavioral problems (17/20), seizures (5/20), cleft lip or palate or bifid uvula (3/20) and minor skeletal anomalies. *FBXO11* encodes a member of the F-Box protein family, constituting a subunit of an E3-ubiquitin ligase complex. This complex is involved in ubiquitination and proteasomal degradation and thus in controlling critical biological processes by regulating protein turnover. The identified *de novo* aberrations comprise two large deletions, ten likely gene disrupting and eight missense variants distributed throughout *FBXO11*. Structural modeling for missense variants located in the CASH or the Zinc-finger UBR domains suggests destabilization of the protein. This in combination with the observed spectrum and localization of identified variants and the lack of apparent genotype-phenotype correlations is compatible with loss-of-function or haploinsufficiency as an underlying mechanism. We implicate *de novo* missense and likely gene disrupting variants in *FBXO11* in a neurodevelopmental disorder with variable intellectual disability and various other features.

1174T

Systematic scoring of candidate genes for neurodevelopmental disorders has a higher yield than large meta-analyses. R. Jamra¹, S. Martin¹, I. Krey¹, D. Le Duc¹, T. Bartolomaeus¹, T. Polster^{1,4}, D. Huhle¹, W. Kiess², A. Merckenschlager³, M. Bernhard³, R. Pfäffle³, F. Hornemann³, D. Wieczorek⁴, S. Hoffjan⁵, Y. Hellenbroich⁶, A. Küchler⁷, K. Platzer¹, S. Syrbe⁸, J.U. Schlump⁹, A. Rolfs¹⁰, S. Biskup¹¹, S. Leye¹², R. Ewald¹², I. Schanze¹³, M. Zenker³, P. Muschke¹³, J. Hentschel¹, B. Liesfeld¹², J. Lemke¹, B. Büttner¹. 1) University of Leipzig, Leipzig, Germany; 2) Praxis für Humangenetik Leipzig, Leipzig, Germany; 3) Hospital for Children and Adolescents, University Medical Center Leipzig, Leipzig, Germany; 4) Institute of Human Genetics and Anthropology, Heinrich Heine University Düsseldorf, Düsseldorf, Germany; 5) Department of Human Genetics, Ruhr-University, Bochum, Germany; 6) Institute of Human Genetics, University Lübeck, Lübeck, Germany; 7) Institute of Human Genetics, University Hospital Essen, Essen, Germany; 8) Division for Neuropaediatrics and Metabolic Medicine, Center for Paediatric and Adolescent Medicine, Heidelberg University Hospital, Heidelberg, Germany; 9) Evangelical Hospital Oberhausen, Division for Children and Adolescents, Oberhausen, Germany; 10) Centogene AG, Rostock, Germany; 11) CeGaT GmbH, Center for Genomics and Transcriptomics, Tübingen, Germany; 12) Limbus Medical Technologies GmbH, Rostock, Germany; 13) Institute of Human Genetics, University Hospital Magdeburg, Magdeburg, Germany; 14) Bethel Epilepsy Center, Hospital Mara GmbH Bielefeld, Germany.

Introduction Large meta analyses of trio exome sequencing in neurodevelopmental disorders (NDD) identify novel disorder genes. However, the yield is relatively low. E.g. the DDD study in the UK described in a cohort of over 7000 trio exome analysis only 94 genes, most of them were known in advance. **Materials and Methods** We performed trio exome sequencing of 199 undiagnosed NDD cases. Sequencing was performed by Centogene (Rostock, Germany) or CeGaT (Tübingen, Germany) using well-established enrichment kits and sequencing platforms. Bioinformatic analyses were performed by Limbus (Rostock, Germany). If exome sequencing did not reveal a clear causative variant in an established NDD gene, we searched for candidate genes on scientific basis. We developed a scoring system using 13 parameters, based on inheritance, gene and variant attributes, and published literature. **Results** A reliable clinical genetic diagnosis was made in 64 (32%) trios. In the remaining 135 (68%) trios, scientific re-evaluation revealed 181 potentially causative variants in candidate genes. We applied our scoring system to prioritize the relevance of these variants. The scores varied between 2 and 11.9. The top 10% of the scores were made by the genes *ASIC1*, *TANC2*, *FBXL19*, *KMT2E*, *GLS* (twice), *ACTL6B*, *GRIN3B*, *CUX1*, *PUM1*, *UNC13A*, *GRIA4*, *SPEN*, *PUM2*, *TOP1*, *NPTX1*, *ETV5*, and *WDFY3*. Within few months, international cooperation has been set up and clinical and functional analyses were performed, and the majority of these candidates is validated and is already published or under review, or a manuscript is under preparation. **Discussion and Conclusion** Our results illustrate the enormous scientific value of the re-evaluation of the substantial amount of negative exome sequencing samples. In comparison to large studies that have meta-analyzed thousands of trio exomes, we achieve a higher yield (15 novel genes out of 135 families) through individual evaluation of each case, systematically scoring, and an approachable and cooperative attitude. We want to present to the conferees our results and encourage them re-evaluate and share in order to fully decipher the genetics of NDD within few years.

1175F

Two novel mutations in DYRK1A gene cause microcephaly and learning difficulties. WT: Keng¹, Rachel. Challis², Clare. Logan², Jennie. Murray², Andrew. Jackson², GS. Chng¹, LH. Ngu¹, HY. Leong¹. 1) Genetic Dept, Hospital Kuala Lumpur, Kuala Lumpur, Malaysia; 2) MRC Human Genetics Unit, MRC Institute of Genetics & Molecular Medicine, Western General Hospital, Edinburgh.

DYRK1A codes for a member of the dual-specificity tyrosine phosphorylation-regulated kinase family. Triplication of this highly conserved gene located on chromosome 21q22.1 is thought to play a role in cognitive deficits associated with Down syndrome. Haploinsufficiency of the DYRK1A protein can lead to moderate to severe learning difficulties, microcephaly, epilepsy as well as characteristic facies, scoliosis and short stature. We describe 2 Malaysian patients who have microcephaly and learning difficulties caused by novel de novo heterozygous mutation [c.469delG, p.Asp157IlefsTer8; c.843_864dup-TAGTATCATTCACTGTGATCTA, p.Lys289Ter] in DYRK1A gene. Both have significant speech deficit. One of them also had additional abnormalities including dystrophic disc with myelinated nerve fibre layer in her left eye, cleft dimple complex on the right cerebrum in brain MRI imaging.

1176W

Expanding the spectrum of BAF-related disorders – *de novo* variants in *SMARCC2* cause a syndrome with intellectual disability, developmental delay and growth retardation. K. Machol¹, J. Rousseau², S. Ehresmann², R.C. Spillmann³, J.A. Sullivan³, V. Shashi³, Y.H. Jiang⁴, S. Petrovski⁵, E. Fiala⁶, R. Pfundt⁷, T. Kleefstra⁷, M.T. Cho⁸, M. Rosello Piera⁹, C. Orellana⁹, T. Roscioli¹⁰, M. Buckley¹¹, A. Turner¹², W. Jones¹³, P.M. van Hasselt¹⁴, A.S. Brooks¹⁵, K. Lachlan¹⁶, J. Sebastian¹⁷, D. Sonal¹⁸, T. Julien¹⁹, N. Stong²⁰, S. Petrovski²¹, I.D. Krantz²², J.A. Rosenfeld¹, B. Lee¹, P. Campeau², *Undiagnosed Diseases Network*. 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Department of Pediatrics, CHU Sainte-Justine Research Center and University of Montreal, Montreal, QC, H3T 1C5, Canada; 3) Department of Pediatrics, Duke University School of Medicine, Durham, North Carolina 27710, USA; 4) Department of Pediatrics and Neurobiology, Program in Genetics and Genomics, Duke University School of Medicine, Durham, North Carolina 27710, USA; 5) Department of Medicine, The University of Melbourne, Melbourne, Australia; 6) Department of Pediatrics, Washington University School of Medicine, St. Louis, MO, USA; 7) Human Genetics Department, Radboud University Medical Center, HB, Nijmegen, The Netherlands; 8) GeneDx, Gaithersburg, Maryland, USA; 9) Unidad de Genética, Hospital Universitari i Politècnic La Fe, Valencia, Spain; 10) Centre for Clinical Genetics, Sydney Children's Hospital, Neuroscience Research Australia, University of New South Wales, Sydney, Australia; 11) New South Wales Health Pathology, Randwick, Australia; 12) Centre for Clinical Genetics, Sydney Children's Hospital, Sydney, Australia; 13) Locum Consultant in Clinical Genetics, Great Ormond Street Hospital for Children NHS Foundation Trust, London WC1N 3JH, UK; 14) Department of Metabolic Diseases, Wilhelmina Children's Hospital, University Medical Center Utrecht, The Netherlands; 15) Department of Clinical Genetics, Erasmus Medical Center, Rotterdam, The Netherlands; 16) Wessex Clinical Genetics Service, Princess Anne Hospital, Southampton, UK; 17) Department of Medical Genetics, Children's Hospital of Pittsburgh of UPMC, Pittsburgh, PA; 18) Department of Neurogenetics, Kennedy Krieger Institute, Baltimore, MD, USA; 19) Centre de Génétique, Centre de Référence Anomalies du Développement et Syndromes Malformatifs de l'Interrégion Est et FHU TRANSLAD, CHU Dijon, F-21079 Dijon, France; 20) Institute for Genomic Medicine, Columbia University, New York, NY 10032, USA; 21) AstraZeneca Centre for Genomics Research, Precision Medicine and Genomics, IMED Biotech Unit, AstraZeneca, Cambridge CB2 0AA, UK; 22) Department of Pediatrics and Division of Human Genetics, The Children's Hospital of Philadelphia and the Perelman School of Medicine at The University of Pennsylvania, Philadelphia PA 19104, USA.

SMARCC2 (BAF170, MIM:601734) is one of the invariable core subunits of the ATP-dependent chromatin remodeling BAF (BRG1-associated factor) complex and plays a crucial role in embryogenesis and corticogenesis. Pathogenic variants in genes encoding other components of the BAF complex have been associated with intellectual disability syndromes. Despite its significant biological role variants in *SMARCC2* have not been associated with human disease previously. Using whole exome sequencing and a web-based gene-matching program, we identified 15 individuals with variable degrees of neurodevelopmental delay and growth retardation harboring one of 13 heterozygous variants in *SMARCC2*, most of them novel and proven *de novo*. The clinical presentation overlaps with other BAF-related syndromes (BAFopathies) and includes prominent speech impairment, hypotonia, feeding difficulties, behavioral abnormalities, and dysmorphic features such as hypertrichosis, thick eyebrows, thin upper lip vermillion, and upturned nose. Nine out of the fifteen individuals harbor variants in the highly conserved *SMARCC2* DNA-interacting domains (SANT and SWIRM) and present with a more severe phenotype. Two of these individuals present cardiac abnormalities. Transcriptomic analysis of fibroblasts from affected individuals highlights a group of differentially expressed genes with possible roles in regulation of neuronal development and function, namely *H19*, *SCRG1*, *RELN* and *CACNB4*. Our findings suggest a novel *SMARCC2*-related syndrome which overlaps with the spectrum of BAF-related neurodevelopmental disorders.

1177T

A novel X-linked dominant mutation of *MED12* causes Ohdo syndrome in a female patient. H. Murakami¹, Y. Kimura¹, Y. Enomoto², Y. Tsurusaki², Y. Kuroda², Y. Sugio³, K. Kurosawa¹. 1) Division of Medical Genetics, Kanagawa Children's Medical Center, Yokohama, Japan; 2) Clinical Research Institute, Kanagawa Children's Medical Center, Yokohama, Japan; 3) Department of Pediatrics, Tsudumigaura Medical Center for Children with Disabilities, Yamaguchi, Japan.

MED12 mutations have been shown to cause X-linked form of Ohdo syndrome, which comprises a heterogeneous group of disorders characterized by intellectual disability (ID) and blepharophimosis. We identified the novel *de novo* missense mutation in *MED12* [MIM 300188](NM_005120:exon4:cG-514C:p.E172Q) in a female patient with severe ID and congenital anomalies. Patient is a five-year-old female who was first seen by medical genetics at about twelve months due to severe developmental delay (DD). She presented with severe ID/DD and congenital anomalies including ocular hypertelorism, strabismus, epicanthus, low nasal root, wide and thick alae nasi, pointed chin, hyperextension of limb joints, hypoplasia of nails, scoliosis and vitiligo. She was born at full term to non-consanguineous parents after an uncomplicated pregnancy and birth parameters were normal. Her developmental milestones were severely delayed: rolling over at 12 months, sitting at 2 years, walking at 3 years and 6 months. She could not speak meaningful words at 5 years of age. Standard karyotyping was normal. A brain MRI did not reveal abnormalities. We performed exome sequence and identified a novel heterozygous mutation in *MED12* gene. *MED12* encodes a subunit of the Mediator complex, which is required for transcription by RNA polymerase II. Sanger sequencing confirmed that this mutation was *de novo*. This mutation affects evolutionary highly conserved amino acid and is predicted damaging to protein function. Analysis of X-chromosome-inactivation status showed skewed pattern. There has been a very few reports of X-linked dominant mutation in *MED12*. This is a first report about *MED12* mutation that causes severe ID and congenital abnormalities in a female patient.

1178F

Novel biallelic mutations in *SZT2* cause mild intellectual disability and epilepsy: Expanding the phenotypic spectrum. Y. Nakamura¹, Y. Togawa², Y. Okuno³, H. Muramatsu⁴, K. Nakabayashi⁵, Y. Kuroki⁶, D. Ieda¹, I. Hori¹, Y. Negishi¹, T. Togawa¹, A. Hattori¹, S. Kojima¹, S. Saitoh¹. 1) Department of Pediatrics and Neonatology, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan; 2) Department of Pediatrics, Toyohashi Municipal Hospital, Aichi, Japan; 3) Center for Advanced Medicine and Clinical Research, Nagoya University Hospital, Nagoya, Japan; 4) Department of Pediatrics, Nagoya University Graduate School of Medicine, Nagoya, Japan; 5) Department of Maternal-Fetal Biology, National Research Institute for Child Health and Development, Tokyo, Japan; 6) Department of Genome Medicine, National Research Institute for Child Health and Development, Tokyo, Japan.

Background: The *SZT2* gene is transcribed in many tissues, with predominant expression in brain. Recent reports have revealed that dysfunctional *Szt2* results in hyperactivation of mTORC1 signaling under nutrient-deprived conditions. The first clinical report in 2013 revealed biallelic mutations in *SZT2* caused early-onset epileptic encephalopathy and dysmorphic corpus callosum, followed by the second report of patients with non-syndromic intellectual disability without seizures. Since then, 6 individuals with *SZT2* mutations, all of whom showed severe early-onset epileptic encephalopathy, have been reported. We herein report an additional patient with biallelic mutations in *SZT2*, which were confirmed by genetic and biochemical analyses. **Case report:** The proband is a 4-year-old girl. Neonatal period was uneventful, but she showed developmental delay with holding up her head at 7 months, sitting at 2 years, and walking at 3 years of age. She was not able to speak meaningful words. Seizures began at 3 years, which were readily controlled by valproate. An EEG showed epileptic discharges which rarely occurred. Brain MRI revealed a short and thick corpus callosum. At 4 years of age, she showed macrocephaly, high forehead, hypertelorism, and pectus carinatum. Whole-exome sequencing detected compound heterozygous *SZT2* mutations, namely c.8596dup (p.Tyr2866Leufs*42) and c.2930-17_2930-3delinsCTC-GTG, all of which transcripts were experimentally confirmed to result in loss of function. Immunoblotting using the lysates of patient's peripheral leukocytes cultured under amino acid-deprived condition showed increased phosphorylation of S6K protein, indicating abnormal hyperactivation of mTORC1 signaling. **Conclusion:** The severity of this case appears to be intermediate between the previously reported severe and mild phenotypes, which suggests a broad phenotypic spectrum arising from *SZT2* mutations, forming a continuum from epileptic encephalopathy and severe developmental delay to mild intellectual disability without epilepsy. Furthermore, biochemical analyses using peripheral leukocytes might be useful for determining the pathogenicity of *SZT2* variants.

1179W

The first report on a patient with OGT catalytic core mutation causing O-GlcNAc transferase defect. K. Ounap^{1,2}, S. Pajusalu^{1,2}, R. Zordania¹, M.H. Wojcik^{3,4}, V. Pravata⁵, D. Van Aalten⁵. 1) Department of Clinical Genetics, United Laboratories, Tartu University Hospital, 2 L. Puusepa street, Tartu 50406, Estonia; 2) Department of Clinical Genetics, Institute of Clinical Medicine, University of Tartu, 2 L. Puusepa street, Tartu 50406, Estonia; 3) Divisions of Newborn Medicine and Genetics and Genomics, Department of Medicine, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA; 4) Broad Institute of MIT and Harvard, Cambridge, MA, USA; 5) School of Life Sciences, University of Dundee, Dow Street, Dundee DD1 5EH, Scotland, United Kingdom.

Five missense mutations in O-GlcNAc transferase (OGT) have recently been reported to cause X-linked intellectual disability (XLID). All of these previously-described cases, missense mutations are located in the tetratricopeptide (TPR) repeats of OGT. Here we report for the first time a missense variant identified in the OG catalytic core. **Case report:** At birth, the patient presented with remarkable facial asymmetry and almost missing toenails were described. Clinical problems started at 4 months of age with very serious stomach pain and gases, and frequent otitis. Development was subsequently delayed between 6 months and 24 months due to recurrent infections and feeding problems. He started to stand at 24 months and to walk at 28 months. At 7 years of age, his height was 127 cm (+1 SD), weight 23.4 kg (0 SD) and OFC 52 cm (-0.5 SD). He presently has moderate to severe intellectual disability. In addition, he has coarse facial features, strabismus, increased drooling, hyperactivity, no speech, muscular hypotonia and mild T2-3 syndactyly. Brain MRI at 5 years of age showed brain atrophy and a mega cisterna magna. X ray investigation showed cone shape epiphyses of T2-T5. Biochemical analyses showed increased thyroglobulin and FSH, and low percentage of activated T lymphocytes (2.0, normal 2.3-7.0 %). **Methods and results:** Trio whole genome sequencing and data processing were performed by Genomics Platform at the Broad Institute of Harvard and MIT (Broad Institute, Cambridge, MA, USA), which identified a *de novo* variant in *OGT* gene's protein domain: C-terminal O-GlcNAc transferase: ChrX(GRCh38): g.71561865A>T; NM_181672.2: c.1942A>T p.(Asn648Tyr). The variant appeared as *de novo* in protein domain: O-GlcNAc transferase, C-terminal. The mutated p.Asn648 amino acid is highly conserved up to *C. elegans*, and multiple *in-silico* pathogenicity predicting algorithms indicated a damaging effect. Unlike previous OGT mutations described, this mutation is located in the catalytic domain. We have determined the crystal structure of the enzyme to reveal this mutation is proximal to the active site. However, *in vitro* tests of the mutated form of the enzyme has not revealed changes in O-GlcNAc transferase activity. We are currently investigating the effects of this mutation on differentiation of stem cells into neurons and I will report the latest results of this.

1180T

Molecular analysis of *CC2D1A* and *PRSS12* genes in affecteds with cognitive dysfunction. I.G. Rasool, M.Y Zahoor, H.Q Abbas. Institute of Biochemistry and Biotechnology, University of Veterinary and Animal Sciences, Lahore, Punjab, Pakistan.

Introduction: Cognitive disability is a group of genetically heterogeneous abnormality that leads to variable degrees of intellectual disability. This disease affects 1-3% of overall population in globe and majority of monogenic variants are not well characterized. It has been shown that such inherited disorders can be caused by mutations in over a large number of different genes and there is evidence for the presence of as yet unknown genes in a significant proportion of patients. **Materials & Methods:** The study was carried for mutational analysis of genes reported for involvement in cognitive dysfunction. We have screened *CC2D1A* and *PRSS12* genes in ten patients of cognitive dysfunction after enrolling with informed written consents. DNA sequencing of selected exons of the genes were carried out by Sanger method through ABI 3130 sequencer. Bioinformatics analysis performed for polymorphism analysis and protein prediction. **Results:** We mapped with one novel mutation in three patients. This is a deletion of A at 331N position (Chr.19: 14034167/14034169) of *CC2D1A* and it leads to frame-shift and production of truncated protein and thus causing moderate level of cognitive dysfunction. This is the third report in the globe for a novel mutation in *CC2D1A* gene in three unrelated affecteds with cognitive dysfunction. **Conclusion:** The cognitive dysfunction is a lifetime burden in the affecteds of this group and their families. This condition advocates the establishment of expanded newborn screening program for this disorder in indigenous and neighboring populations.

1181F

Three individuals with neurodevelopmental disorders caused by heterozygous protein-truncating variants in *KMT5B*. K. Takano^{1,2}, T. Fukuyama³, M. Motobayashi⁴, M. Hosoya⁵, T. Yamaguchi^{1,2}, R. Kawamura⁶, K. Wakui^{1,2}, Y. Fukushima¹, T. Kosho^{1,2}. 1) Department of Medical Genetics, Shinshu University School of Medicine, Matsumoto, Nagano, Japan; 2) Center for Medical Genetics, Shinshu University Hospital, Matsumoto, Japan; 3) Neonatology and Developmental Medicine, Shinshu University School of Medicine, Matsumoto, Japan; 4) Division of Neurology, Nagano Children's Hospital, Azumino, Nagano, Japan; 5) Division of Pediatrics, Saku Central Hospital Advanced Care Center, Saku, Japan; 6) Division of Molecular Genetics, Institute for Comprehensive Medical Science, Fujita Health University, Toyoake, Japan.

Mutations in the genes encoding histone lysine methyltransferases (KMTs) or histone lysine demethylases (KDMs) play a role in the pathogenesis of neurodevelopmental disorders (NDDs). The *KMT5B* (*SUV420H1*) gene, encoding a KMT, was first reported to be a candidate gene for autism spectrum disorder (ASD) in 2012. In 2017, *KMT5B* was reported to cause intellectual disability/developmental delay (ID/DD), ASD, febrile seizures, and overgrowth in seven patients (OMIM#617788). Including these reports, only eleven patients have been published. Here, we report a familial case and a de novo case with heterozygous protein-truncating variants in *KMT5B*. **A familial case:** the proband is a 4-years-old boy with infantile hypotonia, mild to moderate DD, autistic features, and epilepsy. His father is healthy. His mother was not good at studying in childhood, though she has never been diagnosed with NDDs. She receives support for housework and child-rearing from her parents now. The elder sister of the proband died from sudden infant death syndrome at age 8 months. Targeted next-generation sequencing (NGS) identified a heterozygous splice-acceptor site variant, c.544-2A>G in *KMT5B* (NM_017635) in the proband and his mother, which resulted in aberrant splicing evidenced by reverse transcriptase-PCR. **A de novo case:** the proband is a 4-years-old boy with ASD, mild DD, and macrocephaly. There is no family history of NDDs or macrocephaly. Targeted NGS identified a de novo heterozygous frameshift variant (c.2260delG: p.D754Ifs*6) in *KMT5B* in the proband. The two novel variants in *KMT5B* identified in this series were predicted to be loss-of-function through disrupting the gene. All three individuals had various degrees of ID/DD with or without ASD/seizures/overgrowth. These findings suggest a high phenotypic variability even in the same family associated with heterozygous *KMT5B* variants. Further study would be needed to delineate the disorder including the genotype-phenotype correlations.

1182W

Targeted next-generation resequencing analysis in 105 Japanese subjects with undiagnosed intellectual disability and multiple congenital anomalies. D.T. Uehara¹, K. Tanimoto², J. Inazawa^{1,3}. 1) Department of Molecular Cytogenetics, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan; 2) Genome Laboratory, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan; 3) Bioresource Research Center, Tokyo Medical and Dental University, Tokyo, Japan.

Intellectual disability (ID) is a complex neurodevelopmental disorder, estimated to affect 1-3% of the general population. ID can be non-syndromic or syndromic, the latter of which is usually referred as multiple congenital anomalies (MCA). Due to the high genetic heterogeneity of ID, the diagnosis of a large subset of cases remains unknown. The present study had the goal of detecting disease-associated single nucleotide variants (SNVs) in ID/MCA subjects previously negative for pathogenic copy number variants, through targeted resequencing by a custom panel of 75 genes. In this panel, 61 genes are the most frequently mutated known genes causing ID/MCA, and the remaining 14 genes were included as candidate genes. In total, we have screened 105 patients, in which causative or likely pathogenic variants in known genes were found in 20% of the cases (21/105). Mutations were found in 19 different genes, in which *DDX3X* and *PPP2R5D* were found mutated in more than one patient. Among the cases with available parental samples, *de novo* mutations were identified in *GNAO1* (p.E246K), *SATB2* (p.W535*), *PPP2R1A* (p.R182W) and *STXBP1* (p.D255Y). As to candidate genes, a missense variant was detected in the *GAP43* (p.A179V) gene. However, as the inheritance could not be determined so far, this variant remains with uncertain significance. The diagnostic yield of 20% is consistent with previous studies and highlights the utility of targeted resequencing for the screening of variants in the most frequently mutated ID/MCA genes.

1183T

Expanding phenotypes of patients with constitutional pathogenic variants in *KAT6A*: Four patients with syndromic intellectual disability. T. Uehara¹, S. Mizuno², A. Kondo³, T. Kirino³, H. Ootani⁴, T. Okazaki⁵, Y. Yamaguchi¹, H. Suzuki¹, Y. Sakaguchi^{1,6}, T. Takenouchi⁶, K. Kosaki¹. 1) Center for Medical Genetics, Keio University School of Medicine, Tokyo, Japan; 2) Department of Pediatrics, Central Hospital, Aichi Human Service Center, Aichi Japan; 3) Department of Obstetrics and Gynecology, Shikoku Medical Center for Children and Adults, Kagawa, Japan; 4) Department of Pediatrics, Tottori prefectural Tottori Prefecture Chuo Hospital, Tottori, Japan; 5) Division of Child Neurology, Department of Brain and Neurosciences, Faculty of Medicine, Tottori University, Tottori, Japan; 6) Department of Pediatrics, Keio University School of Medicine, Tokyo, Japan.

KAT6A, which is a member of the MYST family of histone acetyltransferases, has attracted attention as a gene associated with human malignant diseases. Recently, *KAT6A* has also been reported as a gene associated with human neurodevelopmental disorders. The phenotypic spectrum of *KAT6A* haploinsufficiency has yet to be delineated. Here, we report four unrelated patients with pathogenic variants in *KAT6A*. Patient 1 was an 8-year-old male with a *de novo* heterozygous *KAT6A* p.S985R mutation identified in a trio exome analysis. He had a severe intellectual disability (ID), speech disturbance, and dysmorphic facial features including a smooth philtrum, puffy eyebrows, a wide nasal bridge, epicanthal folds, a high arched palate, and preauricular pits. He also had MP joint contractures, cryptorchidism, and epilepsy. Patient 2 was a 10-year-old male with a *de novo* heterozygous *KAT6A* p.V1425TfsTer13 mutation. He had an ID, an atrial septal defect, a hydronephrosis, and dysmorphic facial features including hypertrichosis, a long nose, posteriorly rotated ears, low-set ears, and blepharophimosis. Patient 3 was a 7-year-old female with a *de novo* heterozygous *KAT6A* p.L1456QfsTer9 mutation. She had an ID, an atrial septal defect, and hypothyroidism requiring thyroid hormone replacement therapy. She also had loose skin and dysmorphic facial features including down-slanted palpebral fissures, a flat nose, and a small chin. Patient 4 was a 2-year-old male with a *de novo* heterozygous *KAT6A* p.R1024X mutation. He had a severe ID and a developmental delay. He also had an atrial septal defect, cryptorchidism, hypospadias, an equinus foot, long fingers, and dysmorphic facial features including low-set ears, a long nose, and left ptosis. The four documented patients had ptosis or blepharophimosis and congenital heart defects, both have been previously reported. In addition, three of the four documented patients had genitourinary problems and one had hypothyroidism. Constitutional mutations in *KAT6B*, which is similar in structure to *KAT6A*, cause neurodevelopmental disorders characterized by developmental delay, intellectual disability, ptosis, blepharophimosis, congenital heart defects, hypothyroidism, and genitourinary anomalies. Thus, patients with *KAT6A* mutations may have phenotypes similar to those of patients with *KAT6B* mutations. In conclusion, both hypothyroidism and genitourinary problems should be included in the expanded phenotypes of patients with *KAT6A* mutations.

1184F

Identifying disorder-specific methylation signatures in patients with severe developmental disorders. J.E. Handsaker, E.J. Gardner, P.J. Short, G. Gallone, L. Parts, J.C. Barrett, M.E. Hurles on behalf of the Deciphering Developmental Disorders study. Human Genetics, Wellcome Sanger Institute, Hinxton, Cambridgeshire, United Kingdom.

Chromatin-modifying genes (CMGs) are a diverse set of genes whose protein products bind to chromatin and regulate cellular processes including proliferation and the regulation of gene expression via modification of histones. Deleterious mutations within CMGs are significantly enriched in patients with severe developmental disorders (DD) and mutations in over 50 CMGs have been robustly associated with DDs. As previous studies have shown that perturbations of the histone code impact DNA methylation, we have characterised methylation profiles in blood-derived DNA from 128 DD probands with *de novo* mutations in one of 22 disease-associated CMGs from the Deciphering Developmental Disorders (DDD) study, a UK-wide cohort of 13,600 children with DD. A median of 5 patient samples were selected per CMG disorder and, along with a selection of age and sex matched negative controls, were methylation profiled using an Illumina EPIC array. Using our methylation data as input to a simple linear correlation model, we sought to identify a CpG methylation signature discriminating between case and control individuals for each CMG disorder. For many CMGs (including *NSD1*, *DNMT3A*, and *EP300*) we were able to establish a signature based on differential methylation, which indicates that such a perturbed phenotype is a general characteristic of CMG disorders, and that these signatures could be diagnostically predictive. We will describe validation of these methylation signatures in additional independent cases run on the same methylation array, as well as quantification of the strength and specificity of these methylation signatures, and exemplification of the diagnostic use of these signatures when the *de novo* status of a variant is not available. We will also describe analyses of the transferability of methylation signatures between centres and across different arrays, using methylation signatures of CMG disorders previously published by other groups. Overall, our study highlights the utility of DNA methylation profiles as biomarkers supporting the diagnosis of CMG disorders.

1185W

Learn as we go: Studying Xia-Gibbs Syndrome through a disease specific registry and community engagement. Y. Jiang^{1,2}, D. Murdock^{1,2}, M. Wangler^{2,3}, J. Posey², Q. Meng^{1,2}, M. Khayat^{1,2}, M. Murugan^{1,2}, A. McGuire⁴, F. Xia², J. Lupszki^{1,2,5,6}, R. Gibbs^{1,2}, 1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 3) Department of Pediatric Radiology, Texas Children's Hospital, Houston, TX; 4) Center for Medical Ethics and Health Policy, Baylor College of Medicine, Houston, TX; 5) Department of Pediatrics, Baylor College of Medicine, Houston, TX; 6) Texas Children's Hospital, Houston, TX.

Background: The rapid increase in the discovery rate of genes underlying rare Mendelian disease has challenged the pace at which we can systematically study phenotypic and genotypic features that yield insight into disease mechanism. We have combined scientific efforts with interactions with families in order to speed our understanding of Xia-Gibbs Syndrome (OMIM: 615829: XGS), a Mendelian disorder discovered in 2014 and characterized by a *de novo* heterozygous loss of function mutation in the *AHDC1* gene located on chromosome 1p36 (Xia *et al.*, 2014). The disorder was first identified in four individuals, but now approximately 80 families worldwide are known via personal communication and social media. Molecular studies and interaction with the families enable exploration of nuances that can both explain laboratory observations and suggest new experiments. **Methods:** We obtained institutional review board (IRB) approval at Baylor College of Medicine to recruit families to an XGS Registry. For families who consent to join, we provide HIPAA compliant secure links to collect clinical information including: clinical reports, MRI imaging, clinical surveys specifically designed to assess XGS related phenotypes. **Results:** As of June 2018, 34 of the 83 known families have joined the registry and 28 have provided detailed clinical information. We also evaluated 19 families through face-to-face meetings or clinical visits. In addition to consolidating the initial phenotypic descriptions (Jiang *et al.*, 2018) we have (i) identified a 56-year-old individual with XGS (the oldest patient previously known was 21) (ii) established a cohort of individuals with *de novo* missense *AHDC1* variants to further study if this is the unique explanation for pathogenicity and (iii) identified cases with deletions involving *AHDC1* to further explore the XGS phenotype. Families have reported under-appreciated clinical features that now warrant study to identify molecular correlates. Among these are (i) irregular sleep-wake rhythm (ii) sensitivity to barometric pressure, (iii) lack of sweat and/or tears, (iv) sometimes hypersocial behavior. **Conclusion:** Systematic research on rare Mendelian disorders is a learning-as-we-go process and requires extensive community and stakeholder (patient, family, physician, researcher) involvement. A disease gene specific registry is an effective platform to systematically study rare Mendelian disorders and has been used here to advance our knowledge of XGS.

1186T

De novo missense variants in *RAC3* cause a novel neurodevelopmental syndrome. D. Chitayat^{1,16}, G. Costain¹, B. Callewaert², H. Gabriel³, T.Y. Tan⁴, S. Walker⁵, J. Christodoulou^{4,6}, T. Lazard⁷, B. Menten², J. Orkin^{8,9,10}, S. Sadedin⁴, M. Snell^{1,11}, A. Vanlander^{1,2}, S. Vergult², S.M. White⁴, S.W. Scherer^{13,14}, R.Z. Hayeems^{10,11}, S. Blaser¹⁵, S.J. Wodak⁷, C.R. Marshall^{5,11,17,18}, M.S. Meyn^{1,8,11,14,19}.

1) Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, Ontario, Canada; 2) Center for Medical Genetics, Ghent University, Ghent, Belgium; 3) Center for Genomics and Transcriptomics, Eberhard Karls University of Tübingen, Tübingen, Germany; 4) Victorian Clinical Genetics Services, Murdoch Children's Research Institute, & Department of Paediatrics, University of Melbourne, Melbourne, Australia; 5) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Ontario, Canada; 6) Neurodevelopmental Genomics Research Group, Murdoch Children's Research Institute, Melbourne, Australia; 7) VIB-VUB Structural Biology Research Center, Pleinlaan 2, 1050 Brussels, Belgium; 8) Department of Paediatrics, University of Toronto, Toronto, Ontario, Canada; 9) Division of Paediatric Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada; 10) Child Health Evaluative Sciences, The Hospital for Sick Children, Toronto, Ontario, Canada; 11) Centre for Genetic Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada; 12) Department of Pediatrics, Ghent University, Ghent, Belgium; 13) Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada; 14) Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada; 15) Department of Diagnostic Imaging, The Hospital for Sick Children, Toronto, Ontario, Canada; 16) The Prenatal Diagnosis and Medical Genetics Program, Mount Sinai Hospital, University of Toronto, Toronto, Ontario, Canada; 17) Genome Diagnostics, Department of Paediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada; 18) Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada; 19) Center for Human Genomics and Precision Medicine, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin, United States of America.

Purpose: *RAC3* is an under-examined member of the Rho GTPase gene family that is expressed in the developing brain and linked to key cellular functions. *De novo* missense variants in the homolog *RAC1* were recently associated with developmental disorders. In the *RAC* subfamily, transforming missense changes at certain shared residues have been observed in human cancers and previously characterized in experimental studies. The purpose of this study was to determine whether constitutional dysregulation of *RAC3* is associated with human disease. **Methods:** We discovered a *RAC3* variant in the index case using whole-genome sequencing, and searched for additional variants using international data-sharing initiatives. Functional effects of the variants were assessed using a multifaceted approach generalizable to most clinical laboratory settings. **Results:** We rapidly identified five individuals with *de novo* monoallelic missense variants in *RAC3*, including one recurrent change. Every participant had severe intellectual disability and brain malformations. *In silico* protein modeling, and prior *in vivo* and *in situ* experiments, supported a transforming effect for each of the three different *RAC3* variants. All variants were observed in databases of somatic variation in cancer. **Conclusion:** Missense variants in *RAC3* cause a novel brain disorder, likely through a mechanism of constitutive protein activation.

1187F

Next generation phenotyping of 7,697 individuals with intellectual disability. S. Jansen¹, J. van Reeuwijk¹, R. van der Donk², C.J.M. van Amen-Helbrekers³, A.T. Vulto-van Silfhout⁴, B.W.M van Bon⁵, D.A. Koolen¹, N. de Leeuw¹, R. Pfundt¹, T. Rinne¹, C. Gilissen¹, H.G. Brunner¹, J.A. Veltman^{1,3}, R.F. Kooy⁴, E.A. Sistermans⁶, J.Y. Hehir-Kwa², L.E.L.M. Vissers¹, L.B.A. de Vries¹.

1) Department of Human Genetics, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, P.O. Box 9101, 6500 HB, Nijmegen, The Netherlands; 2) Princess Máxima Center for Pediatric Oncology, Postbus 113, 3720 AC, Bilthoven, the Netherlands; 3) Institute of Genetic Medicine, International Centre for Life, Newcastle University, Central Parkway, Newcastle, NE1 3BZ, United Kingdom; 4) Department of Medical Genetics, University Hospital and University of Antwerp, Universiteitsplein 1, 2610 Antwerp, Belgium; 5) Department of Clinical Genetics, VU University Medical Center, 1081 HV, Amsterdam, The Netherlands.

Next generation sequencing requires next generation phenotyping. Especially in genetically and clinically heterogeneous diseases, such as intellectual disability (ID), phenotyping should be detailed and standardized. Human Phenotype Ontology (HPO) provides >11,000 standardized terms to describe human disease and has previously been combined with literature on disease-gene annotation to provide potential genotype-phenotype correlations. However, since the specificity of an HPO feature in the ID population is unknown, the value of these observations in novel syndromes remains inconclusive. Here we present the results of a unique phenotype database, with detailed and standardized clinical information of 7,697 individuals with developmental delay (DD)/ID who had a SNP-array and/or whole exome sequencing. For 5,935 individuals (77%), clinical information of the diagnostic request form was used, whereas for 1,762 individuals (23%) all clinical data from the medical record was included. As expected, the group of individuals of whom the medical record was checked was represented by more features per individual than the other group (18.4 and 6.6 respectively, *p*-value <0.05, independent samples *t*-test). For all 3,256 reported HPO features the frequency was calculated, which will provide an essential tool for better evaluation and diagnosis of individuals with novel ID syndromes. To correlate phenotypic features in this ID cohort to the general population, we used the frequencies of 65 major congenital malformations in the general population provided by Eurocat, a European network of population-based registries for congenital anomalies and compared these to the frequencies in our database. 58 of 65 (89%) malformations were reported in the DD/ID cohort and for the majority of these (*n*=36, 62%), the frequency in the DD/ID cohort was significantly enriched, showing that major congenital abnormalities are more often seen in individuals with DD/ID than in the general population. Moreover, we translated the HPO features to a phenotypic score for each individual and confirmed that a higher phenotypic score is correlated to obtaining a conclusive genetic diagnosis (*p*-value <0.05, Kruskal Wallis-test). In conclusion, we provide frequencies of HPO features in a large representative cohort of individuals with DD/ID and show that quantification of phenotyping will not only lead to better interpretation of clinical heterogeneity, but also to better understanding of genomic variation.

1188W

Novel c.5535-1G>A variant in a patient with a mild features of CHARGE syndrome. E. Siavriene¹, G. Petraityte¹, V. Mikstiene^{1,2}, T. Rancellis^{1,2}, Z. Maldziene¹, A. Morkuniene², E. Preiksaitiene^{1,2}, V. Kucinskas¹.

1) Department of Human and Medical Genetics, Institute of Biomedical Sciences, Faculty of Medicine, Vilnius University, Vilnius, Lithuania; 2) Center for Medical Genetics, Vilnius University Hospital Santaros Klinikos, Vilnius, Lithuania.

Background: CHARGE syndrome (MIM# 214800), which is characterized by a number of congenital anomalies including coloboma, ear anomalies, deafness, facial anomalies, heart defects, atresia choanae, genital hypoplasia, growth retardation, and developmental delay, is caused by heterozygous mutation in the *CHD7* (MIM# 608892) on chromosome 8q12. **Clinical report:** In the present study, we report a female individual with a novel heterozygous splice c.5535-1G>A variant in *CHD7* gene, that has not been reported previously. The patient has many characteristic features of CHARGE syndrome such as visual disturbance, hearing impairment, facial anomalies, congenital heart defect, choanal atresia, and delayed puberty. **Methods:** Whole-exome sequencing (WES) followed by functional genome analysis was conducted for the proband with suspected CHARGE syndrome. In order to elucidate the pathogenicity of the detected variant in *CHD7* gene, proband's total blood RNA was isolated and template cDNA was synthesized. Subsequently, *CHD7* gene 26-27 exon junction was sequenced in cDNA sample using Sanger sequencing. **Results:** Heterozygous *CHD7* gene variant c.5535-1G>A located in the acceptor splice site of intron 26 was identified in proband's DNA sample after analysis of WES data. *In silico*, the alteration of splice acceptor site was predicted to affect pre-mRNA splicing. Sanger sequencing of cDNA revealed that c.5535-1G>A disrupts the original splice acceptor site and activates a cryptic splice site only one nucleotide downstream of the pathogenic variant site. This change causes the deletion of the first nucleotide of exon 27 and leads to a frameshift in cDNA and protein. Our results suggest that the alteration leads to the premature truncation of the protein thus resulting in CHARGE syndrome. **Conclusion:** Functional genomic approach of cDNA analysis provides a unique possibility to understand the etiology and pathophysiology as well as to identify the molecular basis of many hereditary diseases and conditions including those associated with congenital anomalies. The work was funded by the Research Council of Lithuania (No. S-MIP-17-19/LSS-150000-1179, Ingenu project).

1189T

3-Hydroxy-3-methylglutaryl-coenzyme A lyase deficiency presenting with rhabdomyolysis. E.E. Conboy¹, M.K. Myers², P. Rinaldo¹, K.M. Raymond¹, D. Matern¹, S. Tortorelli¹, D. Gavrilov¹, B.C. Lanpher³. 1) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA; 2) Critical Care Medicine (Internal Medicine), Mayo Clinic, Rochester, MN, USA; 3) Department of Clinical Genomics, Mayo Clinic, Rochester, MN, USA.

3-Hydroxy-3-methylglutaryl-coenzyme A lyase deficiency (HMGCLD) is a rare inborn error of ketone body synthesis and leucine degradation. The mitochondrial enzyme 3-Hydroxy-3-methylglutaryl-coenzyme A lyase converts 3-Hydroxy-3-methylglutaryl-coenzyme A to acetoacetate and acetyl-coenzyme A. Patients with this disorder are prone to hypoketotic hypoglycemia during a metabolic disturbance. Our patient is a 29 year old male with known HMGCLD and a history of seizures associated with hypoglycemia, last occurring approximately 10 years prior to presentation. He was found down in the field with hypoglycemia (glucose of 20 mmol/L). He was intubated and sedated with propofol for several days. He had significant rhabdomyolysis with a creatine kinase above the detectable limit of 197,500 U/L causing renal failure, and was transferred to our institution for further management. At the time of his hospitalization a plasma acylcarnitine profile was obtained. The concentrations of 3-hydroxyisovaleryl-/2-methyl-3-hydroxybutyryl- carnitine (C5-OH) and 3-methylglutaryl carnitine (C6-DC) were prominently elevated, consistent with the known diagnosis of 3-hydroxy 3-methylglutaryl-(HMG)-CoA lyase deficiency. Rhabdomyolysis is not a known complication of HMGCLD. It is a component of propofol infusion syndrome (PRIS), a combination of findings including metabolic derangements and organ failure that can occur in patients receiving propofol. The current theory regarding the pathophysiology of PRIS proposes that propofol inhibits intracellular energy production by inhibiting the membrane-bound transporter carnitine palmitoyl transferase I (CPT1) as well as a direct inhibitory effect on the mitochondrial respiratory chain. Since HMG-CoA lyase deficiency is also an abnormality of fatty acid oxidation, it may be that patients with this rare genetic disorder are predisposed to PRIS. It seems reasonable to hypothesize that the propofol infusion may have exacerbated or precipitated rhabdomyolysis. This may warrant avoiding propofol in patients with HMG CoA Lyase deficiency.

1190F

A multicenter, double-blind, placebo-controlled, multiple-dose study assessing the safety and efficacy of MNK-1411 in patients with Duchenne muscular dystrophy. *B.R. Due¹, P. Becker², T.O. Crawford³.* 1) Mallinckrodt ARD Inc., Bedminster, New Jersey; 2) Former employee of Mallinckrodt ARD Inc., Bedminster, New Jersey; 3) Department of Neurology, Johns Hopkins University, Baltimore, Maryland, United States.

Background: Duchenne muscular dystrophy (DMD) is a recessive X-linked neuromuscular disorder caused by mutations in the gene for dystrophin, a structural protein essential for the normal development and survival of muscle. Frame-shifting and nonsense mutations in the dystrophin gene are associated with DMD. As the most common muscle disease in children, DMD typically manifests prior to 6 years of age, and results in progressive muscle degeneration and loss of ambulation. Corticosteroids are the only therapy with demonstrable effectiveness that are approved for treatment of all patients with DMD indifferent to the type of mutation. There are substantial off-target effects associated with long-term use of corticosteroids, however, highlighting the need for better therapies. MNK-1411 is a 24-amino-acid synthetic adrenocorticotrophic hormone analogue and melanocortin receptor (MCR) agonist with the potential to slow DMD progression via MCR-mediated properties by reducing inflammation and/or attenuating muscle damage. **Objective:** To evaluate the safety and efficacy of MNK-1411 in male patients aged 4 to 8 years with DMD. **Methods:** This is a phase 2, multicenter, double-blind, placebo-controlled, multiple-dose study (ClinicalTrials.gov Identifier: NCT03400852). Patients randomized 2:2:1:1 into the 24-week double-blind phase will receive weight-based doses of subcutaneous MNK-1411 or volume-matched placebo (2x/week). The 4 treatment groups comprise Group A, 0.5-mg MNK-1411 (0.5 mL for patients >20 kg) or 0.4-mg MNK-1411 (0.4 mL for patients ≤20 kg); Group B, 0.25-mg MNK-1411 (0.25 mL for patients >20 kg) or 0.2-mg MNK-1411 (0.2 mL for patients ≤20 kg); Group C, 0.5-mL or 0.4-mL placebo; and Group D, 0.25-mL or 0.2-mL placebo. Patients who complete the blinded phase will be eligible to enter a 24-week open-label extension phase and continue receiving the same volume of MNK-1411. The primary efficacy endpoint is change from baseline in the 10-meter walk/run at Week 24. Adverse events will be monitored throughout the study. **Results:** Approximately 130 patients will be enrolled at 50 global sites; the trial will have 80% power to detect a treatment difference between groups at a significance level of 0.05. **Conclusions:** This study will potentially provide data to support the safety and efficacy of MNK-1411 for treatment of DMD.

1191W

Persistent metabolic abnormalities and hypertrophic cardiomyopathy in a patient with multiple acylCoA dehydrogenase deficiency associated with a frameshift variant in *SLC52A1*. *R.M. Zambrano¹, H.R. Meddaugh¹, S. Sernich², M. Marble¹.* 1) Department of Pediatrics, Division of Genetics, Louisiana State University Health Science Center, New Orleans, Louisiana; 2) Department of Pediatrics, Division of Pediatric Cardiology, Louisiana State University Health Sciences Center, New Orleans, Louisiana.

Multiple acylCoA dehydrogenase deficiency (MADD) is associated with elevations of metabolites of compounds oxidized by enzymes that depend on Flavin Adenine Dinucleotide (FAD) as a cofactor. MADD can be caused by pathogenic variants in a number of genes involved in riboflavin metabolism and transport including: *ETF*, *ETFDH*, *FLAD1*, *SLC25A32*, *SLC52A1*, *SLC52A2* and *SLC52A3*. MADD associated with pathogenic variants in *SLC52A1* is rare, having been reported in only 2 patients: a transient form associated with maternal riboflavin deficiency (Chiong et al., 2007; Ho et al., 2011) has been reported and the patient responded well to riboflavin supplementation; and a neonate with a maternally inherited intronic variant in *SLC52A1* (Mosegaard et al., 2017). This patient's MADD reportedly resolved with riboflavin supplementation which was discontinued at 30 months of age. MADD phenotype is quite variable and ranges from severe neonatal metabolic crisis to later onset forms presenting with weakness. Some patients develop hypertrophic cardiomyopathy (HCM). We report a new case of MADD associated with a maternally inherited frameshift variant in *SLC52A1* in a neonate with HCM and newborn screening suggestive of MADD. Fat and protein restriction were initiated along with riboflavin and carnitine supplements. The HCM improved, however despite elevated riboflavin levels, metabolic testing continued to show elevated metabolites. She was stable until 4 months old of age when she presented with respiratory failure, pericardial effusion, weakness and worsening HCM and a persistent MADD metabolic profile after a viral infection. We propose that pathogenic variants in *SLC52A1* cause some cases of MADD, with variable severity. The fact that the variant was inherited from an asymptomatic mother in our case and in Mosegaard's case suggests that pathogenic variants in *SLC52A1* are of reduced penetrance and may require other factors (environmental or genetic) to cause disease. Whole exome sequencing failed to identify any other explanation for the phenotype.

1192T

Synergistic effect of mutations in dystroglycanopathies-associated genes? A Brazilian case report.

M. Lazar¹, K.M. Rocha¹, M.C. Varela¹, G.L. Yamamoto^{1,2}, V.N. Takahashi¹, K. Bagatini¹, S. Ezquina¹, M.O. Scliar¹, J.Y.T. Wang¹, R. Pavanello¹, A.S. Bulle Oliveira³, M. Zatz¹, M. Vainzof¹, M.R. Passos-Bueno¹. 1) Human Genome and Stem Cell Research Center (HUG-CELL), University of São Paulo, São Paulo, Brazil; 2) Department of Clinical Genetics, Children's Hospital, Medical School, University of São Paulo, Brazil; 3) Department of Neurology, Neuromuscular Disorders Unit, Federal University of São Paulo, Brazil.

HUG-CELL has implemented targeted massively parallel sequencing (MPS) for diagnosis of several human diseases. Although the high sensitivity rate of ~70% experienced by our laboratory for neuromuscular disorders (NMD), some patients remain undiagnosed. Recently, a 43 year-old-woman presenting muscle weakness was referred to our center for genetic investigation. According to the neurologist request, the patient has relatives affected with FSHD1, and despite she presents proximal tetraparesis and myopathic EMG, normal D4Z4 alleles were reported. Informed consent was signed, in accordance with our ethical regulations. Using a commercial panel with ~6700 genes (TruSight One Expanded - Illumina), we focused our analysis on those involved in NMD. After variant filtering, three heterozygous variations in genes associated with autosomal-recessive dystroglycanopathies were prioritized: *GMPPB* [NM_013334.3: c.859C>T p.(Arg287Trp); pathogenic rs142908436], *POMT2* [NM_013382.5: c.1271A>G p.(His424Arg); located in MIR2 domain] and *DAG1* [NM_004393.5: c.2571T>G p.(Asp857Glu); located in cytoplasmic topological domain of β -dystroglycan] (hg19/HGVS). Both *POMT2* and *DAG1* alterations were absent from disease-associated databases as well as from public variant repositories (gnomAD, ABraOM), and could have deleterious effects, according to multiple lines of computational evidence. Considering that: (1) mannose-1-phosphate guanyltransferase β and protein O-mannosyltransferase 2 (respectively encoded by *GMPPB* and *POMT2*) are implicated in dolichol-phosphate mannose (Dol-P-Man) synthesis and (2) the relevance of Dol-P-Man as a mannose-donor in mannosylation processes of proteins like α - and β -dystroglycan (encoded by *DAG1*), we hypothesized the identified variants could be interacting in a synergistic manner. Since the available data are insufficient to conclude the molecular diagnosis of the patient, we intend to perform: (1) segregation analysis of those variants in her parents and (2) protein analysis on her muscle biopsy to verify the possible occurrence of defective glycosylation process and/or secondary deficiency of the dystrophin-associated glycoprotein complex. The usage of MPS in clinical practice has increased the detection of patients who carry variants in functionally correlated genes and therefore reinforces the requirement for complementary studies to determine the deleteriousness of concurrent alterations. Financial support: Fapesp/CEPID and CNPq/INCT.

1193F

KI/KO mice carrying a BAG3 mutation homologous to the human P209L myofibrillar myopathy mutation develop a clinical and pathological skeletal muscle phenotype.

R. Robertson¹, T. Conte¹, M.J. Dicaire¹, R. Bryson-Richardson², J.N. Lavoie³, E. O'Ferrall¹, J.C. Young¹, B. Brais¹. 1) McGill, Montreal, Quebec, Canada; 2) Monash University, Australia; 3) Laval University, Quebec, Quebec.

BAG3 is a co-chaperone to heat-shock proteins important for the degradation of misfolded proteins through chaperone assisted selective autophagy. The recurrent dominant BAG3^{P209L} mutation results in a severe myofibrillar myopathy associated with progressive muscle weakness, cardiomyopathy and respiratory failure. Since a homozygote knock-in line (KI) for the mP215L mutation homologous to the human P209L mutation did not have a gross phenotype, we generated heterozygote knock-out (KO) and KI mP215L mice to establish if a further loss of BAG3 expression would lead to a phenotype. Indeed, the KI/KO mice show a significant decrease in voluntary movement compared to wildtype KI/KI mice in the open field starting at 6 months, reminiscent to the decreased activity observed in aging patients. Pathologically, the KI and KI/KO lines both show significantly smaller fiber size, while only the KI/KO mice show clear skeletal muscle histological changes consisting of: centralized nuclei, splitting fibers and inflammatory infiltrate in particular at nerve entry zones and myotendinous junctions. These changes are increased following downhill running only in the KI/KO mice. The KI/KO mP215L line is the first murine model of BAG3 myopathy that may resemble the human skeletal muscle condition. It supports the data generated in zebrafish that to develop the myopathy some significant loss of the protein is needed and not only a gain of function mutation. The observation that the KI/KO mice are more vulnerable to muscle injury may explain in part the slow progression of this myopathy in humans. The further characterization of the BAG3 KI/KO mice will provide insight into the pathophysiology of this myofibrillar myopathy and may demonstrate its usefulness for pre-clinical trials.

1194W

Dilated cardiomyopathy: A novel finding in a patient with ADSSL1-related myopathy. D. Bonner¹, J.N. Kohler¹, M. Majcherska¹, D.B. Zastrow¹, C. Reuter¹, L. Fernandez¹, C. McCormack¹, S. Marwaha¹, C. Curnin¹, L. Fre-sard², K.S. Smith², E. Worthey², UDN. Undiagnosed Diseases Network⁴, J.B. Sampson^{1,5}, S.B. Montgomery⁶, P.G. Fisher^{1,5}, E.A. Ashley^{1,6,7}, J.A. Bernstein^{1,6,8}, M.T. Wheeler^{1,7}. 1) Stanford Center for Undiagnosed Diseases, Stanford, CA; 2) Department of Pathology, School of Medicine, Stanford, CA; 3) HudsonAlpha Institute for Biotechnology, Huntsville, AL; 4) NIH Undiagnosed Diseases Network, Office of the Director and the National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 5) Department of Neurology, School of Medicine, Stanford, CA; 6) Department of Genetics, School of Medicine, Stanford, CA; 7) Division of Cardiovascular Medicine, School of Medicine, Stanford, CA; 8) Department of Pediatrics – Medical Genetics, School of Medicine, Stanford, CA.

ADSSL1 encodes a muscle and cardiac-specific enzyme involved in the interconversion of inosine monophosphate and adenosine monophosphate in the purine nucleotide cycle. Deficiency of *ADSSL1* is thought to impair energy production, metabolic regulation, and synthesis of DNA and RNA. Biallelic loss of function variants in *ADSSL1* have recently been established as the cause of distal myopathy 5 (MPD5; OMIM:617030), an autosomal recessive condition characterized by adolescent onset distal myopathy, facial muscle weakness, elevated serum CK, and rimmed vacuoles. To date, only 3 pathogenic variants have been reported in 9 patients. We evaluated a 33-year-old male at the Stanford Center for Undiagnosed Diseases for slowly progressive myopathy, with involvement of the proximal and distal limb, ocular, facial, ventilatory and cardiac musculature. He presented to care at 23 years old with proximal muscle weakness and elevated CK, although he noted onset of diffuse weakness in childhood. Echocardiogram at 28 years old revealed mild global hypokinesis, decreased left ventricular function (EF=44%) and mild right ventricle dilation. Muscle biopsy showed type 2 fiber predominance, necrotic fibers, rimmed vacuoles and foci of chronic inflammation. His physical exam was notable for atrophic scarring and contractures of the deep finger flexors and elbows. Extensive testing for inflammatory and genetic etiologies were unrevealing. Genome sequencing (GS) revealed a homozygous pathogenic variant in the *ADSSL1* gene: c.910G>A, p.D304N. All 9 patients reported with MPD5 carry the c.910G>A variant in *trans* with a second pathogenic variant. While many aspects of our patient's phenotype overlap with MPD5, dilated cardiomyopathy (DCM) was not reported among the 7 of 9 patients screened with echocardiogram. Targeted-analysis of our patient's GS data did not reveal a pathogenic variant related to DCM. Thus, the association of DCM with the c.910G>A variant remains unclear. Functional studies including RNA sequencing and in vitro assay using iPSC-derived cardiomyocytes and skeletal muscle are ongoing. This case contributes to the emerging phenotype associated with biallelic loss of function variants in *ADSSL1*. Functional studies will clarify the association with DCM and underlying mechanism of disease. Ongoing cardiac evaluation of patients with MPD5 will help determine the natural history of cardiac muscle involvement.

1195T

Homozygous recessive mutations in the myogenic factor MYF5 cause a novel syndrome characterized by external ophthalmoplegia, rib, and vertebral anomalies. S.A. Di Gioia^{1,2,3,4}, S. Shaaban^{1,2,3,14}, B. Tüysüz⁵, N.H. Elcioglu^{6,7}, W.M. Chan^{1,2,4,8}, C.D. Robson^{9,10}, K. Ecklund^{9,10}, N.M. Gillette^{1,2}, A. Hamzaoglu¹¹, G.A. Tayfun⁴, E.I. Traboulsi¹², E.C. Engle^{1,2,3,4,8,13}. 1) Department of Neurology, Boston Children's Hospital, Boston, MA 02115, U.S.A.; 2) F.M. Kirby Neurobiology Center, Boston Children's Hospital, Boston, MA 02115, U.S.A.; 3) Department of Neurology, Harvard Medical School, Boston, MA 02115, U.S.A.; 4) Broad Institute of M.I.T. and Harvard, Cambridge, MA 02142, U.S.A.; 5) Istanbul University, Cerrahpasa Medical School, Department of Pediatric Genetics; 6) Department of Pediatric Genetics, Marmara University Medical School, Istanbul, Turkey; 7) Eastern Mediterranean University Medical School, Cyprus, Mersin 10, Turkey; 8) Howard Hughes Medical Institute, Chevy Chase, MD 20815, U.S.A.; 9) Department of Radiology, Boston Children's Hospital, Boston, MA 02115, U.S.A.; 10) Istanbul Spine Center at Florence Nightingale Hospital, Abide-i Hurriyet Cad. No:166, Sisli, 34381, Istanbul, Turkey; 11) Cole Eye Institute, Cleveland Clinic, Cleveland, OH 44195, U.S.A.; 12) Department of Ophthalmology, Boston Children's Hospital, Boston, MA 02115, U.S.A.; 13) Department of Ophthalmology, Harvard Medical School, Boston, MA 02115, U.S.A.; 14) Sanford Imagenetics, Sioux Falls, SD, 57105.

MYF5 is member of the Myc-like basic helix-loop-helix transcription factor family and, in cooperation with other myogenic regulatory factors MYOD and MYF5, is a key regulator of early stages of myogenesis. Here, we report three consanguineous families with biallelic homozygous loss-of-function mutations in *MYF5* who define a new clinical disorder characterized by congenital ophthalmoplegia with scoliosis, vertebral, and rib anomalies. The clinical phenotype overlaps strikingly with that reported in several *Myf5* knockout mouse models. Affected members of two families share a haploidentical region that contains a homozygous 10bp frameshift mutation in exon 1 of *MYF5* (NM_005593.2: c.23_32delAGTTCTCACC, [p.Gln8Leufs*86]) predicted to undergo nonsense-mediated decay. Affected members of the third family, who have external ophthalmoplegia harbor a homozygous missense change in exon 1 of *MYF5* (NM_005593.2: c.283C>T, [p.Arg95Cys]). Using *in-vitro* assays we show that this missense mutation acts as a loss-of-function allele by impairing MYF5 DNA binding and nuclear localization. We performed whole genome sequencing in one affected individual with the frameshift mutation and did not identify additional rare variants in the haploidentical region that might account for differences in severity among the families. These data support the direct role of *MYF5* in rib, spine, and extraocular muscle formation in humans.

1196F

Possible effect of sarcolemmal complex genes on the phenotype of Duchenne muscular dystrophy. M. Garcia-Acero^{1,3}, T. Gamez¹, T. Buitrago¹, M. Guerra¹, R. Garcia-Robles², F. Suarez^{1,3}, P. Ayala-Ramirez¹. 1) Instituto de Genética Humana, Pontificia Universidad Javeriana, Bogota, Colombia; 2) Departamento de Ciencias Fisiológicas, Facultad de Medicina, Pontificia Universidad Javeriana, Bogota, Colombia; 3) Servicio de Genética, Hospital Universitario San Ignacio, Bogota, Colombia.

Introduction: Duchenne muscular dystrophy (DMD), an X-linked recessive inheritance condition, caused by the absence of dystrophin, is characterized by atrophy and progressive muscle weakness. Despite the same altered gene, individuals with DMD have wide clinical variability, initially associated to the altered isoform related to the exon involved, however, individuals with the same mutation have a different clinical course. Given the complexity of the sarcolemma and considering the presence of multiple proteins involved in the maintenance and repair of muscle fiber, we have considered the search for variants in other genes considering a possible oligogenic effect modifier of the phenotype. **Methodology:** Next-generation Sequencing analysis of genes involved in the sarcolemmal complex in individuals with a pathogenic variant in the *DMD* gene using the Invitae Comprehensive Muscular Dystrophy Panel, which analyze sequence changes and exonic deletions/duplications in 55 genes associated with muscular dystrophies. Kaplan-Meier analysis of the clinical course in the presence of other variants. **Results:** 62 patients with pathogenic variants in *DMD* gene were studied. Most patients with codon stop mutations (43.13%), followed by frameshift mutation (27.45%) and intron variants (27.45%), the remaining percentage corresponds to individuals with missense variants (1.97%). In 42 (67.74%) individuals, at least one variant (uncertain significance, probably pathogenic and pathogenic) was documented in another gene, the most frequent being in *PLEC* (n=8), *SYNE1* (n=13) and *SYNE2* (n=7) genes. We mentioned two patients with the same pathogenic variant c.2804-1delG in the *DMD* gene, one of them with additional variants in the *SYNE1* and *FKTN* genes, with early developmental delay and cardiac conduction defect since the age of 4 years. The other patient with other variants in the genes *DMD*, *DES*, *DYSF*, *LAMA2*, *PLEC* and *POMT2* currently with 8 years with difficulty walking but without cardiac compromise. Other patients of 22 and 23 years with the pathogenic variant c.9040dupC in *DMD* and a in *SYNE1* currently with limitation for walking without cardiac or respiratory compromise. **Discussion:** the specific effect of variants in other associated to the sarcolemmal complex in individuals with Duchenne muscular dystrophy is unknown; however, the wide clinical variability may be partially explained due to the dysfunction of other muscle proteins, predominantly involved in muscle repair.

1197W

Recurrent rhabdomyolysis and hyperCKemia: Consideration of broad molecular testing such as WES/WGS is indicated due to the broad differential of conditions. L. Massingham^{1,2}, M.K. Cushing^{1,2}, J. Schwab^{1,2}, W. Campbell^{1,2}, M. Mori^{1,2}, C. Phornphutkul^{1,2}. 1) Department of Pediatrics, Division of Genetics, Hasbro Children's Hospital/Rhode Island Hospital, Providence, RI; 2) Alpert Medical School of Brown University, Providence, RI.

Recurrent rhabdomyolysis and and hyperCKemia without muscle overt muscle weakness can be challenging cases as there are no clear guidelines for best practice. These are often difficult cases as the differential is broad and includes acquired causes, ethnic variation, autoimmune, lab error, and of course hereditary conditions. Even among the hereditary conditions the pathology can stem from metabolic, muscular, or endocrine origin. There are multiple different next generation sequencing panels for rhabdomyolysis and metabolic myopathies, but the genes included in these panels vary greatly by laboratory. We present multiple cases of hyperCKemia without weakness, rather in the setting of episodic pain or rhabdomyolysis, who underwent broad genetic testing, such as whole exome sequencing. For example, a 16 year old girl presented to our clinic with multiple episodes of rhabdomyolysis with CKs >20,000 after strenuous basketball practice. Whole exome sequencing identified a partial deletion of *DMD* involving at least exons 8-11, consistent with her being a "manifesting carrier". She does not report muscle weakness symptoms or exhibit muscle weakness on examination. She doesn't meet the criteria to consider Eteplirsén and the side effects of steroids were thought to outweigh the benefits. The *DMD* gene is not included in all rhabdomyolysis NGS panels. This unexpected diagnosis significantly changes her management, fortunately she does not have any cardiac manifestations of Duchenne muscular dystrophy at this time but she will need to continue with serial surveillance. Another example is a 20 year old young man with elevated CK incidentally found on routine bloodwork. His symptoms included daytime sleepiness and muscle weakness and fatigue after walking for 1 hour. Muscle biopsy revealed no evidence of myositis or dystrophy and no increase in glycogen. Metabolic workup was normal. He previously had a myopathy panel which only revealed one pathogenic variant in the *TK2* gene. Whole exome sequencing identified compound heterozygous changes in the *ANO5* gene (c.191dupA and c.1963T>C). These cases demonstrate the importance of consideration of inherited causes for rhabdomyolysis and hyperCKemia. Broad testing, such as whole exome sequencing or whole genome sequencing should be considered in these patients, even if a targeted panel has already been completed and is negative or inconclusive.

1198T

Diagnosis of congenital myopathies through exome sequencing and a collaborative analysis platform. Z.M. Valivullah^{1,2}, A.K. Lovgren^{1,2}, C. Genetti³, M. Wilson^{1,2}, J. Chang^{1,2}, S. Baxter^{1,2}, H. Arachchi^{1,2}, H. Snow^{1,2}, B. Weisburd^{1,2}, A.H. Beggs³, H.L. Rehm^{1,2}, A.H. O'Donnell-Luria^{1,2,3}, D. MacArthur^{1,2}.

1) Broad Center for Mendelian Genomics and Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA; 2) Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA; 3) The Manton Center for Orphan Disease Research, Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA, USA.

The Broad Institute Center for Mendelian Genomics (CMG) performs next generation sequencing to discover new disease genes underlying Mendelian disorders. In collaboration with investigators worldwide, the CMG analyzes diverse rare disease cohorts, including neuromuscular disorders. Congenital myopathies compose a tenth of all neuromuscular disorder cases and are found in around 1 in 16,000 live births. The typical phenotype includes congenital hypotonia and generalized skeletal muscle weakness that is non- or slowly progressive, often accompanied by facial weakness and respiratory insufficiency. Using our open-source, web-based analysis platform, *seqr*, we have achieved an overall diagnostic rate of around 30% through exome sequencing in our neuromuscular cohorts, which include many patients with prior negative gene panel testing. We have identified several high-confidence discoveries, including a novel mode of inheritance for *KDM5B* and a novel phenotype association for *MYLK*. Histone demethylase *KDM5B* has recently been associated with an autosomal recessive syndrome characterized by developmental delay, facial dysmorphism, and camptodactyly. We found a rare de novo nonsense variant in *KDM5B* in a patient with delayed gross motor development, hypotonia, arachnodactyly, and predominance of small type 1 myofibers. If proven pathogenic, this will expand the previously-reported phenotype and mode of inheritance for *KDM5B* variants. *MYLK*, a myosin light chain kinase, has been previously associated with autosomal dominant aortic aneurysm and intestinal/urinary smooth muscle defects. In a patient with muscle weakness resulting in mobility impairment, we discovered compound heterozygous missense and splice variants in *MYLK*. These candidate genes have been submitted to the Matchmaker Exchange through *matchbox* integration in *seqr*. *Seqr* is proving to be an effective collaboration tool in diagnosing patients and discovering new Mendelian disease genes. It features a user-friendly interface that facilitates collaborative analysis by providing a reproducible filter-based variant search; variant- and gene-specific information and links; and variant tagging and annotation to highlight and track discoveries. In the future, we plan to implement analysis of additional types of variation in *seqr*, including uniparental disomy and copy number and structural variation.

1199F

Expanding the phenotype of *SLC25A42*-associated mitochondrial encephalomyopathy. M. Almannai¹, A. AlAsmari¹, A. Alqasbi², E. Faqeh¹, F. AlMutairi³, M. AlOtaibi⁴, M. Samman⁴, W. Eyaid⁵, Y. Aljadhai⁶, H. Shamseldin⁶, W. Craigen⁷, F.S. Alkuraya^{6,8,9}. 1) King Fahad Medical City, Riyadh, Saudi Arabia; 2) Children's Hospital, King Saud Medical City, Riyadh, Saudi Arabia; 3) Medical Genetics Division, Department of Pediatrics, King Abdullah International Medical Research Centre, King Saud bin Abdulaziz University for Health Sciences, King Abdulaziz Medical City, Riyadh, Saudi Arabia; 4) Pathology and Clinical Laboratory Administration, Section of Molecular Pathology, King Fahad Medical City, Riyadh, Saudi Arabia; 5) Department of Neuroimaging and Intervention, Medical Imaging Administration, King Fahad Medical City, Riyadh, Saudi Arabia; 6) Department of Genetics, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia; 7) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas; 8) Department of Anatomy and Cell Biology, College of Medicine, Alfaisal University, Riyadh, Saudi Arabia. Saudi Human Genome Program, King Abdulaziz City for Science and Technology, Riyadh, Saudi Arabia; 9) Saudi Human Genome Program, King Abdulaziz City for Science and Technology, Riyadh, Saudi Arabia.

SLC25A42 gene encodes an inner mitochondrial membrane protein that imports Co-enzyme A (CoA) into the mitochondrial matrix. A mutation in this gene was recently reported in a subject born to consanguineous parents who presented with mitochondrial myopathy with muscle weakness and lactic acidosis. In this report, we present 12 additional individuals with the same founder mutation who presented with variable manifestations ranging from asymptomatic lactic acidosis to a severe phenotype characterized by developmental regression and epilepsy. Our report confirms the link between *SLC25A42* and mitochondrial disease in humans, and suggests that pathogenic variants in *SLC25A42* should be interpreted with the understanding that the associated phenotype may be highly variable.

1200W

Expanding the clinical spectrum of EARS2-associated mitochondrial disease. P. Prasun¹, B. Webb¹, K. Oishi¹, C. Mintz¹, H. Li². 1) Icahn School of Medicine at Mount Sinai, New York, NY; 2) Department of Human Genetics, Emory School of Medicine, Atlanta, GA.

Introduction: *EARS2* is associated with leukoencephalopathy with thalamus and brain stem involvement and high lactate (LTBL). All previously described cases had characteristic MRI findings of diffuse white matter changes and symmetrical signal abnormalities in the thalamus and brain stem. Clinical presentations fall into two groups- severe and mild. The patients in the severe group present before 6 months with marked neurological regression and then clinical stagnation. Patients in the mild group present later in infancy with mild neurological regression but they can regain some milestones. Here, we describe four patients with atypical and distinctive presentations. **Case studies:** Patient 1 is a 6 month old girl who presented at 3 months of age with seizures. Her growth and development were age-appropriate. She had hepatomegaly, abnormal liver function tests, and high lactate. Abdominal ultrasound showed an enlarged liver and MRI of the brain showed restricted diffusion in the white matter and corpus callosum. Basal ganglia and posterior fossa structures were normal. Two mutations in the *EARS2* gene, c.322C>T (p.R108W) and c.328 G>A (p.G110S) were identified. At 6 months of age, her physical growth, language and social development were normal, however, her motor development showed mild delay. Both lactate and liver enzymes were trending down. Patient 2 & 3 are female siblings (currently 2 and 5 years old), who were identified by newborn screening for SCID. They had persistent T cell lymphopenia, lactate elevation, and macrocytic anemia. Additionally, older sibling had failure to thrive and mild developmental delay before 2 years of age, which has resolved now. Bi-allelic, novel, likely pathogenic variants, c.485+4090_511del5090 and c.667G>A, p.D223N in *EARS2* were detected in both. Patient 4 is 3 year old male with normal development until 12 months of age when he developed rapid regression of milestones. Now he is slowly gaining skills. MRI of the brain at the onset of symptoms was normal. Plasma lactate was elevated to 3.5 at presentation (repeat 1.4). WES identified two VUS in *EARS2* gene. **Conclusion:** Here, we describe patients with *EARS2* mutations with developmental delay, liver abnormalities, macrocytic anemia, and lymphocytopenia without typical "LTBL" presentation. Thus, the clinical spectrum/phenotype of *EARS2*-associated disease is wider than currently recognized. Moreover, *EARS2*-related lymphocytopenia could be detected by newborn screening for SCID. .

1201T

In-Vivo oxidation of 1-¹³C-propionate as a surrogate endpoint for clinical outcomes in propionic acidemia. O. Shchelochkov¹, A. Pass¹, S. Ferry¹, C. Van Ryzin¹, C. Chlebowski², J. Snow², K. Chen³, J. Sloan¹, I. Manoli¹, A. Thurm², C.P. Venditti¹. 1) NHGRI, Bethesda, MD; 2) NIMH, Bethesda, MD; 3) NIDDK, Bethesda, MD.

Propionic acidemia (PA) is an organic acidemia caused by the reduced activity of propionyl-CoA carboxylase (PCC). PA is a multi-systemic disorder characterized by accumulation of propionyl-CoA-derived metabolites in body tissues. To date, our ability to predict clinically relevant outcomes based on genotype or biochemical markers has been lacking. We hypothesized that *in vivo* 1-¹³C-propionate oxidation could represent a surrogate biomarker for enzyme activity, and therefore would closely correlate with severity of pathogenic variants, biochemical parameters, growth, and cognitive functioning in patients with PA. Through a dedicated natural history study (The Natural History, Physiology, Microbiome and Biochemistry Studies of Propionic Acidemia, ClinicalTrials.gov ID NCT02890342), we evaluated 27 PA patients, mainly in the in-patient setting. The oxidation of 1-¹³C-propionate was measured in 24 patients. After a single oral or gastric bolus, ¹³CO₂ enrichment in expired total CO₂ was analyzed using isotopic ratio mass spectroscopy, and the cumulative percentage of the isotopic dose metabolized over two hours was determined. Baseline VCO₂ production (ml/min) was measured by indirect calorimetry. A previous set of healthy volunteers (n=19) served as controls. Compared to healthy controls (47.7 ± 4.5%, mean±SD), PA patients had decreased oxidation of 1-¹³C-propionate to ¹³CO₂ (15.9 ± 13.4% of cumulative dose oxidized, mean ± SD, P<0.0001). Homozygous or compound heterozygous nonsense variants in *PCCA* or *PCCB*, genes that encode the PCC enzyme, were associated with lower 1-¹³C-propionate oxidation (5.7 ± 0.7%) compared to patients with at least one missense or splice site pathogenic variant (19.76±13.1%, P=0.02). 1-¹³C-Propionate oxidation correlated with propionate metabolites including plasma propionylcarnitine (R²= 0.53, P=0.0004) and plasma total methylcitrate (R²= 0.46, P= 0.003) as well as height Z-scores (R²=0.33, P=0.007) and full-scale intelligence quotients (R²=0.55, P=0.007). The two-hour cumulative oxidation of 1-¹³C-propionate in a PA patient with two non-sense *PCCB* alleles, who underwent a liver transplant, was 64.2%. 1-¹³C-propionate oxidation was strongly associated with genetic, biochemical, and clinical parameters in PA patients compared to controls, suggesting that this measurement could represent an appropriate surrogate endpoint in therapeutic trials aiming to improve clinical outcomes by increasing propionyl-CoA carboxylase activity.

1202F

miR-181a/b downregulation protects from mitochondria-associated neurodegeneration.

B. Franco^{1,2}, **A. Indrieri**^{1,2}, **S. Carrella**^{1,3}, **A. Romano**¹, **F.M. Golia**¹, **M. Pizzo**¹, **R. Tammaro**¹, **E. Marrocco**¹, **N. Giordano**^{1,4}, **A. Carboncino**^{4,5}, **A. Spaziano**¹, **L. Ciampi**¹, **J. Henao-Mejia**^{6,7}, **A. Williams**^{8,9}, **R.A. Flavell**^{10,11}, **E. De Leonibus**^{1,4}, **E.M. Surace**^{1,2}, **S. Banfi**^{1,2}. 1) Telethon Institute of genetics and Medicine, Pozzuoli (Naples), Italy; 2) Department of Translational Medicine, Federico II University, Naples, Italy; 3) Medical Genetics, Department of Biochemistry, Biophysics and General Pathology, University of Campania "L. Vanvitelli", Italy; 4) Institute of Genetics and Biophysics (IGB), CNR, Naples, Italy; 5) Department of Neuroscience, Biomedicine and Movement Science, University of Verona, Italy; 6) Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA; 7) Institute for Immunology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA; 8) The Jackson Laboratory for Genomic Medicine, Farmington, Connecticut, USA; 9) Department of Genetics and Genomic Sciences, University of Connecticut Health Center, Farmington, Connecticut, USA; 10) Department of Immunobiology, Yale University School of Medicine, New Haven, Connecticut, USA; 11) Howard Hughes Medical Institute, Chevy Chase, Maryland, USA.

Mitochondrial dysfunction underlies the pathogenesis of a variety of human neurodegenerative disorders, either directly, e.g., in the case of mitochondrial diseases (MD), or indirectly, e.g. in the case of Parkinson's disease (PD). The complexity of these disorders has so far prevented the development of effective therapeutic strategies. Here, we demonstrate that the microRNAs miR-181a and miR-181b (miR-181a/b) regulate key genes involved in mitochondrial biogenesis and function. We also show that these miRNAs are involved in the global regulation of mitochondrial turnover in the central nervous system through the simultaneous and fine tuning modulation of different gene pathways. We found that miR-181a/b downregulation strongly protects neurons from cell death and significantly ameliorates the disease phenotype in *in vivo* models of both primary and secondary mitochondria-mediated neurodegeneration, such as Microphthalmia with Linear Skin Lesions, Leber hereditary optic neuropathy and PD. Altogether our results indicate that miR-181a/b represent novel gene-independent therapeutic targets for a wide-range of neurodegenerative diseases caused by mitochondrial dysfunction.

1203W

Characterizing mtDNA variants from 10,000 germline exomes. **A. Bolze**, **J. Lu**, **F. Mendez**, **Y. Ni**, **N.L. Washington**. Helix, San Carlos, CA.

Pathogenic variants in mitochondrial DNA (mtDNA) have been reported to cause rare diseases such as Leber's hereditary optic neuropathy (LHON), tubulointerstitial kidney disease or mitochondrial encephalomyopathy, lactic acidosis and stroke-like episodes (MELAS). Deciphering the genetic etiology of rare diseases has benefited from the release of large databases reporting allele frequencies for variants observed in 10,000s-100,000s of individuals such as Bravo (<https://bravo.sph.umich.edu/freeze5/hg38/>) and gnomAD (gnomad.broadinstitute.org). These tools can be used to rule variants out as the cause of a disease on the basis of their allele frequency in the general population, taking into account other parameters such as prevalence, penetrance, inheritance and genetic homogeneity. However, these two resources did not yet have information for mtDNA variants. In order to support rare mitochondrial disease studies, we called and analyzed mtDNA variants based on the germline exomes of a cohort of 10K individuals. The mtDNA variants reported supplement broader mitochondrial resources such as Hmtvar (<http://www.hmtvar.uniba.it/>) and Mitomap (<https://www.mitomap.org/MITOMAP>). We compared the allele frequencies of the variants in our cohort with the allele frequencies in the two mitochondrial resources and characterized the mtDNA variants impacting mitochondrial genes, tRNAs and rRNAs.

1204T

Multi-omics analyses and functional studies in a new model of Barth syndrome and in patient derived cells uncover novel mechanisms of pathogenesis and potential targets for therapeutic intervention. A. Franca Anzmann¹, R.N. Cole², B.J. Kirsch³, A. Le³, F.M. Vaz⁴, S.M. Claypool⁵, H.J. Vernon¹. 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD., Select a Country; 2) Proteomic Core for the Hopkins Conte Digestive Diseases Basic and Translational Research Core Center Johns Hopkins University School of Medicine Baltimore, MD; 3) Department of Pathology and Oncology Johns Hopkins University School of Medicine Baltimore, MD; 4) Laboratory Genetic Metabolic Diseases Academic Medical Center Amsterdam, The Netherlands; 5) Department of Physiology Johns Hopkins University School of Medicine Baltimore, MD.

Barth Syndrome (BTHS) is an X-linked inborn error of metabolism of cardiolipin (CL) synthesis caused by genetic defects in *TAZ* on chromosome Xq28. Clinical manifestations of BTHS include childhood-onset cardiomyopathy, neutropenia, growth defects, and skeletal muscle myopathy. Despite knowledge of the primary metabolic defect in this disorder, there is limited knowledge of downstream mechanisms of cellular pathogenesis and targets for treatment are limited. We developed a CRISPR edited *TAZ* deficient model in HEK293 cells, *TAZ*^{Δ45}, and performed semi-targeted lipidomics, shotgun proteomics and targeted metabolomics analyses in the *TAZ*^{Δ45} and *TAZ* WT cells. Through these studies we identified pathways representing major bioenergetic shifts in both mitochondrial and cytoplasmic metabolism. Lipidomics studies showed that the *TAZ*^{Δ45} cells recapitulate the cardiolipin dysregulation seen in BTHS with a decrease in mature CL, an increase in monolysophosphatidylcholine (lysophosphatidylcholine), and an altered content of CL biosynthetic intermediates. DAVID and KEGG analyses, and analysis of regulatory components of differentially expressed proteins, revealed major pathways of dysregulation in *TAZ*^{Δ45} cells including (a) downregulation of mitochondrial respiratory chain complex I (b) alterations in AMPK signaling and fat metabolism pathways and (c) alterations in insulin signaling pathways. Western blotting confirmed key shotgun proteomics findings. Metabolomics analyses intersected with the proteomics studies, showing an increase in the NADH/NAD⁺ ratio, increase in cellular AMP content, and altered glycolytic intermediates. Targeted studies in 5 individual BTHS patient derived lymphoblastoid cell lines with different *TAZ* genotypes recapitulated core proteomic findings found in *TAZ*^{Δ45}. With these studies we have (a) validated the biological relevance of a novel CRISPR edited *TAZ* cellular model (b) identified mitochondrial complex 1 as a major site of bioenergetic dysregulation in BTHS and (c) identified AMPK signaling, insulin signaling, and fat metabolism as areas of major cytoplasmic bioenergetic dysregulation in BTHS. Clinical correlation of these pathways in individuals with BTHS is underway, with the goal of developing associated endpoints for future targeted therapeutic trials.

1205F

Deletion in mitochondrial peptidase results in lethal juvenile brain disorder with status epilepticus in dogs. M.K. Hytonen^{1,2,3}, R. Sarviaho^{1,2,3}, C. Jackson⁴, P. Syrjä³, K. Matisek⁴, T. Jokinen⁵, M. Arumilli^{1,2,3}, M. Anttila⁶, L.A. Bindoff^{7,8}, A. Suomalainen¹, H. Lohi^{1,2,3}. 1) Research Program of Molecular Neurology, University of Helsinki, Helsinki, Finland; 2) The Folkhälsan Institute of Genetics, Helsinki, Finland; 3) Dept Veterinary Biosciences, University of Helsinki, Finland; 4) Section of Clin Comp Neuropathol, LMU Munich, Munich, Germany; 5) Dept Equine Small Animal Medicine, University of Helsinki, Helsinki, Finland; 6) Finnish Food Safety Authority, Evira, Helsinki, Finland; 7) Dept Neurology, Haukeland University Hospital, Bergen, Norway; 8) Dept Clinical Medicine, University of Bergen, Bergen, Norway.

Dogs have emerged as an excellent comparative model for human medicine due to shared genes and because they phenocopy human physiology and dysfunction. Canine complementary models help uncovering new candidate genes, improving the understanding of the molecular mechanisms and opening possibilities for therapeutic trials. In this study, we have investigated an unknown rapidly progressing and lethal canine brain disorder with epileptic seizures beginning at 6 to 12 weeks of age. Detailed clinical investigations and post-mortem examination were carried out and tissues collected for pathology and functional studies. DNA samples were obtained from 7 affected dogs and ~400 unaffected dogs by owner's consent and with ethical permission. A combined approach of homozygosity mapping and exome sequencing was performed to identify the genetic defect. Candidate variants were validated by Sanger sequencing and TaqMan assay. We discovered a novel rapidly progressing recessive juvenile mitochondrial encephalopathy resulting in lethal status epilepticus. Pathology showed a diffuse necrotizing panpolioencephalopathy with extensive neuronal mitochondrial crowding in two of the affected dogs available for detailed necropsy. Genetic analyses revealed an in-frame 6-bp deletion in a gene encoding for mitochondrial matrix peptidase. Mitochondrial function was assessed by respirometry in the affected and unaffected brain tissues, which demonstrated a significant decrease in respiration. In conclusion, we have characterized a novel canine brain disorder, juvenile mitochondrial encephalopathy with status epilepticus, and identified the genetic cause in a gene coding for a mitochondrial matrix peptidase, that degrades mitochondrial targeting sequences cleaved from imported proteins. Mutations in this gene have been found in one family with a slowly progressing neurodegeneration in humans. In contrast to the human disease, the canine condition is rapidly progressing after birth leading to status epilepticus and death. Further functional and pathological studies aim to uncover pathophysiological details of the canine disease and possible as yet unknown functions of the gene.

1206W

Atypical mitochondrial Leigh syndrome associated with a novel *MT-ATP6* gene variation. E. Rai¹, A. Angural¹, I. Sharma¹, P. Pandoh², V. Sharma¹, A. Spolia¹, V. Singh¹, S. Razdan³, K.K. Pandita⁴, S. Sharma¹. 1) Human Genetics Research Group, School of Biotechnology, Shri Mata Vaishno Devi University, Katra, Jammu and Kashmir, India, 182320; 2) Acharya Shri Chander College of Medical Sciences and Hospital, Sidra, Jammu, Jammu and Kashmir, India, 180017; 3) Neurology Clinic, 7 Bhagwati Nagar, Jammu, Jammu and Kashmir, India, 180001; 4) Health Clinic, H. No. 62, Lane 11, Swam Vihar, Muthi, Jammu, Jammu and Kashmir, India, 181205.

Leigh Syndrome is a rare, hereditary progressive neurodegenerative disorder of infancy or early childhood that affects all the levels of neuroaxis in affected individuals. Owing to its clinical and genetic heterogeneity, the disorder is mostly misdiagnosed, unfortunately leading to no or delayed optimal medical intervention to the patients. However, through a combinatorial approach including use of advanced clinical diagnostic and advanced molecular biology techniques, identification and characterization of the underpinning molecular etiology of Leigh syndrome has become a possibility. Here, we present a case report on a 9 year old boy having a severe, atypical clinical presentation of Leigh Syndrome- late survival age and bilateral basal ganglia calcifications (in the caudate nuclei)- validated through their genetic screening. Whole Mitochondrial Genome Sequencing of the recruited patient and his mother at different read coverage (first at 100X and later repeated at 500X for validation) revealed a novel, disease associated *MT-ATP6* variation, but with different allele counts between them. A higher allele count in the patient than his mother was concluded as disease-causing. Further, an exhaustive study is required to evaluate the clinical association of bilateral basal ganglia calcification to mitochondrial form of Leigh Syndrome. In conclusion, the present study indicates significance of a combinatorial diagnostic strategy including use of neuro-imaging and Next Generation Sequencing technologies for proper diagnosis of atypical cases of Leigh Syndrome.

1207T

Non-coding variants in *MECR*: Case report and molecular phenotype.

J.N. Kohler¹, D.E. Bonner¹, D.B. Zastrow¹, J. Davidson¹, L. Fernandez¹, C. McCormack¹, M. Majcherska¹, S. Marwaha¹, C. Curnin¹, M.R.Z. Ruzhnikov^{1,2,3}, G.M. Enns^{1,3}, L. Fresard⁴, K.S. Smith⁴, E. Worthey⁵, U.D.N. Undiagnosed Diseases Network⁶, S.B. Montgomery^{4,7}, P. Fisher^{1,2}, E.A. Ashley^{1,7,8}, J.A. Bernstein^{1,3}, M.T. Wheeler^{1,7,8}. 1) Stanford Center for Undiagnosed Diseases, Stanford, CA; 2) Department of Neurology, School of Medicine, Stanford, CA; 3) Department of Pediatrics – Medical Genetics, School of Medicine, Stanford, CA; 4) Department of Pathology, School of Medicine, Stanford, CA; 5) HudsonAlpha Institute for Biotechnology, Huntsville, AL; 6) NIH Undiagnosed Diseases Network, Office of the Director and the National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 7) Department of Genetics, School of Medicine, Stanford, CA; 8) Division of Cardiovascular Medicine, School of Medicine, Stanford, CA.

The mitochondrial fatty acid synthesis (mtFASII) pathway is an organelle-specific pathway for fatty acid synthesis with roots in the mitochondrion's bacterial evolutionary lineage. The mtFASII system remains conserved in mammals despite their independent, eukaryotic cytosolic system for fatty acid synthesis (FASI). FASI operates through a single enzyme whereas mtFASII uses a series of sequential enzymes. The final enzyme in the mtFASII pathway is encoded by *MECR*. Two brothers, ages 4 and 5 years, were evaluated at the Stanford Center for Undiagnosed Diseases (CUD) for a neurodegenerative disease of unknown etiology. They each presented at 6 months with delayed motor milestones and hypotonia, which evolved to include spasticity, an ataxic gait, and progressive loss of motor skills. Additional features include dystonia with hyperkinetic movements (elder brother) as well as ptosis and neck muscle weakness (younger brother). Both have relatively intact receptive language skills and cognition but minimal expressive language. Brain MRIs show bilateral symmetric T2 hyperintensity in the basal ganglia. Extensive biochemical and genetic evaluations, including exome sequencing, were unrevealing. Genome sequencing (GS) sent by the CUD identified two variants in *trans* in *MECR*, present in both siblings: c.830+2dupT and c.-39G>C. Pathogenic variants in *MECR* are associated with mitochondrial enoyl CoA reductase protein-associated neurodegeneration (MEPAN). MEPAN is characterized by (1) childhood-onset movement disorder, (2) signal hyperintensity in the basal ganglia, (3) optic atrophy, and (4) relatively preserved cognition. To date only seven individuals with MEPAN have been reported in the literature. While the c.830+2dupT variant has been described previously, the c.-39G>C variant is novel. It lies in the 5'UTR, a region typically known to influence gene expression levels. Functional studies exploring the effect of the c.-39G>C variant, including RNAseq and untargeted metabolomics, are ongoing. Given *MECR*'s established role in lipoic acid synthesis, the boys began a trial of alpha lipoic acid and MCT oil supplement to their mitochondrial cocktail. This case contributes to the emerging clinical and molecular characterization of MEPAN and expands the scope of potentially disease-causing variants. Additional molecular phenotyping will enhance understanding of the mtFASII pathway and can ultimately inform development of therapeutic drugs for MEPAN patients.

1208F

Preclinical efficacy of systemic messenger RNA (mRNA) therapy as a treatment for propionic acidemia. L. Jiang, J. Park, L. Yin, A. Frassetto, J. Zhuo, X. Yan, X. Zhu, K. Burke, K. Benenato, C. Mihai, A. Graham, K. Hoar, C. Lukacs, L. Guey, P. Martini. Moderna Therapeutics, Cambridge, MA.

Background: Propionic acidemia (PA) is a rare, inherited metabolic disorder with significant morbidity and mortality. The vast majority of PA patients present in the neonatal period with severe metabolic decompensations. Longer-term sequelae include neurologic complications and cardiomyopathy. PA is caused by a deficiency in propionyl CoA carboxylase (PCC), a mitochondrial enzyme composed of 6 alpha (PCCA) and 6 beta (PCCB) subunits. Lack of PCC activity due to mutations in PCCA (PA type I) or PCCB (PA type II) genes results in accumulation of toxic metabolites in all body fluids and tissues. There is no approved therapy for PA and the standard of care is restricted to dietary and other supportive measures. Liver transplant has emerged as an effective yet radical therapeutic option for severely affected patients to normalize PCC activity in the liver, and has been shown to reduce the risk of metabolic decompensations and in some patients, halt the progression of cardiomyopathy. However, liver transplant has significant limitations include the risks of the procedure itself. **Results:** Here we report a novel therapy for PA using a combination of two mRNAs encoding PCCA and PCCB to restore enzymatically active PCC in liver. In vitro studies demonstrated that co-transfection of PCCA and PCCB mRNAs increased PCC activity and encoded protein with the proper subcellular localization in the mitochondria in patient fibroblasts. Single and repeat IV administration of PCCA and PCCB mRNA encapsulated in biodegradable lipid nanoparticles in PA (PCCA hypomorphic) mice resulted in restoration of hepatic PCC activity with a concomitant decrease in plasma levels of disease-associated metabolites. The decrease in disease-associated metabolites was rapid and significant (>50%), persisted for several weeks after a single dose, and was sustained after multiple doses. **Conclusions:** These studies demonstrate preclinical proof-of-concept of systemic PCCA and PCCB mRNA therapy as a treatment for PA..

1209W

A splice mutation in *ATP11A* causes progressive sensorineural hearing loss and maps to the *DFNA33* locus. J.A. Pater, C. Penney, A. Griffin, N. Abdelfatah, J. Houston, D. O'Rielly, S.S. Stanton, J. Squires, T.L. Young. Craig L. Dobbin Genetics Research Centre, Faculty of Medicine, Discipline of Genetics, Memorial University of Newfoundland, St. John's, Newfoundland and Labrador, Canada.

Although more than 200 hearing loss loci have been mapped, dominant loci are particularly difficult to solve due to the rarity of large extended families, broad critical regions and the large number of heterozygous variants in individual genomes. As part of an ongoing study, we recruited a six-generation family from the genetic isolate of Newfoundland (Canada) with sixteen members affected by bilateral sensorineural hearing loss. The rate of hearing threshold decline was variable with onset as early as one year of age, progressing to a profound loss between the third and sixth decades. All family members screened negative for the 23 deafness mutations that we have previously identified in this founder population. Genome-wide SNP genotyping and two-point linkage analysis on a core pedigree identified a 3.6 Mb linked region (LOD=4.77) that overlapped with *DFNA33* on 13q34. Whole genome sequencing and filtering of variants mapping to the 13q34 critical region identified several candidate variants that co-segregated with hearing loss. Extensive haplotype analysis on the extended family using polymorphic markers revealed two key recombination events that markedly reduced the critical region leaving only one variant in *ATP11A*, which encodes for an integral membrane ATPase. Absent from 340 ethnically-matched population controls, *ATP11A* c.*11G>A was a predicted splice variant. Whole RNA extracted from blood, followed by reverse transcription PCR, TA cloning and Sanger sequencing revealed that *ATP11A* c.*11G>A causes alternative splicing in three *ATP11A* isoforms, identifying the *DFNA33* gene. .

1210T

Unraveling the genomic and mutational complexity of syndromic and non-syndromic deafness. H. Azaiez¹, K. Booth^{1,2}, B. Crone¹, S. Ephraim¹, A. Black-Ziegelbein¹, R. Marini¹, K. Frees¹, A. Weaver¹, J. Hendon¹, E. Shearer³, C. Sloan-Heggen⁴, D. Kolbe¹, E. Renkes¹, D. Wang¹, C. Nishimura¹, T. Braun⁵, R. Smith^{1,2,3,6}. 1) Molecular Otolaryngology and Renal Research Laboratories, Department of Otolaryngology, University of Iowa, Iowa City, IA; 2) The Interdisciplinary Graduate Program in Molecular Medicine, Carver College of Medicine, University of Iowa, Iowa City, IA; 3) Department of Otolaryngology, University of Iowa, Iowa City, IA; 4) Medical Scientist Training program, Carver College of Medicine, University of Iowa, Iowa City, IA; 5) Center for Bioinformatics and Computational Biology, Departments of Electrical and Computer Engineering and Biomedical Engineering, University of Iowa College of Engineering, Iowa City, IA; 6) Iowa Institute of Human Genetics, University of Iowa, Iowa City, IA.

Advances in sequencing technologies have revolutionized the fields of genetics and genomics, offering the promise of precision medicine tailored to disease prevention and treatment based on an individual's genomic data. Integral to precision medicine is the interpretation of variant effect in the context of clinical phenotypes. Such task represents a major challenge because of the extraordinary number of genetic variants and the complexities inherent in determining their clinical significance, especially for disorders exhibiting great phenotypic, genetic and allelic heterogeneity. To address this challenge for hearing loss, we have developed the Deafness Variation Database, a comprehensive, configurable, open-access bioinformatics pipeline that integrates all available genetic, genomic and clinical data together with expert curation to generate a single evidence-based classification for each variant in 152 genes implicated in syndromic and non-syndromic deafness. Of the 876,139 variants collated, we classify over 8100 as pathogenic or likely pathogenic, 172,257 as benign or likely benign and 695,709 as variants of uncertain significance; 1270 variants are re-categorized based on expert curation and for 24% of those; the change is of medical significance and impacts patient care. We also define the genomic landscape and mutational signatures of known deafness-causing genes and show that over 96% of coding variants are rare and novel and that pathogenicity is driven by minor allele frequency thresholds, variant effect and protein domain. Our findings provide an intricate and complex portrait of the genomic and mutational landscape of deafness genes and expose unrecognized nuances in specific genes which have major impact on tailoring variant classification guidelines in a disease- and gene-specific manner. This will greatly empower clinical decision-making and improve our understanding of deafness biology.

1211F

Usher syndrome diagnostics in the genomics era: Successes, challenges and opportunities. K.T. Booth^{1,2}, B. Crone¹, E. Renkes¹, B. Anderson¹, H. Azaiez^{1,3}, R.J. Smith^{1,3}. 1) Molecular Otolaryngology Renal Research Laboratories, Department of Otolaryngology-Head Neck Surgery, University of Iowa, Iowa City, Iowa; 2) The Interdisciplinary Graduate Program in Molecular Medicine, Carver College of Medicine, University of Iowa, Iowa City, Iowa; 3) Co-corresponding author.

Accurate variant interpretation is at the heart of genomic medicine. Both missed and mis-diagnoses yield non-ventured personalized therapeutic and intervention opportunities for patients. In genes that exhibit pleiotropy, such as the genes associated with Usher syndrome (USH) which could also cause non-syndromic deafness, elucidating phenotypic outcome based on genetic data often alters treatment avenues and quality of life. We therefore sought to elucidate the genomic landscape of the Usher-causing genes to lay the foundation for more accurate genetic variant interpretation. We collated over 127,300 variants across nine genes (*MYO7A*, *USH1C*, *CDH23*, *PDCH15*, *USH1G*, *USH2A*, *ADGVR1*, *WHRN* and *CLRN1*) and annotated them with minor allele frequency (MAF), variant effect, conservation, deleteriousness predictions and variant classification obtained from the Deafness Variation Database. Pathogenic and likely pathogenic variants represented ~1% of all variants, whereas likely benign and benign accounted for ~21% and variants of unknown significance ~78.2%. Of the known pathogenic variants, ~88% are exonic and ~12% alter splicing. Missense variants are most common and comprise ~38% of reported pathogenic variants, followed by frameshift indels, nonsense, splice-altering and in-frame indels (~26.5%, ~19.7%, ~12%, ~2.4%, respectively). Start/stop loss, synonymous and UTR variants make up <1% each. Novel and rare variants (MAF <0.5%) constitute 81% of all variants and >95% of coding variants, showcasing the complexities of distinguishing disease-causing variants from variants of unknown significance. In conclusion, we provide for the first time a comprehensive overview of the genetic landscape of the Usher syndrome genes. We established informative gene- and phenotype-specific MAF cutoffs for pathogenicity and specific variant types which will ultimately enhance variant classification and improve genotype-phenotype associations, thereby improving genetic diagnostics and advancing precision medicine for individuals affected with this disorder.

1212W

Genetic heterogeneity of hearing loss in sporadic cases from Pakistan.

M. Ramzan, S. Naz, *Baylor-Hopkins Center for Mendelian Genomics*. University of the Punjab, Lahore, Punjab, Pakistan.

Hearing loss affects up to 8% of Pakistani population. Genetic and allelic heterogeneity is one of the key features of recessively inherited nonsyndromic familial deafness. Although sporadic individuals with hearing loss are common in Pakistan, a comprehensive evaluation of genes involved in their phenotype has not been reported. We ascertained 21 unrelated individuals affected with sensorineural, moderate to severe hearing loss who were all born to consanguineous parents. Whole-exome sequencing was performed on DNA obtained from these individuals. We examined homozygous exonic and splice-site variants from the exome data for which the allele frequencies were <0.01 in public databases including ExAC and gnomAD. Both known and new variants in genes previously implicated with hearing loss; *GJB2*, *SLC26A4*, *OTOF*, *RDX*, *ESRRB*, *BSND* and *CABP2*, were found responsible for the phenotype in 13 participants. The maximum number of individuals had mutations in *OTOF*, followed by those with *GJB2*, *BSND* and *SLC26A4* variants. While variations in all other genes are known to cause moderate to severe hearing loss, those affecting *ESRRB* have only been described associated with profound deafness. This finding further strengthens the involvement of modifiers which control the degree of hearing loss in humans. Variants predicted to be pathogenic in genes not known to cause disorders were found in two individuals. These may represent newly implicated genes for hearing loss and remain to be evaluated further. No homozygous variant in a known deafness gene or a predicted pathogenic variant in any other gene was found for six participants. This could be due to the involvement of compound heterozygous or de novo variants in the etiology of their phenotype. Our work reveals similar genetic heterogeneity of hearing loss in sporadic individuals with hearing loss as that observed in those affected with familial deafness. Funded by HEC 3288, Pakistan.

1213T

“Enhancing” the DFNB1 locus: Targeted sequencing and functional characterization of non-coding elements affecting *GJB2* expression. S. Rentas¹, S.E. Raible², D. Mceldrew¹, I.D. Krantz², N.B. Spinner¹, A.N. Abou Tayoun³. 1) Division of Genomic Diagnostics, Department of Pathology and Laboratory Medicine, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Division of Human Genetics, Department of Pediatrics, The Children's Hospital of Philadelphia, Philadelphia, PA; 3) Al Jalila Children's Specialty Hospital, Dubai, UAE.

Bi-allelic pathogenic variants in the DFNB1 locus cause 50% of all recessively inherited non-syndromic sensorineural hearing loss (SNHL). While genomic diagnostics for this disorder is highly successful in identification of pathogenic variants in the coding sequence, up to 30% of individuals present with only a single deleterious variant in *GJB2*, giving an inconclusive answer for their phenotype. Importantly, this high percentage of *GJB2* carriers is significantly greater than the coincidental carrier rate in the general population, indicating these individuals likely carry a second pathogenic variant that is not detected by conventional testing. Multiple large deletions upstream of *GJB2* have been shown to contribute to deafness by significantly reducing its expression, and recent work from our group mapped a 95 kb critical region that is shared across reported upstream deletions in individuals with SNHL. This noncoding region has a high degree of histone acetylation (H3K27Ac) indicating the presence of uncharacterized distal regulatory elements that control *GJB2* expression. We hypothesize variants within this 95 kb critical region will be found as the second hit in carriers of a single pathogenic variant in *GJB2* and clinical diagnosis of non-syndromic SNHL. To test for novel pathogenic variants in the 95 kb critical region, we designed and carried out a targeted capture of this region followed by next-generation sequencing in a cohort of SNHL individuals who are heterozygous for a pathogenic variant in *GJB2*. The variants identified were compared to 15,496 whole genomes, using data in the Genome Aggregation Database (gnomAD). To complement the sequencing analysis, we have undertaken an unbiased approach to screen for functional enhancer elements in the 95 kb critical region. For these studies, we are generating BAC-focused libraries containing this region of interest and performing STARR-seq (Self-Transcribed Active Regulatory Region Sequencing) in a model cell line that abundantly expresses *GJB2*. Altogether, the combination of approaches taken here provide novel insights into the uncharacterized enhancer elements that promote *GJB2* expression, and aim to improve diagnostic yield in hereditary hearing loss testing.

1214F

Genomic characterization of hereditary sensorineural hearing loss in the Louisiana Acadians and Mexican Mayans. A. Umrigar¹, D. Mercer¹, A. Hurley², M. Norman³, A. D'Angelo¹, C. Hicks¹, R. Hernandez¹, J. Leshner¹, F. Tsien¹. 1) Department of Genetics, Louisiana State University Health Sciences Center, New Orleans, LA; 2) Department of Communication Disorders, Louisiana State University Health Sciences Center, New Orleans, LA; 3) Louisiana Deaf Blind Project, Louisiana State University Health Sciences Center, New Orleans, LA; 4) Division of Genetics, Universidad Juárez Autónoma de Tabasco, Tabasco, Mexico.

In the United States, at least 1/1000 children are born with complete sensorineural hearing loss (SNHL) per year, with 50-80% of cases due to genetic mutations. Cases of congenital and pre-lingual SNHL are caused by mutations in over 100 known genes that contribute to the development, function, and maintenance of inner ear structures. Many developed nations have implemented successful newborn screening programs to identify infants with hearing loss and provide early intervention services. Although these programs are effective in the identification of hearing loss, they are not always able to determine the etiology. Next generation sequencing and diagnostic panels have become increasingly reliable and less expensive for clinical use. Despite these advances, the diagnosis of genetic SNHL still presents challenges for health care providers due to genetic heterogeneity and variable expressivity. In some cases, genetic hearing loss is associated with other clinical features that may not become apparent until later in life. Pursuing a genetic diagnosis is especially important for individuals who come from ethnic backgrounds with high rates of consanguinity. In order to further our understanding of genetic SNHL, our group is studying individuals from two founder populations, the Louisiana Acadians and the Mexican Mayans. Samples are screened for genetic abnormalities using cytogenetic analysis, fluorescence *in situ* hybridization (FISH), PCR, and targeted sequencing. When indicated, whole exome sequencing (WES) is employed followed by a filtration of variants based on phenotypic effect. Pathway analysis is then used to discover affected cellular processes. Preliminary data from patients and families recruited from rural Acadian Louisiana and remote Mexican Mayan communities indicate the presence of hereditary, congenital and pre-lingual syndromic and non-syndromic SNHL. Evaluation of more than 300 family members and patients demonstrate a higher incidence of rare and novel variants than in the general population. The majority of the families had more than one gene alteration detected, usually within the same individual. The genetic analysis of these founder populations will further our understanding of rare and/or novel intra-familial genetic mutations that affect these populations, providing caregivers and health professionals with the information necessary for care and education of deaf individuals and families in their communities.

1215W

Expansion of the phenotype associated with variants in the cilium intraflagellar transport RAB28. C. Jespersgaard¹, T.D. Hjortshøj¹, M. Fang², T. Ilginis³, M. Bertelsen³, N. Bech³, H. Jensen³, C.A.B. Hey¹, L.B. Møller¹, Z. Tümer¹, T. Rosenberg³, K. Brøndum-Nielsen¹, K. Grønskov¹. 1) Kennedy Center, Department of Clinical Genetics, Rigshospitalet, Copenhagen, Denmark; 2) BGI-Shenzhen, Shenzhen 518083, China; 3) Department of Ophthalmology, Copenhagen University Hospital, Rigshospitalet-Glostrup, Denmark.

Background: Retinal dystrophy encompasses a range of diagnoses and modes of inheritance, and over 250 genes have been associated. Retinitis pigmentosa (RP), a group of generalized retinal dystrophies, characterized by degeneration of rod and cone photoreceptors, has been associated with many causative genes. Some of the affected genes in RP encode proteins in the primary cilium, a sensory organelle present on most cell types. Pathogenic variants in these genes causes ciliopathy and several organs can be affected, such as heart, eye, kidney, the skeletal system and obesity. Many of the genes associated with ciliopathies are involved in the intraflagellar transport, a process for the assembly and maintenance of cilia. **Methods:** Two brothers from a consanguineous family were referred for genetic counselling. The brothers presented with RP, excessive myopia and polydactyly. The combination of RP and polydactyly led to a suspicion of Bardet Biedl syndrome although the clinical criteria were not fulfilled. Previous investigations included array analysis and sequencing of at the time known BBS and related ciliopathy genes (*BBS1-17*) without any causative variants detected. We performed sequencing of 125 genes previously associated with retinal dystrophies on the index patient. **Results:** The variants were evaluated and classified according to the ACMG guidelines. A homozygous variant in the index in *RAB28* was detected and verified by Sanger sequencing in both brothers. The c.55G>A (NM_004249.3) p.(Gly19Arg) variant, located in the GTPase domain, has not previously been described in individuals with RP. It is reported with very low frequency in the gnomAD database (1 in 30824 alleles). **Discussion:** We present two brothers from a consanguineous family homozygous for a missense variant in *RAB28*. Pathogenic variants in *RAB28* have been reported previously in patients with cone-rod dystrophy and although *RAB28* has been identified as a GTPase in the cilium formation no other phenotypes than cone-rod dystrophy have been associated with the gene. This resembles the phenotypic spectrum of another intraflagellar transport protein - IFT172. In *IFT172* the phenotypic spectrum spans from isolated RP to severe ciliopathy as in Mainzer-Saldino syndrome, where patients have skeletal, renal, hepatic, brain and retinal abnormalities. It has been speculated that modifying alleles may be involved and explain the phenotypic spectrum although it requires cohorts with large sample size.

1216T

Autosomal recessive non-syndromic keratoconus: Homozygous

frameshift variant in the candidate novel gene *GALNT14*. T. Froukh¹, A. Hawwari². 1) Philadelphia university, Amman, Jordan; 2) Sight and insight eye clinic, Amman, Jordan.

Keratoconus (KC) is usually bilateral, noninflammatory progressive corneal ectasia in which the cornea becomes progressively thin and conical. Despite the strong evidence of genetic contribution in KC, the etiology of KC is not understood in most cases. In this study, we used whole exome sequencing to identify the genetic cause of KC in two siblings to first cousin parents. We identified a homozygous frameshift variant (NM_001253826.1:c.60del;p.Leu21Cysfs*6) in the gene N-acetylgalactosaminyltransferase 14 (*GALNT14*). We have prioritized this variant as a strong candidate causative to KC for two reasons. Firstly, the variant does not exist in the public databases: EXAC, gnomAD, 1000 genome project, Exome variant server (EVS), and Greater Middle East (GME) Variome project and neither in our in-house exome database (~3000 exomes). Moreover, none of the public or in-house exome databases harbor homozygous loss of function variant in the candidate gene. Secondly, *GALNT14* catalyzes the initial reaction in O-linked oligosaccharide biosynthesis, the transfer of an N-acetyl-D- galactosamine residue to a serine or threonine residue on mucins including ocular mucins. As a result, 55% of the ocular mucin is O-linked glycan. It was proved that alteration of mucin O-glycosylation at the ocular surface is described with eye diseases. We have linked in this study an association between the truncated protein *GALNT14* and KC. This study is one of the very few studies that identifies and links homozygous loss of function mutations to KC.

1217F

Biallelic CPAMD8 variants associated with congenital glaucoma and

anterior segment dysgenesis. E. Souzeau¹, O.M. Siggs¹, D.A. Taranath¹, T. Zhou¹, A. Dubowsky², S. Javadiyan¹, A. Chappell¹, A. Narita³, J.E. Elder^{4,5}, J. Pater¹, J.B. Ruddle^{4,6,7}, J.E.H. Smith^{8,9,10}, L.S. Kearns^{5,7}, S.E. Staffier^{4,6,7}, A.W. Hewitt^{6,7,12}, D.A. Mackey^{11,12}, K.P. Burdon^{1,12}, J.E. Craig¹. 1) Department of Ophthalmology, Flinders University, Flinders Medical Centre, Adelaide, Australia; 2) SA Pathology, Flinders Medical Centre, Adelaide, Australia; 3) Geelong Eye Centre, Geelong, Australia; 4) Department of Ophthalmology, Royal Children's Hospital, Melbourne, Australia; 5) Department of Paediatrics, University of Melbourne, Melbourne, Australia; 6) Centre for Eye Research Australia, Royal Victorian Eye and Ear Hospital, Melbourne, Australia; 7) Department of Ophthalmology, University of Melbourne, Melbourne, Australia; 8) Department of Ophthalmology, Children's Hospital at Westmead, Sydney, Australia; 9) Discipline of Ophthalmology, University of Sydney, Sydney, Australia; 10) Department of Ophthalmology, Macquarie University, Sydney, Australia; 11) Centre for Ophthalmology and Visual Science, University of Western Australia, Lions Eye Institute, Perth, Australia; 12) Menzies Institute for Medical Research, University of Tasmania, Hobart, Australia.

Purpose: Childhood glaucoma is a significant cause of irreversible blindness in children. The condition is clinically and genetically heterogeneous, with both primary and secondary causes. Anterior segment dysgeneses are associated with developmental defects of the ocular anterior chamber and can cause secondary glaucoma by impairing drainage of the aqueous humor, thereby increasing intraocular pressure. Biallelic pathogenic variants in the *CPAMD8* gene have previously been associated with anterior segment dysgenesis, myopia and ectopia lentis but not with childhood glaucoma. **Methods & Results:** We performed whole genome sequencing on an individual with a suspected diagnosis of primary congenital glaucoma and his unaffected parents, and identified biallelic predicted deleterious variants in *CPAMD8*. We then performed whole exome sequencing on 84 additional probands with a suspected diagnosis of primary congenital glaucoma recruited in the Australian and New Zealand Registry of Advanced Glaucoma, from which we identified biallelic variants in an additional two families. In total, we have identified five novel predicted deleterious variants in *CPAMD8* in four individuals from three families. *CPAMD8* variants explained 3.6% (3/84) of cases in our cohort. Common ocular features included corectopia, iris stromal hypoplasia, iris transillumination defects, iridodonesis, ectopia lentis and myopia. Cataracts were present in all individuals above 20 years of age. All four individuals required surgery to control intraocular pressure and glaucoma. *CPAMD8* expression was highest in tissue of neuroectodermal origin (including the ciliary body and corneal epithelium of adult anterior segment), suggesting a role for *CPAMD8* in drainage structures of the eye. **Conclusion:** We extend the phenotypic spectrum associated with *CPAMD8* variants to anterior segment dysgenesis with congenital glaucoma, adding *CPAMD8* to the list of childhood glaucoma-associated genes.

1218W

The *PMEL* gene and merle in the domestic dog: A continuum of insertion lengths leads to a spectrum of coat color variations in Australian shepherds and related breeds. B.C. Ballif, C.J. Ramirez, C.R. Carl, K. Sundin, M. Krug, A. Zahand, L.G. Shaffer, H. Flores-Smith. Paw Print Genetics, Genetic Veterinary Sciences, Inc.

Merle is a distinct coat color and pattern characterized by patches of diluted eumelanin (black pigment) interspersed among areas of normal pigmentation. In dogs, this variegated pattern is caused by an insertion of a SINE element into the canine *PMEL* gene. Although variation in the length of the SINE insertion (due to a variable-length poly(A)-tail) has been observed to be associated with variation in merle coat color and patterning, no systematic evaluation of the correlation between the length of merle insertion variants and coat color and pattern has been conducted and published in the scientific literature. We performed high resolution analysis of the SINE insertion lengths in 175 dogs (99 Australian shepherds, 45 miniature Australian shepherds, and 31 miniature American shepherds) and compared the genotypes with the coat phenotypes. Lengths varied from 201 – 277 base pairs (bp) indicating that variants can occur in virtually any size along the entire continuum. Genotype-phenotype correlation of 126 dogs with only a single SINE insertion (m/M) identified at least four major phenotypic clusters designated as 'cryptic', 'atypical', 'classic' and 'harlequin' merle. However, we identified several phenotypic outliers that did not cluster within these major groupings suggesting that insertion size is not the only factor responsible for merle phenotypic variability. In addition, we identified 25 dogs with two SINE insertions (M/M) and 24 dogs with more than two *PMEL* (merle) alleles, indicating mosaicism. Genotype-phenotype correlation of M/M dogs indicate that cryptic merle alleles often act like non-merle (m) alleles when combined with atypical, classic and harlequin sized alleles. The finding of mosaicism has important implications for the dog's phenotype and the ability to potentially transmit various alleles to its offspring. Furthermore, we identified examples of the SINE insertion poly(A)-tail expansion and contraction between generations, which also has important implications for breeding practices and determining mating pairs to avoid producing 'double merle' dogs that often have large patches of white fur and auditory and ophthalmologic defects. These findings demonstrate that there is a continuum of merle insertion lengths associated with a spectrum of coat color and patterns and that genotype-phenotype exceptions and overlap make it difficult to strictly assign insertion sizes with an expected coat color, although some generalizations are possible.

1219T

Biallelic loss-of-function variants in *RAX2*, encoding a homeobox-containing Rax transcription factor, cause autosomal recessive inherited retinal disease. S. Van de Sompele¹, C. Smith², M. Karal^{3,4}, M. Corton^{5,6}, K. Van Schil¹, F. Peelman⁷, T. Cherry⁸, T. Rosseel¹, H. Verdin¹, J. Derolez¹, T. Van Laethem¹, K.N. Khan⁹, M. McKibbin⁹, C. Toomes³, M. Al¹⁰, A. Torella³, F. Testa¹⁰, F. Simonelli¹⁰, J. De Zaeytijd¹¹, J. Van den Ende¹², B.P. Leroy^{11,13}, F. Coppie¹⁴, C. Ayuso^{3,6}, C.F. Inglehearn², S. Banfi^{3,4}, E. De Baere¹. 1) Center for Medical Genetics, Ghent University and Ghent University Hospital, 9000 Ghent, Belgium; 2) Section of Ophthalmology and Neuroscience, Leeds Institute of Biomedical and Clinical Sciences, St James's University Hospital, LS9 7TF Leeds, UK; 3) Medical Genetics, Department of Precision Medicine, Università degli Studi della Campania "Luigi Vanvitelli", 80138 Naples, Italy; 4) Telethon Institute of Genetics and Medicine, 80078 Pozzuoli, Italy; 5) Genetics Department, Instituto de Investigación Sanitaria-Fundación Jiménez Díaz University Hospital (IIS-FJD, UAM), 28040 Madrid, Spain; 6) Center of Biomedical Network Research on Rare Diseases (CIBERER), ISCIII, 28040 Madrid, Spain; 7) Flanders Institute for Biotechnology (VIB), Department of Medical Protein Research, Faculty of Medicine and Health Sciences, Ghent University, 9000 Ghent, Belgium; 8) Center for Developmental Biology and Regenerative Medicine, Seattle Children's Research Institute, 98101 Seattle, USA; 9) Department of Ophthalmology, St. James's University Hospital, LS9 7TF Leeds, UK; 10) Eye Clinic, Multidisciplinary Department of Medical, Surgical and Dental Sciences, Università degli Studi della Campania "Luigi Vanvitelli", 80131 Naples, Italy; 11) Department of Ophthalmology, Ghent University and Ghent University Hospital, 9000 Ghent, Belgium; 12) Center for Medical Genetics, Antwerp University Hospital, 2650 Antwerp, Belgium; 13) Division of Ophthalmology, The Children's Hospital of Philadelphia, 19104 Philadelphia, USA.

RAX2 encodes a member of the homeobox-containing Rax family of transcription factors, which play a pivotal role in late retinogenesis in vertebrate species by regulating the spatial expression of photoreceptor-specific genes. So far, only four monoallelic *RAX2* variants have been tentatively implicated in autosomal dominant cone-dominated retinal disease. Here we report biallelic missense, frameshift and structural variants in *RAX2* identified by whole exome sequencing, in five unrelated index cases of Belgian, British, Italian and Spanish origin, diagnosed with non-syndromic autosomal recessive retinitis pigmentosa (RP) with a variable age of onset. Protein structure analysis of the two novel missense variants revealed a loss of Rax2 protein folding and/or stability, in agreement with a loss-of-function effect. Modeling of the previously reported dominant *RAX2* missense variant on the other hand demonstrated potential roles in homeodomain/DNA higher order complex formation with no effect on protein stability or DNA binding, thus compatible with a gain-of-function effect. Haplotype analysis in three Belgian RP cases sharing the same *RAX2* frameshift variant c.335dup suggested a common ancestry. Fine-mapping and bio-informatics analysis of the two identified structural variants affecting *RAX2* disentangled their underlying mechanisms. In summary, our findings support a role for *RAX2* as a novel disease gene for autosomal recessive RP. This study uncovered the first structural variants affecting *RAX2* and a founder allele in Belgian RP patients. The identification of biallelic pathogenic *RAX2* variants in five unrelated families may suggest a role in other autosomal recessive RP cases with an unknown molecular diagnosis.

1220F

Identification of genetic causes of paediatric cataracts, including novel cataract genes, achieved using whole genome sequencing and linkage analysis. J.L. Jones¹, B. McComish¹, M.A. Corbett², J. Charlesworth¹, J. Gecz³, J.E. Craig³, K.P. Burdon¹. 1) Menzies Institute for Medical Research, University of Tasmania, Hobart, Tasmania, Australia; 2) Adelaide Medical School, Robinson Research Institute, University of Adelaide, Adelaide, South Australia, Australia; 3) Department of Ophthalmology, Flinders University, Bedford Park, South Australia, Australia.

Whole genome sequencing is enhancing the ability to identify disease causing mutations in inherited paediatric cataracts (PC), a heterogeneous rare disease that causes visual impairment. We are investigating families from Australia's largest repository of PC DNA. Nine large families, with at least 3 affected individuals each, were selected for linkage analysis using genome-wide genotyping on Illumina OmniExpress SNP arrays. Parametric linkage analysis, using Merlin, was performed with a relevant disease model for each family. Whole genome sequencing of an affected individual from each family was performed on an Illumina X10 platform. The Churchill pipeline was used to align fastq files to hg19, and variant calling was performed using SAMtools for SNPs and Platypus for indels. Structural variants will also be considered. Variants within putative linkage regions are filtered for frequency (MAF <0.01) and predicted functionality (using tools such as CADD). Novel cataract genes were identified as candidates in the two largest families and are undergoing further investigation, in a zebrafish model. In family 66, the PC have been linked to a 6.8Mb region at Xq24 (LOD=2.53) and two additional regions at 1q42.2-1q43 and 3q26.31-3q26.32 (LOD=2.44). A whole genome sequence (Complete Genomics™) of the proband enabled the identification of a 127kb deletion at Xq24, which truncates the *PGRMC1* gene following exon 1. This variant was not detectable by prior exome or Sanger sequencing and highlights the utility of whole genome sequencing for identifying putative disease causing structural and non-coding variants. The *ERO1B* c.662C>T (p.Ala221Val) variant on chr1 segregates, but is not highly conserved and a mini-gene splicing assay was unable to detect altered splicing in human lens epithelial cells despite in silico predictions. In family 24, segregating variants have been identified in novel cataract genes *SH3KBP1* and *PHEX*, both in the Xp22.2-p22.11 linkage region (LOD=1.63). In the smaller families, linkage analysis has reduced the search area down to a manageable 1-12% of the whole genome and analysis is ongoing. This work has identified novel disease causing genes for PC and can provide these Australian families with a molecular diagnosis and enable genetic screening of family members. The discovery of novel genes will expand the suite of genes available for genetic testing in other patients and improve our understanding of cataract pathogenesis.

1221W

Two candidate genes for rare cases of autosomal recessive non-syndromic hearing impairment in the gene loci DFNB33 and DFNB81. R. Birkenhager, M. Trabandt, S. Arndt, A. Aschendorff, R. Laszig. Department of Otorhinolaryngology, Head and Neck Surgery, University Medical Center Freiburg, Killianstrasse 5, D-79106 Freiburg, Germany.

Hearing impairment is the most prevalent sensorineural disorder and also the most common birth defect. Approximately one of every 1000 newborns is born profoundly deaf and one of every 300 newborns has a permanent mild-to-profound congenital hearing impairment. More than half of pre-lingual hearing impairment cases occur due to genetic factors. Currently, over 120 genes and over 180 loci have been identified for non-syndromic hearing impairment. Autosomal recessive non-syndromic hearing loss (ARNSHL) is a genetically heterogeneous sensorineural disorder, with prelingual hearing loss and absence of other clinical manifestations. Based on the clinical diagnosis it is not possible to recognize in which gene mutations are present. This is only feasible in exceptional cases (7/180). The aim of this study is to identify the responsible genes in rare cases of German families with prelingual ARNSHL. Hearing testing BERA/Electrocochleography and radiological a high-resolution CT scan was carried out. Mutational analysis of two affected family members was performed using direct sequencing of the coding exon and intron transitions of the genes *GJB2* (MIM 121011) and *GJB6* (MIM 604418), including deletion analysis. For investigation of autosomal recessive nonsyndromic hearing loss genes, whole exome sequencing was performed, with the "INVIEW HUMAN EXOME" platform; array Agilent Genomics SureSelectXT All Exon V5. No mutations could be identified in the *DNFB1* gene locus, containing the genes *GJB2* and *GJB6*. A further targeted analysis of other genes was not possible; therefore complete exome sequencing took place. All known genes and loci for autosomal recessive and dominant hearing impairment were analyzed. Only in the genes *DPP7* (MIM 610537), located on Chromosome 9q34.3 (DFNB33), and the gene *MUC16* (MIM 606154), located on chromosome 19p (DFNB81) two heterozygous mutations were detected; in the gene *DPP7* two splice site mutations and in the gene *MUC16* missense mutations c.40666G>A, p.D13556N and c.40655C>T, p.T13552I.

1222T

Novel mutations of *SCN9A* gene in patient with congenital insensitivity to pain identified by whole genome sequencing. *G.B. Christensen, K. Day, S. Austin, M. Miller, H. Escobar, L. Nadauld.* Intermountain Precision Genomics, Intermountain Healthcare, St. George, UT.

Congenital Insensitivity to Pain (CIP) is a very rare condition distinguished by inability to perceive pain, beginning at birth. Biallelic loss-of-function mutations in *SCN9A* are a known cause of CIP. *SCN9A* encodes the voltage-gated sodium channel Nav1.7, which is essential to nociceptor function and pain sensation. Here we describe whole genome sequencing results for a Caucasian male child symptomatic of CIP. DNA was collected from the child and both parents and sequenced on the Illumina NovaSeq platform. Sequence alignment and joint variant calling was performed with Sentieon. Structural variation analysis and consensus calling was performed with the Parliament method on the DNAnexus platform. Golden Helix VarSeq software was used for variant annotation and analysis. Whole genome sequencing identified two potential loss-of-function variants in *SCN9A*: a *de novo* missense variant (NM_002977.3:c.4651T>C, NP_002968.1:p.Cys1551Arg) in exon 26; and a paternally inherited deletion resulting in the complete loss of exon 17 (NC_000002.11:g.167128623_167131901del, NP_002968.1:p.Val948_Val1106del). Both variants are heterozygous, and neither has been reported previously. Point mutations in exons 17 and 26 of *SCN9A* have previously been associated with CIP and other pain disorders. One previous study also associated CIP with the loss of exon 17 due to a splice donor mutation. Whole genome sequencing proved particularly valuable for this case by enabling clear identification of the intronic breakpoints of the exon 17 deletion. We are working to confirm the phase of the identified variants with long-range haplotyping technology. If confirmed as biallelic, we believe that these variants may explain the patient's phenotype.

1223F

A novel *CRYGD* gene mutation loci results in non-syndromic autosomal dominant congenital cataract and high myopia in a Chinese family. *R. Zheng, D. Luo, L. Jiang.* University of Electronic Science and Technology of China, Chengdu, China.

Background Almost one-third of congenital cataracts are autosomal dominant congenital cataract, resulting in blindness and opacity of the lens. Mutations in *CRYGD* have been reported in families of cataract patients, and this gene encodes the lens gamma D protein which is the main protein of the vertebrate eye lens and maintains the lens's transparency and refractive index. The purpose of this study was to identify the disease-causing mutation in a Chinese family affected by bilateral, autosomal dominant congenital cataract. **Methods** Detection of candidate gene mutations and linkage analysis of microsatellite markers were executed for known candidate genes. Using the candidate gene in all affected family members Molecular mapping and clone screening potential genetic mutations and mutations confirmed by a single enzyme digestion. **Results** The proband was diagnosed with congenital cataract and high Myopia, who have a novel mutation in *CRYGD* gene, c.104G>A(Ser35Leu), and variation from grandmother (grandmother's sisters also have the mutation), after by his father inherited Index (the patient). This mutation was not found in unaffected family members. Sequence analysis confirmed that the Ser35 amino acid residue is highly conserved. **Conclusions** The novel missense mutation c.104G>A of the *CRYGD* gene is the first report of a non-syndromic, autosomal dominant congenital cataract along with high myopia, thereby underline the important role of lens gamma D protein in the physiological and the refractive effect of the eye.

1224W

Punctiform and polychromatic pre-Descemet corneal dystrophy: Report of two families, clinical findings, and the identification of a segregating intronic variant using whole-exome sequencing. D.D. Chung¹, J. Alio del Barrio², A.B. Barrington¹, K. Jatavallabhula¹, V. Swamy¹, A.J. Aldave¹. 1) Ophthalmology, Stein Eye Institute, UCLA, Los Angeles, CA; 2) Visum Alicante, C/ Cabañal, N1 03016 Alicante, Spain.

Purpose: To report the clinical features and genetic basis of two previously unreported families with punctiform and polychromatic pre-Descemet corneal dystrophy (PPPCD). **Methods:** Slit lamp examination and diagnostic corneal imaging were performed for members of two families with PPPCD. DNA was collected from examined individuals of both families and whole exome sequencing (WES) was performed on 10 members of the first family. Alignment, variant calling and annotation were performed using commercially available software programs. Novel or rare variants that segregated with the affected status were validated by Sanger sequencing and were screened for in the family members who did not undergo WES and in the second family. Identified non-coding variants that segregated with the affected status in both families were characterized using in silico prediction tools. **Results:** Eight of the 13 members of the first family and three of the six members of the second family were diagnosed with PPPCD based on the presence of polychromatic opacities in the posterior corneal stroma of each eye. No novel or rare non-synonymous coding variants that segregated with the affected status in the first family were identified with WES. An intronic variant on Chromosome 10 segregated with the affected phenotype in both families with PPPCD. In silico prediction tools predicted the intronic variant to create a cryptic donor splice site. **Conclusions:** We report only the 9th and 10th families with PPPCD, demonstrating an autosomal dominant inheritance pattern. Whole exome sequencing led to the identification of an intronic variant in both PPPCD families that is predicted to create a cryptic donor splice site.

1225T

Disparities in discovery of pathogenic variants within autosomal recessive non-syndromic hearing impairment by ancestry. I. Chakchouk, D. Zhang, Z. Zhang, H. Dai, R. Lyn P. Santos-Cortez, I. Schrauwen, S. M. Leal. Molecular & Human Genetics, Baylor College of Medicine, Houston, TX.

Hearing impairment (HI) is characterized by an extensive genetic heterogeneity. Although the explosion in HI gene discovery accredited to technological advances has yielded 104 Non-Syndromic (NS) HI genes, among which 68 are involved in autosomal recessive (AR) NSHI, we demonstrate that for most populations the majority of genes and variants explaining ARNSHI etiology remain to be elucidated. Pathogenic and likely pathogenic ARNSHI variants were selected based on ClinVar, Deafness Variation Databases, and allele frequencies from gnomAD were used to determine the contribution of ARNSHI genes to HI etiology per population. The estimated amount of ARNSHI explained by known pathogenic or likely pathogenic variants varies greatly by population, e.g. 88.1% for Ashkenazi Jews but only 4.7% for Africans/African Americans. Even within European populations, there is variability, although known ARNSHI genes and variants explain 24.2% of ARNSHI in non-Finnish Europeans they only explain 8.7% in Finns. There was also variability amongst Asians with 15.5% of ARNSHI explained in East Asians, but twice as much for South Asians (30.6%). This variability may be explained by the number of studies and clinical sequencing performed within specific populations in addition to the unique genetic background of isolated populations, e.g. Finns and East Asians and the rarity of admixture with other populations. As expected *GJB2* has the highest contribution to ARNSHI in all populations represented in gnomAD, with the exception of *MYO15A* in South Asians. Furthermore, although *GJB2* c.35delG (p.G12fs) is very frequent amongst several populations, due to an ancient founder, other common *GJB2* variants were population-specific e.g. p.W24X in South Asians [minor allele frequency (MAF)=4.45x10⁻³] and p.L79fs in East Asian (MAF=6.15x10⁻³). We also reveal for several rare ARNSHI variants, that were previously observed in one population/family, to have higher frequencies in other ancestry groups, e.g. *USH1C* p.Q723X was previously reported in a single Chinese family, although in gnomAD it is prevalent in Africans/African Americans MAF=2.16x10⁻³. In conclusion, we illustrate the need for additional gene discovery for ARNSHI particularly for specific ancestries, e.g. individuals of African-descent. The knowledge on the population specific spectrum of ARNSHI variants which this study has uncovered helps to define the genetic etiology of ARNSHI and aids in genetic testing.

1226F

A pathogenic haplotype, common in Europeans, causes autosomal recessive albinism and uncovers missing heritability in OCA1. K.

Grønskov¹, C. Jespersgaard¹, G.H. Bruun², K. Brøndum-Nielsen¹, B.S. Andersen², T. Rosenberg³. 1) Kennedy Center, Rigshospitalet, Glostrup, Hovedstaden, Denmark; 2) Department of Biochemistry and Molecular Biology and the Villum Center for Bioanalytical Sciences, University of Southern Denmark, Odense, Denmark; 3) Department of Ophthalmology, Kennedy Center, Copenhagen University Hospital, Rigshospitalet, Glostrup, Denmark.

Oculocutaneous albinism (OCA) is a genetically heterogeneous disorder which affects the eyes, skin and hair. Six genes are associated with autosomal recessive OCA (*TYR*, *OCA2*, *TYRP1*, *SLC45A2*, *SLC24A5* and *LRMDA*), while *GPR143* is associated with X-linked ocular albinism (OA) which only affects the eyes. Molecular genetic analysis of the seven genes provides a genetic diagnosis in approximately 60% of individuals with a clinical diagnosis of OA/OCA. This leaves a considerable number of individuals with albinism without a genetic diagnosis; a large number of these are heterozygous for a causative sequence variation in *TYR*. The number is in excess of what could be explained by the carrier frequency. To identify missing causative sequence variants we used a NGS based approach, genotyping and segregation analysis. We report a putative pathogenic haplotype segregating in accordance with an autosomal recessive inheritance pattern. This haplotype includes two rare SNVs in non-coding regions (100 kb upstream of the *TYR* translation start site and in intron 3) and two common coding SNVs (rs1042602A (p.S192Y) and rs1126809A (p.R402Q)). The haplotype covering 143 SNPs in *TYR* and surrounding regions is designated: "GYGQ". It is found in 15% of affected individuals in our cohort either as compound heterozygotes with a pathogenic *TYR* sequence variant or as homozygotes. Individuals homozygous for the haplotype have a partial albinism phenotype. The haplotype was not detected among 32 unaffected family members who are carriers of a pathogenic *TYR* sequence variation (measured by the presence of one of the rare SNVs). An explanation for the pathogenicity of the haplotype could be the combination of p.S192Y and p.R402Q, although we cannot completely rule out that the two rare SNVs or a yet undetected mutation could be causative. However, the SNV in intron 3 was analysed by minigene analysis and showed no effect on splicing. The upstream SNV is located in a region not evolutionarily conserved and not known to possess enhancers. CNV analysis of the NGS data showed no pathogenic CNVs. In conclusion, the haplotype GYGQ can provide a molecular genetic diagnosis to a substantial number of individuals with albinism. Furthermore, as individuals homozygous for GYGQ have only a partial albinism phenotype, these could be clinically diagnosed as unexplained bilateral subnormal vision and/or nystagmus and should be analysed clinically and molecularly for albinism.

1227W

The mutation detections in two Chinese families with aniridia. M. Gu¹,

Y. He¹, H. Ying², X. Wei¹, L. Chen¹, W. Yang², C. Tao², Z. Wang^{1,3}. 1) Shanghai Jiao Tong University School of Medicine, Shanghai, China; 2) Department of Ophthalmology, Ren-ji Hospital at SJTUSM, Shanghai, China; 3) Research Center for Experimental Medicine, Rui-Jin Hospital at SJTUSM, Shanghai, China.

Aniridia (MIM 106210) is a kind of ocular disorders with a bilateral congenital absence or malformation of the iris. The disorder usually occurs in childhood with a prevalence of 1:40 000. Frequently associated ocular abnormalities include cataract and corneal abnormalities. About two-thirds of cases are familiar with dominant inheritance; the others are sporadic. Most cases are associated with mutations or deletions of the *PAX6* gene. However, aniridia is also a heterogeneous disorder with more than 5 disease-associated genes identified to date. These genes are *ELP4*, *TRIM44*, *FOXE3*, *FOXC1*, *PITX2*. We investigated two Chinese families with autosomal dominant aniridia. The study was approved by the committee of Shanghai Jiao Tong University School of Medicine. After informed consent was obtained, all family members who agreed to participate in this study were evaluated by experienced ophthalmologist. All affected members underwent interrogation and ophthalmological examinations. We found that five patients from family 1 were completely consistent with the clinical features of aniridia, and nine patients from family 2 had the phenotype of aniridia, ptosis and cataract. Screening for *PAX6* gene mutations was carried out in the 2 families by Sanger sequencing. The result showed that there was a known *PAX6* gene mutation C. 718C>T (p. Arg240) in Family 1, but no mutation of *PAX6* gene occurred in family 2. Considering the genetic heterogeneity of aniridia, we directly performed whole exome sequencing. We further selected eight highly suspicious genes for sequencing analysis. We finally identified a missense mutation [c.542C>T (p.Ser180Tyr)] in myosin heavy chain13 (*MYH13*, MIM 603487) at chromosome 17p13 in all affected individuals of family 2. Such genetic mutation was not detected in unaffected members in the family or in 300 unrelated, ethnically matched individuals (i.e., 600 chromosomes). *MYH13* was specifically expressed in the extraocular muscles for function in eye movement. We also checked *MYH13* sequence information in the database of the 1000 Genomes Project. No such variation (c.542C>T) in *MYH13* was found. These data suggest that the *MYH13* mutation is a disease-associated genetic variation cosegregating with the aniridia phenotype in this family. *In vitro* and *in vivo* studies on the molecular mechanism of *MYH13* gene mutations are underway. This work is partially supported by National Natural Science Foundation of China (Proj. No. 30470951, 31071107).

1228T

Novel genes implicated in rare congenital inner ear and cochleovestibular nerve malformations. I. Schrauwen^{1,2}, E. Kari³, J. Mattox⁴, L. Llac⁵, J. Smeeton⁶, I. Chakchouk⁷, M. Naymik⁶, D.W. Raible⁶, J.A. Knowles⁷, J.G. Crump⁷, S.M. Leal¹, M.J. Huettelman², R.A. Friedman³. 1) Center for Statistical Genetics, Molecular and Human Genetics Department, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030; 2) Neurogenomics Division and Center for Rare Childhood Disorders, Translational Genomics Research Institute, 445 N 5th str, Phoenix, AZ 85004; 3) Division of Otolaryngology, Head and Neck Surgery, Department of Surgery, University of California, San Diego, ECOB- East Campus Office Building Room 3-013, 9444 Medical Center Drive, Mail Code 7220, La Jolla, CA 92037; 4) Tina and Rick Caruso Department of Otolaryngology- Head and Neck Surgery, Keck University of Southern California School of Medicine, 1975 Zonal Ave., Los Angeles, CA 90033; 5) Department of Stem Cell Biology and Regenerative Medicine, University of Southern California Keck School of Medicine, 1975 Zonal Ave., Los Angeles, CA 90033; 6) Department of Biological Structure, University of Washington, 1959 NE Pacific Street, Seattle, WA 98195; 7) SUNY Downstate Medical Center, Department of Cell Biology - MSC 5, 450 Clarkson Avenue, BSB 2-5, Brooklyn, NY 11203.

Congenital inner ear malformations affecting both the osseous and membranous labyrinth can have a devastating impact on hearing and language development. With the exception of an enlarged vestibular aqueduct, non-syndromic inner ear malformations are rare, and their underlying molecular biology has thus far remained understudied. In the era of next generation sequencing and with the increasing sensitivity of imaging techniques, rare inner ear malformations are now more effective to identify and characterize. Each of these cases can provide a unique and invaluable insight into the development of the ear or the normal hearing process. In addition, abnormalities in cochleovestibular anatomy can provide challenges for cochlear implantation, and a genetic diagnosis and understanding of the molecular basis might help with identifying the correct therapeutic intervention for these unique cases. By using a family-based trio exome sequencing approach, we are studying children with rare malformations to identify molecular factors that might be important in the developing inner ear and/or cochleovestibular nerve. We previously identified compound heterozygous variants in *MASP1* in a child with absent cochlear nerves. Here, we will present the identification of *de novo* loss-of-function variants in *GREB1L* [c.4368G>T;p.(Glu1410fs) and c.982C>T;p.(Arg328*)] in two affected individuals with absent cochleae and 8th cranial nerve malformations. The cochlear aplasia in these affected individuals suggests that there is a developmental arrest or problem at a very early stage of inner ear development, e.g. during the otic pit formation. The crucial function of *Greb1l* in craniofacial development is also evidenced in knockout mice, which develop severe craniofacial abnormalities. In addition, we show that *Greb1l*^{-/-} zebrafish exhibit a loss of or abnormal sensory epithelia innervation. In conclusion, we demonstrate that *GREB1L* variants underlie rare congenital inner ear and/or cochleovestibular nerve malformations. We will present an overview of these results in addition to any newly identified novel candidates from our ongoing studies. The combination of imaging and the molecular basis of rare inner ear malformations can ultimately lead to a more effective individually tailored therapeutic and 'personalized medicine' approach for affected children. .

1229F

miR-204 overexpression exerts a protective role in inherited retinal diseases. S. Banfi^{1,2}, M. Karali^{1,2}, E. Marrocco¹, I. Guadagnino¹, M. Pizzo¹, A. Carissimo¹, I. Conte¹, E.M. Surace^{1,3}. 1) Telethon Institute of Genetics and Medicine (TIGEM), Pozzuoli (NA), Naples, Italy; 2) Medical Genetics, Department of Precision Medicine, University of Campania "L. Vanvitelli", Naples, Italy; 3) Department of Translational Medicine, University of Naples "Federico II", Naples, Italy.

Introduction: Inherited retinal diseases (IRDs) are progressive neurodegenerative conditions of the retina that represent a main cause of genetic blindness in the Western world. Their high genetic heterogeneity hinders the development of effective gene-based therapies. We have recently demonstrated that the microRNA miR-204 is essential for retinal function and plays a pathogenic role in IRD in humans. Therefore, we assessed the potential therapeutic action of this microRNA in IRDs. **Materials and Methods:** We delivered by subretinal injection an adeno-associated viral vector carrying the miR-204 precursor to the *Aip1*1knockout and the P347S-*RHO*transgenic mouse lines, models of autosomal recessive and dominant forms of IRDs, respectively. The impact on retinal function and degeneration was assessed by electroretinographic and immunohistological analyses. **Results:** We detected a notable improvement of the ERG response in the miR-204-injected eyes of P347S-*RHO*mice. This effect persisted for two months post-injection and was prevalent under photopic conditions that reflect cone activity. We also observed a preservation of cone photoreceptors and a significant decrease in apoptotic photoreceptor cells. Photoreceptors were better preserved also in miR204-injected eyes of *Aip1*1knockoutmice. Transcriptome analysis suggested that dampening of microglia activation represents one of the main mechanisms underlying the neuroprotective effect of miR-204. **Conclusions:** Our findings indicate that the subretinal delivery of miR-204 attenuates retinal degeneration in IRD mouse models and preserves retinal function, supporting the gene-independent therapeutic potential of this microRNA.

1230W

Lack of association between CTG repeats in the *DMPK* gene and audiological and vestibular findings. M. Arenas-Sordo, D. Gutiérrez-Tinajero, M. Trujillo-Bracamontes, B. Rivera-Mercado, A. Martínez-García Ramos, O. Hernández-Hernández, J. Magaña-Aguirre. Genetics, Instituto Nacional de Rehabilitación, Mexico City, Mexico City, Mexico.

Introduction. Myotonic dystrophy type 1 (DM1) is a neuromuscular disease, progressive and multisystemic, caused by an unstable mutation of CTG repeats in the *DMPK* gene locus 19q13.3. **Purpose.** The present work shows the results from audiovestibular findings and time of evolution from the disease, and their correlation with a CTG repeats in the *DMPK* gene. **Methods.** 58 patients with molecular diagnosis for DM1 and their controls were analyzed. The controls were matched exact in sex and age. All of them had a comprehensive audiological and vestibular evaluation. Audiograms classified the range, nature and degree of hearing loss and speech audiometry detected the sensorial or conductive nature of hearing loss. Tympanometry, acoustic reflex threshold, and transient otoacoustic emissions were part of the evaluation. The impact of the vestibular disease in daily life was evaluated by the Dizziness Handicap Inventory (DHI) test. Falling in patients with balance deficiency and vestibular disorders was measured by the dynamic gait index (DGI). We made a correlation index looking for if there is or not association between the repeats number and each one of the audiological and vestibular characteristics. **Results.** From the 58 patients, 64% were male and 36% women, 69% had familial background of the disease. Almost half of the patients (48%) had some degree of sensorial hearing loss, of them, with a conventional audiometry, 53% of the patients had some manifestation comparing with just 8.6% of the controls ($p=0.0001$); falls in high frequencies were also more important in patients than in controls ($p=0.001$). In High frequencies audiometry there were also differences ($p=0.00001$). The vestibular manifestations were present in 21 patients (36.2%), $p=0.00001$. Nistagmus was present in 2 patients, none of the controls, the DHI test, in its 3 scales showed that 53.4% had some degree of disability none of the controls ($p=0.00001$). The 3 correlations that show p statistically significant were DGI and DHI with CTG number repeats ($-0.3215/p=0.0313$ and $0.3143/p=0.0355$ respectively) and high frequency audiometry ($0.3422/p=0.0230$), with evolution time of the disease. **Conclusions.** The falls in high frequencies and superficial hearing loss were the findings more frequent. It was low correlation between CTG repeat number with ocular saccades test and in the DGI; with the time of evolution association was only with high frequency audiometry.

1231T

A patient with early-onset high myopia caused by a homozygous mutation in *FAM161A*. A. Bouman¹, M. van Tienhoven¹, N. Dzino², A.A.H.J. Thiadens², L.H. Hoefsloot¹, V.J.M. Verhoeven¹, J.A. Kievit¹. 1) Department of Clinical Genetics, Erasmus Medical Center, Rotterdam, the Netherlands; 2) Department of Ophthalmology, Erasmus Medical Center, Rotterdam, the Netherlands.

Myopia is the most common ocular disease occurring worldwide. In the last few years many new monogenetic causes of isolated early-onset high myopia have been discovered. Whole Exome Sequencing has proven to be a powerful tool in identifying underlying causes of non-syndromic high myopia. Recently, *FAM161A* has been identified as a new candidate gene for this condition. Mutations in this gene have previously been described in patients with retinitis pigmentosa (RP28, OMIM#606068). In this report we present an additional case of a young girl with early onset myopia (over -5 diopters in both eyes at the age of 4 years) without additional ocular or systemic features. No signs of RP were observed by routine ophthalmologic exams. A homozygous nonsense mutation was identified in *FAM161A* by Whole Exome Sequencing (gene-panel analysis). This particular mutation has been described in literature previously in a young patient with a similar phenotype. At this point, it is hard to speculate about the (potential) risk on developing RP in a later stage of life for this particular group of young patients. Additional ophthalmologic exams and follow-up will be performed for our patient. We conclude that mutational screening of *FAM161A* in young patients with early onset (high) myopia could be helpful in identifying its underlying cause. Additional patients with early onset high myopia caused by a mutation in *FAM161A* could be contributive regarding a future cohort description. Email: a.bouman@erasmumc.nl.

1232F

A heterozygous heterozygous variant p.S319G c.955 A>G of *DNMT3A* causing Tatton-Brown-Rahman in a mother and son. A. Tsai, K. Lenhart, A. Meher. peds/genetics, university of colorado , Aurora , CO.

Tatton-Brown-Rahman syndrome (TBRS) is a genetic overgrowth disorder characterized by tall stature, dysmorphic facial features and moderate intellectual disability. While *de novo* heterozygous variants in *DNMT3A* (2p23.3) are known to be causative for TBRS, recent cases of familial transmission have also been reported. We report a boy referred to genetic services due to severe anxiety, developmental delay, sleeping disorder and abnormal gait. Upon initial physical examination, he showed >98 percentiles on growth parameters, prominent glabella ridges, brushy eyebrows, hyperflexible joints, mild 2/3 syndactyly, pes planus, and mild skin acanthosis nigricans. Neurological examination also showed significantly poor balance, aggression and irritability. Initial genetic testing by chromosomal microarray showed the proband carried paternally-inherited gains in 2p22.1 and 3p26.3. As the father was asymptomatic, additional testing was warranted. Indeed, follow-up testing by an expanded trio ASD/ID NGS panel revealed a maternally-inherited variant of uncertain significance (VUS) in the *DNMT3A* gene (c.955A>G; p.Ser319Gly). Interestingly, the mother shared characteristic TBRS physical features with her son, and self-reported a learning disability, as well as a history of migraines, seizures, anxiety and depression. Family history was also significant for three siblings with learning disabilities, although *DNMT3A* status has not been confirmed. Although the c.955A>G; p.Ser319Gly VUS has not been previously reported, the TBRS-like phenotype seen in this family indicate a likely causal relationship. Therefore, we maintain clinical correlation and an in-depth dysmorphological examination remains a powerful tool in clinical diagnoses.

1233W

A novel neurogenetic syndrome with *RALA* mutation. N. Okamoto^{1,2}, Y. Shibukawa², A. Takata³, N. Miyake³, N. Matsumoto³. 1) Department of Medical Genetics, Osaka Women's and Children's Hospital, Izumi, Osaka, Japan; 2) Department of Molecular Medicine, Osaka Women's and Children's Hospital, Izumi, Osaka, Japan; 3) Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan.

[INTRODUCTION] RAL (RAS like proto-oncogene) signaling pathway is a RAS down-stream effector. RAL belongs to the RAS superfamily and shares a structural similarity with RAS. Human cells have two genes, *RALA* and *RALB*, both of which are associated with cancers. We report a girl with a novel pathogenic variant in *RALA*. She showed severe intellectual disability (ID) and autism spectrum disorder (ASD). She also shared some features with Noonan syndrome. **[CLINICAL REPORT]** The 8-year-old female was the first child of healthy and non-consanguineous Japanese parents. She was hypotonic and her developmental milestones were markedly delayed. Physical examination identified relative macrocephaly (+1.5SD) and dysmorphic features including frontal bossing, epicanthal folds, low-set ears, flat nasal bridge, and high arched palate. Her physical growth was within normal range. She spoke no meaningful words. She was diagnosed with severe ID and ASD. Her walking was unsteady. Routine laboratory investigations and brain MRI revealed no significant abnormalities. **[METHOD]** With the approval of our institutional ethics committee, the samples were analyzed using Whole Exome Sequencing. Precise methods have been reported by Takata et al. (Cell Rep. 2018). **[RESULTS]** The patient had a *de novo* variant in *RALA* (chr7:39726339,G>A(;hg19);:p.V25M). This variant was classified as a consensus-damaging missense *de novo* mutation in ASD cohort (Takata et al.). It was not found among public databases and in house data. **[DISCUSSION]** There have been no reports on the association of the *RALA* variant and genetic disorders. It has been reported that *RALA* is activated in various cancers. We suppose that germline mutation of *RALA* causes a novel Noonan-like neurogenetic syndrome.

1234T

A recognizable specific facial phenotype during adolescence and early adulthood in three unrelated patients with Dravet syndrome. *M.K. Thong¹, S. Chowdhury², Y.C. Ying³, M. Vaithalingam³, N.K. Batumalai⁴, Y. Yakob⁵, S.A. Abdul Wahab⁵, T.B. Khoo⁶.* 1) Department of Pediatrics, University of Malaya, Kuala Lumpur, Malaysia; 2) Newcastle University Malaysia; 3) Department of Pediatrics, Melaka-Manipal Medical College, Melaka, Malaysia; 4) Department of Pediatrics, Hospital Melaka, Malaysia; 5) Molecular Diagnostics and Protein Unit, Institute for Medical Research, Kuala Lumpur, Malaysia; 6) Department of Pediatrics, Institute of Pediatrics, Hospital Kuala Lumpur, Malaysia.

Background: Dravet syndrome (OMIM #607208) is an early-onset epileptic encephalopathy and has not been reported to be associated with dysmorphism or pubertal abnormalities. An earlier study showed no significant differences were detected using photogrammetric indices between children with Dravet syndrome and their unaffected siblings and the study concluded there was absence of specific facial phenotype in children with Dravet syndrome. **Methods:** We report on three unrelated Asian male patients, aged 15 to 27 years with Dravet syndrome confirmed by presence of pathogenic *SCN1A* variants, who presented with specific facial features that only occurred during adolescence and early adulthood. These patients were followed up longitudinally since early childhood. **Results:** The patients showed normal development during the first 2 years of life. Childhood photos showed normal facial features with no dysmorphism. During adolescence, they presented with lack of facial hair, sloping broad foreheads with overhanging supraorbital ridges and deep-set eyes and a tendency to keep an open mouth. These striking features were immediately recognizable as unique and similar in all the patients. **Conclusions:** These dysmorphism and the discordance in the Tanner Staging, the hunched posture with hip and knee flexion when walking, are supportive of an expanded Dravet syndrome phenotype only evident in adolescence and early adulthood.

1235F

A new *SMARCE1* variant in a patient with oculoauriculofrontonasal syndrome. *S. Yano¹, A. Fujimoto¹, N. Matsumoto², N. Miyake², J. Morin-Leisk², M. Gillespie³, H. Gao³.* 1) Genetics/Pediatrics, University of Southern California, Los Angeles, CA; 2) Human Genetics Yokohama City University, Yokohama, Japan; 3) Fulgent Genetics, Temple City, CA.

Introduction: Oculoauriculofrontonasal syndrome (OAFNS) is a rare disorder with features of both oculoauriculovertebral spectrum and frontonasal dysplasia. Carey and Yong first coined the term, OAFNS, in 1981 while Golabi et al. described the first case in 1983. The etiology of OAFNS is unknown although it is thought to be a result of the abnormal development of structures in the midface and in the first and the second branchial arches. A sporadic occurrence and autosomal dominant transmission are suggested based on the past reported cases of OAFNS. Herein, we report a case of OAFNS who was found to have a de novo heterozygous *SMARCE1* variant that is not previously reported in the literature. *SMARCE1* encodes the protein which is a part of the large ATP-dependent chromatin remodeling complex SWI/SNF. Human analogs of SWI/SNF complex are BAF complex. Heterozygous mutations in the BAF complex-related genes are reported in patients with Coffin-Siris syndrome. **Case Study:** Proband was born to a G2P1SAb1 19-year-old female and a non-consanguineous 32-year-old male. She was a full-term baby and was delivered by NSVD. At birth, she was diagnosed with multiple minor anomalies including right microtia and left low set ear as well as congenital heart defects (patent ductus arteriosus (PDA) and bicuspid aortic valve). At age 3 years, she was noted to have failure to thrive, microcephaly, hypotonia, bilateral hearing loss, hypertelorism, broad bifid nose, depressed nasal bridge, widely spaced maxillary central incisors, right ear skin tag, right eye coloboma, a thin upper lip, and macrostomia. She had long slender fingers and toes with normal fingernails and toenails. She had severe delay in intellectual and motor development. Brain MRI studies at one years of age showed hypogenesis of the splenium of the corpus callosum. Whole exome sequencing was performed at age 10 years with her parents and a healthy younger brother. A novel de novo *SMARCE1* missense alteration c.752G>A; p.Arg251Gln, was identified and confirmed by Sanger sequencing in the patient. **Discussion:** We could not verify that the novel de novo missense alteration is the sole cause of the patient's phenotype of OAFNS. Further studies with more cases are necessary to evaluate if *SMARCE1* is responsible for OAFNS. It might be necessary to include the other genes involved in SWI/SNF complex to identify the responsible gene(s) for OAFNS.

1236W

Proposed clinical diagnostic criteria for Myhre syndrome: Preparation for evidence-based guidelines. L.J. Starr¹, A.E. Lin². 1) University of Nebraska Medical Center, Omaha, NE; 2) MassGeneral Hospital for Children, Boston, MA.

Myhre syndrome is a distinctive connective tissue disorder with progressive serious comorbidities caused by de novo missense gain-of-function mutations in SMAD4 (at position 500 and 496). It is increasingly identified by clinical exome sequencing (at least 60 reported, and 15 new patients examined at our centers). Deaths were noted in 20% in the largest reviews, with additional patients known anecdotally. Serious complications are due to pericardial disease, restrictive cardiomyopathy and severe restrictive lung disease. Because of an urgent need to enhance clinical recognition, we propose diagnostic criteria which should lead to more rapid molecular confirmation. It is possible that not all families desire or can afford genetic testing. Furthermore, these criteria will allow assessment whether they can be viewed as evidence-based guidelines.

METHODS: Patients from the literature and new patients were compared to the following proposed clinical diagnostic criteria for Myhre syndrome. We also tested single patients with syndromes resembling Myhre syndrome. I. Diagnosis was based on fulfillment of 5/7 of these defined criteria: •Characteristic facial features •IUGR, short stature, or compact habitus •Hearing loss •Restrictive airway, interstitial lung disease •Cardiac (especially coarctation, PDA, descending aorta hypoplasia, pericarditis, restrictive cardiomyopathy) •Connective tissue/fibroproliferation (limited extension of joints, thick skin, straight spine, post-instrumentation fibrosis, fibrotic and thick scarring) •Neurodevelopmental delay, intellectual disability, autism spectrum disorder, anxiety II. The presence of the typical pathogenic SMAD4 mutation and a score of 2 or more is sufficient for diagnosis of Myhre syndrome. **RESULTS:** Though likely a biased sample, all molecularly confirmed patients with Myhre syndrome fulfilled these diagnostic criteria. **DISCUSSION:** Establishing diagnostic criteria for any syndrome is an iterative process which begins with prior knowledge and is refined as more patients are studied. With a larger cohort, the power of this analysis may be sufficient to suggest clinical diagnosis require all 7 features. **CONCLUSION:** We propose the diagnosis of Myhre syndrome use these clinical criteria to decrease the risk of incorrect or missed diagnoses in a condition where careful anticipatory management is critical for optimal quality of life and survival. We plan to update the current management suggestions into evidence-based guidelines.

1237T

Adopting a gene through Human Disease Genes website series facilitates a clinical diagnosis for rare genetic disorders. B.B.A. de Vries¹, M.J. Nabais Sa¹, S. Jansen¹, A. van Remortele¹, J. Ewals¹, M. Verbruggen¹, D.A. Koolen¹, H.G. Brunner¹, E. Eichler², J. Gecz³. 1) Department of Human Genetics, Nijmegen, Netherlands; 2) Department of Genome Sciences, University of Washington School of Medicine, Seattle, Washington 98185 USA; 3) Adelaide Medical School and the Robinson Research Institute, University of Adelaide, SA 5000 Australia.

Human Disease Genes website series (HDG) is an international library of websites for professional information on the clinical consequences of novel variants in the human genome (<http://humandiseasegenes.nl/>). HDG is an initiative of the Human Genetics Department, Nijmegen, in collaboration with the University of Washington and the University of Adelaide. Each website is moderated by a dedicated team of clinicians and molecular biologists, and collects and provides up-to-date mostly unpublished clinical information about one specific gene or copy number variant. HDG aims to fill the gap between first publication of several cases and consecutive publication of a large review paper. Professionals can use the information of the new and rare disorder for counseling and will have the opportunity to share clinical data. Patients, parents and caregivers will find useful information on the disease and have the opportunity to share detailed clinical information through the website of our partner GenIDA. HDG is also a platform where researchers can collect and share (functional) data. Recently a new version of the websites has been launched which not only allows collection of phenotype data using HPO (<http://humandiseasegenes.nl/wac/professionals/uploadclinicalinformation/>) but also provides an up-to-date overview of the clinical data that have been collected (<http://humandiseasegenes.nl/wac/graphandchart/>). These tools will be highly valuable for the clinical practice. So far, over 100 genes have been adopted (<http://humandiseasegenes.nl/genewebsites/>) by more than 60 moderators worldwide (<http://humandiseasegenes.nl/moderators/>). If you wish to adopt a gene and join the team of moderators please, contact us at the poster or send us a mail to info@humandiseasegenes.com.

1238F

22q11.2 deletion syndrome: An update from the 22q and You Center at the Children's Hospital of Philadelphia. I.M. Campbell, S.E. Sheppard, T.B. Crowley, D.E. McGinn, A. Bailey, M.J. McGinn, M. Unolt, J.F. Homans, H.R. Salmons, J.W. Gaynor, E. Goldmuntz, O.A. Jackson, L.E. Katz, M. Mascarenhas, V.F.X. Deeney, R.M. Castelein, K.B. Zur, T. Kolon, C.B. Solot, E.M. Moss, R. Gur, K.E. Sullivan, B.S. Emanuel, E.H. Zackai, D.M. McDonald-McGinn. Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA.

Background: 22q11.2 deletion syndrome (22q11.2DS) is caused by recurrent low copy repeat mediated copy number losses of chromosome 22. The Children's Hospital of Philadelphia has been involved in the clinical care of individuals with what is now known as 22q11.2DS since our initial report of the association with DiGeorge syndrome in 1982. **Methods:** We utilized our continuously growing longitudinal cohort and reviewed the medical records of 1,421 patients with laboratory confirmed 22q11.2 deletion and evaluated by the 22q and You Center following the introduction of FISH studies from 1992 to 2018. **Results:** The mean age at diagnosis was 3.8 years $\pm$ 7.5 years. The vast majority of patients (84%) had the typical LCR22A-LCR22D deletion, and only 7% of these typical deletions were inherited from a parent harboring the deletion constitutionally. However, 5% of individuals harbored other nested deletions that would not be identified by traditional 22q11.2 FISH, as FISH probes are located between LCR22A-LCR22B. Such cases would require an orthogonal technology such as MLPA, chromosomal microarray, or massively parallel sequencing to diagnose. Major medical problems included immunodeficiency (77%); congenital heart disease (CHD) (69%) including 35% with severe anomalies such as tetralogy of fallot and pulmonary atresia; palatal abnormalities (68%); gastrointestinal difficulties including GERD, dysphagia and constipation (66%); endocrine dysfunction including hypocalcemia, hypothyroidism, and growth hormone deficiency (55%); cervical spine abnormalities (50%); renal anomalies (16%); and airway abnormalities including tracheomalacia and subglottic stenosis (7%). Mean full-scale IQ was 76 $\pm$ 13.5, with no significant IQ difference between individuals with and without CHD. We also analyzed the dermatoglyphic patterns of our cohort and noted no substantial differences from previously reported normal controls. Interestingly, 74% of our cohort described themselves as Caucasian, with only 9% of African descent, differing demographically from our general hospital population of approximately 35% African American. **Conclusions:** These ethnicity results suggest a bias in diagnosis due to failure to recognize facial features, differences in access to our referral center, or possibly population variations in a heritable predisposition to development of 22q11.2 deletions. Further surveillance of our patients will likely elucidate additional clinically relevant findings as they age.

1239W

Germline KRAS mutation-related schwannomatosis. M. Serey¹, L. Mes-siaen², H. Cave³, L. Chamard¹, S. Valmary Degano⁴, L. Van Maldergem¹, J. Piard¹. 1) Centre de Genetique Humaine, Universite de Franche-Comte, Besancon, France; 2) Department of Genetics, University of Alabama, Birmingham, Alabama; 3) Department of Genetics, Robert Debré University Hospital and Paris-Diderot University, Paris, France; 4) Department of Anatomopathology, Université de Franche-Comté, Besançon, France.

Schwannomas are benign tumors of the peripheral nerve sheath. Schwannomatosis is characterized by a predisposition to develop multiple schwannomas and. In case of multiple schwannomatosis, germline mutations of *SMARCB1* and *LZTR1* have been described. However, their identification only accounts for a limited number of cases and in many instances, establishing its molecular basis remains elusive. Its typical clinical presentation is made of multiple benign schwannomas, in early adulthood, associated to a chronic pain syndrome that can be exquisitely located at the site of the tumor or be diffuse and widespread. We describe here a 50 year-old female patient with a full-blown *KRAS*-related Noonan Syndrome and a large series of schwannomas becoming symptomatic from the age of 37 onwards. She underwent multiple surgeries. Later, neurological complications resulted in walking impairment and permanent cervical pain. In her case, a c.40G>A (p.Val14Ile) germline *KRAS* heterozygous missense mutation was identified. Sequencing of *NF2*, *LZTR1*, *SMARCB1* and MLPA on paraffin-embedded blocks taken during surgical procedures was tedious but resulted in identification of the same *LZTR1* frameshift c.2092_2093insA mutation in three independent samples. There is an additional *NF2* c.172_189del mutation with MLPA suggestive of the other *NF2* locus deletion in one of these samples and a loss of the *LZTR1* wild type allele in another sample. Only one patient with a similar association has been reported until now. The observation of two unrelated patients with *KRAS*-associated Noonan Syndrome and schwannomatosis suggests that a constitutional mutation in *KRAS* might be considered the third underlying cause of multiple schwannomatosis. It is a well-known phenomenon that mutations in genes encoding proteins belonging to the RAS-pathway are common to tumors and to constitutional disorders called rasopathies. Only a limited number of these constitutional disorders act as cancer-prone conditions, noticeably in Costello syndrome due to mutations in *HRAS* and, based on our second report of multiple schwannomas, in *KRAS*-related Noonan syndrome. Hence, clinical surveillance of *KRAS*-associated Noonan syndrome patients is strongly recommended for occurrence of these tumors, and conversely, in case of schwannoma, a careful dysmorphological examination may help diagnosing Noonan syndrome.

1240T

De novo *PHF6* mutation in a girl with Borjeson-Forssman-Lehmann syndrome. Y. Kuroda¹, H. Murakami¹, Y. Enomoto¹, Y. Tsurusaki¹, T. Uehara², K. Kosaki², K. Kurosawa¹. 1) Division of Medical Genetics, Kanagawa Children's Medical Center, Yokohama, Kanagawa, Japan; 2) Center for Medical Genetics, Keio University School of Medicine, Tokyo, Japan.

Borjeson-Forssman-Lehmann syndrome (BFLS) is X-linked recessive disorder with moderate to severe intellectual disability, epilepsy, hypogonadism, obesity, and dysmorphic features. BFLS results from a mutation or deletion of *PHF6* gene encoding a plant homeodomain (PHD)-like finger protein. Female carriers are usually unaffected, but twelve female patients with BFLS have been reported with intellectual disability and recognizable dysmorphic features. We reported a female BFLS patient with de novo *PHF6* mutation. Patient is 16 year-old girl who is the first child of healthy and nonconsanguineous parents. She was delivered at 38 week gestation with a birth weight of 2862 g (0.47 SD), length 49.5 cm (0.74 SD), and occipital frontal circumference (OFC) 30.5 cm (-1.81SD). Cranial CT showed the slightly enlargement of ventricles at 1 months. She had atrial septal defect and repair surgery was performed at 1 year and 10 months of age. At six months, she developed the enlargement of OFC and cranial CT indicated a hydrocephalus, subdural hemorrhage, and atrophic cortex. She soon underwent a ventriculoperitoneal shunt due to hydrocephalus. She started visual contact at 3 months, smiling at 5 months, rolling at 9 months, and walking with support at 8 years of age. She could not speak recognizable words at 16 years-of-age, and showed signs of a severe intellectual disability. At the age of 16 years, she was 148.0 cm, with a weight of 42.7 kg. Her dysmorphic features consisted of prominent supraorbital ridges, prominent eyebrows with high arch and synophrys, deep-set eyes, a marked zygomatic arch, a high nasal bridge, a short nose with bulbous nasal tip, a prominent chin, clouded teeth, long shaped, and posterior rotated ears and prominent earlobes, tapering fingers, bilateral syndactyly of the fourth and fifth fingers of the foot, linear hyperpigmentation of skin, and hirsutism. Karyotype was 46,XX. and chromosomal microarray analysis showed no pathogenic copy number variation. In trio exome sequencing, de novo *PHF6* mutation, *PHF6*c.733-3_733-2delTA was identified. X-inactivation study of lymphocytes showed a skewed inactivation pattern. The skewed X-inactivation pattern was noted in almost of female patients of BFLS. This patient had consistent dysmorphic features among previously reported female patients and severe intellectual disability. No patient had been reported with hydrocephalus, which may be new clinical findings in female BFLS.

1241F

Molecular findings of 21 Brazilian patients with Cornelia de Lange syndrome. J.R.M. Ceroni¹, H. Aoi², I. Furquim¹, V.E.H. Kim¹, R.S. Honjo¹, G.L. Yamamoto¹, D.R. Bertola¹, T. Mizuguchi², N. Matsumoto², C.A. Kim¹. 1) Unidade de Genética do Instituto da Criança do Hospital das Clínicas da FMUSP; Av. Dr. Enéas de Carvalho Aguiar, 647 - Cerqueira César, São Paulo - SP, Brazil, ZIPCODE: 05403-000 PHONE: +55(11)2661-8671; 2) Department of Human Genetics, Yokohama City University Graduate School of Medicine; 22-2 Seto, Kanazawa, Yokohama, Kanagawa Prefecture, Japan, 236-0027, PHONE: +81 45-787-2311.

Introduction: Cornelia de Lange syndrome (CdLS) is a rare genetic disorder caused by mutations in five genes - *NIPBL*, *RAD21*, *SMC3*, *SMC1A*, *HDAC8*, with different inheritance patterns. Autosomal dominant CdLS is caused by mutations in *NIPBL*, *RAD21* and *SMC3*, whilst *SMC1A* and *HDAC8* cause X-linked CdLS. Analysis of the five genes combined detects approximately 70% of cases. Although phenotype varies from mild to severe, the typical aspects of the syndrome are characteristic craniofacial appearance with synophrys, highly arched eyebrows, long, smooth philtrum, micrognathia, intellectual disability, limb abnormalities, hirsutism, gastroesophageal reflux, short stature, hearing loss. Moreover, clinical overlap with other syndromes, including different forms of autosomal dominant mental retardation, Rubinstein-Taybi, Fetal alcohol syndrome, Wiedemann-Steiner, Alazami-Yuan syndrome, has been described. **Objective:** To characterize the molecular findings of a Brazilian cohort of patients with clinically suspected CdLS. **Methods:** Informed consent was obtained from all patients or parents/guardians prior to the study. All patients were clinically evaluated at the genetic unit. Whole exome sequencing (WES) was performed in 21 probands (17 trios). **Results:** WES revealed pathogenic variants in CdLS genes in 12/21 (57.1%) patients: *NIPBL* (11 patients) and *SMC1A* (1 patient). Other pathogenic variants were found in 3/21 (14.3%) patients: *SETD5* (Autosomal dominant mental retardation 23), *ANKRD11* (KBG syndrome) and *ZMYND11* (autosomal dominant mental retardation 30). Also, one patient had a variant of unknown significance in *EP300* (Rubinstein-Taybi syndrome 2); this patient was adopted, therefore analysis of biological parents' samples were unavailable. In 5/21 (23.8%) patients no candidate variants for the phenotype were found. **Conclusions:** Most pathogenic variants were found in *NIPBL* 11/21 (52.4%), in accordance with the literature. Syndromes with overlapping features with CdLS, indeed Autosomal dominant mental retardation 23 and KBG syndrome were diagnosed. Moreover, one patient with typical phenotype of CdLS presented a pathogenic variant in *ZMYND11* (autosomal dominant mental retardation 30), a syndrome previously associated only with neurologic phenotype. We suggest *ZMYND11* as an additional gene associated with CdLS, reinforcing the strength of WES in genomic analysis. This is the first report of the molecular aspect of CdLS in a Brazilian Cohort. Support by CNPQ 304852/2016-3.

1242W

A zebrafish model of *foxe3* deficiency demonstrates lens and eye defects with dysregulation of key genes involved in cataract formation in humans.

S. Htun¹, M. Krall¹, D. Anand², D. Hart², S. Lachke^{2,4}, A. Slavotinek¹. 1) Department of Pediatrics, University of California San Francisco, San Francisco, CA; 2) Department of Biological Sciences, University of Delaware, Newark; 3) Department of Biochemistry and Biophysics, Cardiovascular Research Institute, University of California San Francisco, San Francisco; 4) Center for Bioinformatics and Computational Biology, University of Delaware, Newark.

The Forkhead box E3 (*FOXE3*) gene encodes a transcription factor with a forkhead/winged helix domain that is critical for development of the lens and anterior segment of the eye. Monoallelic and biallelic deleterious sequence variants in *FOXE3* cause aphakia, cataracts, sclerocornea and microphthalmia in humans. We used CRISPR/Cas9 injections to target the *foxe3* transcript in zebrafish in order to create an experimental model of loss of function for this gene. Larvae that were homozygous for an indel variant, c.296_300delT-GCAG demonstrated severe eye defects, including small or absent lenses and reduced eye size. This phenotype demonstrated an autosomal recessive pattern of inheritance and was fully penetrant. The lenses of the homozygous *foxe3* indel mutants showed more intense staining with zl-1 antibody compared to control lenses, consistent with increased lens fiber cell differentiation. RNA-Seq result found significant dysregulation of genes expressed in the lens and eye whose orthologues are associated with cataracts in human patients, including *cryba2a*, *cryba111*, *mipa* and *hsf4*. We compared the differentially expressed genes from the *foxe3* indel mutant fish with orthologous genes enriched in the mouse embryonic day (E)12.5 lens according to the bioinformatics database resource iSyTE2.0. We found that 181 genes that were 1.5-fold up- or down-regulated in the *foxe3* indel mutant fish exhibit enriched expression in iSyTE data from the E12.5 lens. Among the *foxe3* down-regulated genes that are also iSyTE lens-enriched are interesting candidate genes for lens development, such as *griffin* (galectin family protein in lens), *mipb* (associated with cataract), *crygmx* (lens gamma crystallin), *cryba2b* (associated with cataract), *tmod1* (associated with lens fiber skeleton organization), *six5* (associated with cataract) and *aldh1a3* (associated with microphthalmia/anophthalmia), as well as new uncharacterized candidates, such as *rnf10* and *tcf7*, among many others. As lens fiber differentiation involves an accumulation of crystallin proteins, the increased crystallin gene expression in the *foxe3* indel larvae is consistent with the hypothesis that *foxe3* inactivation is required for lens fiber differentiation, in that a deficiency of *foxe3* increases the expression of genes involved in this process. These findings demonstrate that this new zebrafish *foxe3* mutant model is highly relevant to the study of the gene regulatory networks conserved in vertebrate lens and eye development.

1243T

A 1.9 MB interstitial deletions of 10q23.2q23.31 in a boy with juvenile polyposis.

H. Numabe¹, Y. Morishima¹, N. Inagaki¹, S. Shiyanagi², Y. Hayashi², H. Kawashima¹. 1) Clinical Genetics Center, Tokyo Medical University Hospital, Tokyo, Japan; 2) Department of Gastrointestinal and Pediatric Surgery, Tokyo Medical University, Tokyo, Japan.

A boy was born after premature rupture of the membranes at 36 weeks and 1 day of gestation. Apgar scores at 1 and 5 min were 7 and 8, respectively. His birth weight was 2910 g (+1.45 SD), height 50.3 cm (+1.99 SD), and head circumference 36.0 cm (+2.75 SD). At 1 month of age, he was referred to us because of macrocephaly and exhibited frontal bossing, low-set ears, high-arched palate, laryngeal stridor, and cryptorchidism. Echocardiography and abdominal ultrasonography revealed no abnormalities. The initial G-banding karyotyping showed reciprocal translocations 46,XY,t(1;10)(p36.2;q23.2),t(2;7)(p13;p13). The karyotypes of his parents were normal 46,XX and 46,XY. At 2 years and 2 months of age, he underwent a colonoscopy that revealed 51 polyps in his colon. Two months later, a polypectomy was performed, removing 35 polyps. Histopathology indicated juvenile polyps in all cases. Whole-genome SNP microarray analysis revealed a 1.9 MB interstitial deletion of 10q23.2q23.31. arr[hg19] 10q23.2q23.31(88,121,043-90,005,755)x1. The Online Mendelian Inheritance in Man database includes 7 (*LDB3*, *BMPR1A*, *GLUD1*, *MINPP1*, *PAPSS2*, *KLLN*, and *PTEN*) in this interval. *LDB3* is an autosomal dominant causative gene of dilated cardiomyopathy 1C with or without LVNC, hypertrophic cardiomyopathy 24, left ventricular noncompaction 3, and myofibrillar myopathy 4. *BMPR1A* is an autosomal dominant causative gene of the infantile form of juvenile polyposis syndrome, hereditary mixed polyposis syndrome 2, and juvenile intestinal polyposis. *GLUD1* is an autosomal dominant causative gene of hyperinsulinism-hyperammonemia syndrome. *MINPP1* is an autosomal dominant causative gene of follicular thyroid carcinoma. *PAPSS2* is an autosomal recessive causative gene of brachyolmia 4 with mild epiphyseal and metaphyseal changes. *KLLN* is a causative gene of Cowden syndrome 4. *PTEN* is an autosomal dominant causative gene of Bannayan-Riley-Ruvalcaba syndrome, Cowden syndrome 1, Lhermitte-Duclos syndrome, and Macrocephaly/autism syndrome, and is an autosomal recessive causative gene of the VATER association with macrocephaly and ventriculomegaly, and a causative gene for somatic endometrial carcinoma, somatic malignant melanoma, PTEN hamartoma tumor syndrome, and head and neck somatic squamous cell carcinoma. In this case, juvenile polyposis is recurrent and aggressive. The boy will need frequent surveillance for not only polyposis, but also thyroid carcinoma and other neoplasms associated with Cowden syndrome.

1244F

A novel variant in *REN* explains a familial case of congenital anomalies of the kidney and urinary tract (CAKUT). A.A. Dillio^{1,2}, J. Wang¹, V.M. Siu^{2,3,4}, E. Brown², G. Singh², R.A. Hegele^{1,2}. 1) Robarts Research Institute, Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada; 2) Department of Biochemistry, Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada; 3) Medical Genetics Program Department of Pediatrics, London Health Sciences Centre London, Ontario, Canada; 4) Children's Health Research Institute, London Health Sciences Centre London, Ontario, Canada.

Congenital anomalies of the kidney and urinary tract (CAKUT) comprises a group of diseases affecting kidney and urinary tract morphology. Although CAKUT is the most common form of congenital malformation, presentation is phenotypically heterogeneous, with a wide range of disease severity and outcomes. Here we present a consanguineous Old-Order Amish couple, who have had two children featuring CAKUT symptoms and severe oligohydramnios by 20 weeks' gestation. Both probands succumbed to pulmonary hypoplasia shortly after birth. We hypothesized the CAKUT phenotype is caused by a single rare, large-effect, autosomal recessive variant. We used a combination of autozygosity mapping and whole-exome sequencing to investigate the genetic basis of the CAKUT phenotype. Both probands' DNA samples were subjected to CytoScan HD microarray to interrogate for regions of autozygosity. Upon confirmation of no genetic dosage abnormalities, the eldest proband's DNA sample was whole-exome sequenced using the Illumina HiSeq 4000. Nonsynonymous rare variant prioritization was performed within Golden Helix VarSeq to identify candidate variants, incorporating the regions of autozygosity. We identified a novel homozygous nonsense variant, p.Tyr297*, in the renin (*REN*) gene and Sanger sequencing confirmed variant co-segregation within the family. At this time, a second closely-related family came forward with their child affected with the CAKUT phenotype, and two unaffected children. Again, Sanger sequencing confirmed co-segregation in this family. All unaffected individuals were heterozygous carriers of the *REN* candidate variant. Fibroblast samples were obtained from both probands for immunoblotting assays, from which we observed significantly decreased *REN* expression when compared to fibroblasts with wild-type *REN*. Transcript expression will be analyzed using RT-PCR. Variants in *REN* have previously been associated with renal tubular dysgenesis; however, functional studies have not confirmed their connection to CAKUT. Additionally, *REN* is not a primary gene screened for mutations in CAKUT patients. Nonetheless, the renin-angiotensin system has a proven role in many aspects of renal branching morphogenesis, which is further confirmed by this finding. These results will add to the current limited knowledge of CAKUT genetics. By stratifying the heterogeneous CAKUT phenotypes by genotype, prognosis prediction will improve and allow for more appropriate genetic counselling for families.

1245W

CHARGE syndrome with atypical clinical phenotypes diagnosed by whole exome sequencing. H. Pang, H. Wang, A. Wright, B. Shah, S. Li, H. Wang. Department of Pediatrics, University of Oklahoma Health Sciences Center, 1100 N. Lindsay, Oklahoma City, OK, 73104, USA.

Introduction: CHARGE syndrome (OMIM# 214800) is a rare autosomal dominant genetic disorder caused by pathogenic variants in *CHD7*. Clinical diagnosis is based on the main features of coloboma, heart defects, choanal atresia, retarded growth and development, genital abnormalities, and ear anomalies. We described here two patients with atypical phenotypes diagnosed by whole exome sequencing (WES). **Case description:** Case 1 is a 1-year-old male with multiple congenital anomalies including interrupted aortic arch, VSD, bilateral cleft lip and palate, and 3rd and 4th syndactyly, L vocal cord paresis, bilateral renal artery stenosis. The FISH, karyotype, and microarray results were normal. After the identification of pathogenic variants in *CHD7* via WES, further investigation revealed bilateral optic nerve colobomas. Case 2 is a 35 weeks premature male with prenatal history of polyhydramnios, PDA, VSD, left-sided congenital diaphragmatic hernia (CDH); postnatal additional findings of conotruncal defects, absent thymus with decreased TRECs, lymphopenia, hypocalcemia, hypoparathyroidism, right-sided facial palsy, seizure-like activities. Considering 22q11.2 deletion syndrome, however, the SNP array result was negative and WES was performed. **WES findings:** WES trio studies were performed. Two heterozygous variants c.4609 A>T (p.Lys1537*) located in exon 20 and c.7922dupT (p.Leu2641Phefs*12) located in exon 21 in the *CHD7* gene (NM_017780.3) was detected for case 1 and case 2, respectively. None of these variants has been reported previously in the literatures. Both variants are *de novo* and are predicted to cause loss of normal protein function. Based on the ACMG guidelines, they were classified as pathogenic variants for CHARGE syndrome. **Conclusions:** Emerging studies reported the wide spectrum of phenotype in CHARGE syndrome and clinical overlap with 22q11.2 deletion syndrome. Immune dysfunction has also been reported in CHARGE syndrome. Our study provided more evidences for the expanded spectrum of phenotype in CHARGE syndrome. WES is a powerful tool in the clinical practice for patients with atypical clinical presentations. We reported two novel pathogenic variants in the *CHD7* gene. Updated diagnostic criteria with inclusion of molecular diagnosis for CHARGE syndrome should be considered for patients with atypical phenotypes.

1246T

Expanding the clinical spectrum of *KDM6A*-associated Kabuki syndrome. C.T. Myers¹, J.T. Bennett^{2,3}, J.L. Merritt^{1,2}, K.D. Tsuchiya¹, J. Vogel¹, L. Rogge¹, J.M. Thies¹, K. Foss¹, C.R. Paschal¹. 1) Seattle Children's Hospital, Seattle, WA, USA; 2) Department of Pediatrics, University of Washington, Seattle, WA, USA; 3) Center for Developmental Biology and Regenerative Medicine, Seattle Children's Research Institute. Seattle, WA, USA.

Kabuki syndrome (KS) is a rare disorder that is recognized clinically by characteristic facial features, minor skeletal and dermatoglyphic abnormalities, intellectual disability, and postnatal growth deficiency. Clinical findings in patients with KS are variable, but can also include congenital anomalies, increased susceptibility to infections, feeding difficulties, seizures, and hearing loss. KS has an autosomal dominant inheritance pattern, with the majority of cases being sporadic. Pathogenic variants in *KMT2D* account for 56-75% of cases, whereas pathogenic variants in *KDM6A* are less frequent (5-8%) and not well-described. We report the clinical and molecular findings of two individuals, each with a pathogenic deletion involving *KDM6A* that was identified by chromosome microarray analysis (CMA). Patient #1 is a 5 year old girl who presented with developmental delay, dysmorphic features, and low serum carnitine beyond the neonatal period. A 519 Kb interstitial deletion of Xp11.3 was identified. This deletion includes the entire coding region of *KDM6A*, and the two other genes in the deleted segment, *DUSP21* and *CXorf36*, have not been well-characterized. Parents were not available for testing. Patient #2 is a female infant who presented with hypoglycemia, hypotonia, bilateral hip dysplasia, feeding difficulties requiring nasogastric feeding, and dysmorphic features. She was found to have a 26 Kb deletion that includes two exons within *KDM6A*, which was confirmed by quantitative PCR. Parental testing revealed that this was a *de novo* change. The two cases presented here highlight the variable phenotype of *KDM6A*-associated KS. The facial features of both individuals were considered to be mildly dysmorphic, but consistent with KS. Both patients reported feeding difficulties, albeit of different severity. To date, approximately 40 patients with *KDM6A* variants have been reported, including 6 with exonic or whole gene deletions. Neonatal hypoglycemia has been reported in a subset of patients with *KDM6A* variants, including one other patient with hyperinsulinemia. Other variable features include long halluces, hypertrichosis, joint laxity, and congenital heart defects. The molecular subtype of *KDM6A* abnormality may contribute to the severity of the phenotype. In females, X-inactivation patterns in different tissues are also likely to contribute to phenotypic manifestations. Larger studies are needed to more fully characterize the spectrum of *KDM6A*-associated KS.

1247F

KBG syndrome diagnosed in a 4-month-old infant with failure to thrive caused by deletion of the *ANKRD11* gene. H. Wang, H. Pang, XF. Wang, YM. Kim, SB. Li, KJ. Wierenga. Department of Pediatrics, The University of Oklahoma Health Sciences Center; 1100 N Lindsay Ave, Oklahoma City, OK 73104.

INTRODUCTION: KBG syndrome (OMIM148050) is an autosomal dominant disorder characterized by multiple congenital anomalies including macrodon-tia, distinctive craniofacial features, short stature, and developmental delay. The diagnosis of KBG syndrome is by detection of pathogenic variant in the *ANKRD11* gene or deletion of the 16q24.3 region encompassing *ANKRD11*. More than 100 individuals have been reported. However, KBG syndrome diagnosed in the infant has rarely been described. **CASE DESCRIPTION:** This is a 4-month-old infant born at 35 weeks with birth weight of 4lbs and 2ozs. She stayed in NICU for 21 days due to difficulty feeding, failure to thrive (FTT) and NG-tube was required. Except small VSD, no other abnormalities were found. At age of 4 months, she was re-hospitalized for FTT with both weight and height below 0 % centiles. G-tube was placed. Mild dysmorphic facial features were noticed including triangular face, hypertelorism, relatively long philtrum and thin lips. The comprehensive laboratory workup were negative. The initial whole genome single nucleotide polymorphism (SNP) array was performed followed by duplication/deletion study of Next Generation Sequencing (NGS). **RESULTS:** SNP array revealed a deletion on chromosome 16q24.3 with size range of 46.7kb to 100.1kb based on the two different values of Log R ratio. The deleted region include intron 2 of *ANKRD11* gene (NM_013275.5) at the minimal or whole coding exons at the maximum. Due to the uncertain size of the deletion, further deletion/duplication study of the *ANKRD11* gene was performed which revealed a heterozygous deletion of all coding exons. **CONCLUSIONS:** The age at the diagnosis in most of the reported cases are from about 1 year to adulthood. Diagnosis of KBG syndrome in the early infant could be challenging as the major feature may be just failure to thrive. Detection of the whole *ANKRD11* gene deletion in our case establish the diagnosis. Except one reported case diagnosed at prenatal stage, the patient we reported here is at the youngest age diagnosed with KBG syndrome. Importantly, uncertain size of deletion found on microarray should be further characterized by alternative methods.

1248W

Mechanistic basis of Takenouchi-Kosaki syndrome: Electron microscopic studies of platelets in patients and functional studies in model organism. T. Takenouchi, T. Uehara, K. Kosaki. Keio University, Tokyo, Japan.

We recently reported two unrelated individuals with thrombocytopenia accompanied with intellectual disability, who had a *de novo* mutation, p.Tyr64Cys, in *CDC42* in both patients. The characteristic hematologic feature, *i.e.* increased platelet size and decreased platelet count (macrothrombocytopenia) was compatible with the existing evidence in mice that homozygous knock- mice in *cdc42* causes similar platelet phenotype. Identification of multiple unrelated patients who had similar phenotype and had a *de novo* *CDC42* mutation established that these patients represent a new syndromic form of thrombocytopenia, which was eponymized as "Takenouchi-Kosaki syndrome" (OMIM #616737). Later, Motokawa M. *et al.* reported the third patient with this condition, and we recently evaluated one additional patient. In the present study, detailed phenotypic comparison was made in a total of four patients with Takenouchi-Kosaki syndrome. Morphologic analysis of platelets derived from three affected individuals was performed using electron microscopy. We further performed the functional analysis of p.Tyr64Cys in *CDC42* through CRISPR/Cas9 gene editing with *Caenorhabditis elegans*. The cardinal features in Takenouchi-Kosaki syndrome include intellectual disability accompanied with structural brain abnormalities and sensorineural deafness, macrothrombocytopenia, and camptodactyly. Electron microscopic analysis of platelet showed increased vacuoles and alpha granules. Functional analysis of p.Tyr64Cys in *CDC42* in model organism suggested hypomorphic effects of the mutant allele. Takenouchi-Kosaki syndrome is clinically recognizable by the intellectual disability, macrothrombocytopenia, and presence of a *de novo* mutation in *CDC42*, *i.e.*, p.Tyr64Cys.

1249T

Deleterious mutation of *NSD1* gene in a Senegalese patient with SOTOS syndrome. R. Ndiaye Diallo¹, K. Diallo¹, B. Niang², E. Pasmant³, JPD. Diop¹, Y. Dia¹, SS. Mbacke¹, PA. Diop¹, O. Faye¹, M. Vidaud³, A. Dieye¹. 1) Faculty of Medicine, Pharmacy and Dentistry University Cheikh Anta Diop Dakar, Senegal; 2) Service de Pédiatrie Hôpital d'Enfant Albert Royer CHUN de Fann, Dakar, Senegal; 3) MCU-PH IEA7331, Université Paris Descartes Service de Génétique et Biologie Moléculaires, Hôpital Cochin, Paris.

Sotos syndrome is a rare autosomal dominant childhood overgrowth condition, characterized by distinctive facial features, intellectual disability, macrocephaly and overgrowth of the body in early life. In almost 90% of cases, mutations occurred in the gene encoding nuclear receptor-binding SET domain containing protein 1 (*NSD1*). Here we report a Senegalese patient diagnosed with SOTOS syndrome at age 3 months. He was the second child of a nuclear family and has excessive body growth (size > Z3DS), macrocephaly and facial dysmorphism (pronounced occipital horns, elongated face, palpebral anti-mongoloid clefts). After informed consent of both parents, blood samples were collected from the index case and each parent for genetic testing. Genomic DNA was extracted and all coding exons and flanking intronic regions of *NSD1* (NM_022455.4), *DNMT3A* (NM_175629.6) and *SETD2* (NM_014159.6) genes, were screened for mutations by multiplex PCR followed by next generation sequencing of amplicons. Identified mutations were systematically verified by Sanger technology. We identified in the index case a heterozygous mutation in exon 5 of the *NSD1* gene, leading to a frameshift and premature termination of protein synthesis: c.2306dup (NM_022455); p. Gly771Trpfs*38 (NP_071900). This mutation occurs *de novo* in the index case and was not detected in either parents. This is the first case of SOTOS syndrome genetically diagnosed in Senegal.

1250F

Complete spectrum of *PIK3CA* mutations in isolated lymphatic malformations. C.V. Cheng¹, D. Jensen², A.E. Timms², G. Shivaram^{3,5}, R.A. Bly^{4,5}, S. Ganti^{4,5}, A. Kirsh², W.B. Dobyns^{1,6}, M.W. Majesky², J.T. Bennett^{4,5,6}, J.A. Perkins^{4,5}. 1) Center for Integrative Brain Research, Seattle Children's Research Institute, Seattle, WA; 2) Center For Developmental Biology And Regenerative Medicine, Seattle Children's Research Institute, Seattle, WA; 3) Interventional Radiology, Seattle Children's Hospital, Seattle, WA; 4) Seattle Children's Hospital, Division of Pediatric Otolaryngology, Department of Otolaryngology/Head and Neck Surgery, University of Washington, Seattle, WA; 5) Vascular Anomalies Program, Seattle Children's Hospital, Seattle, WA; 6) Seattle Children's Hospital, Division Of Genetic Medicine, Department Of Pediatrics, University of Washington, Seattle, WA.

Somatic mosaic activating mutations in *PIK3CA* are the most common cause of isolated lymphatic malformations (LMs). According to a previous study, 4 frequently mutated residues, "hotspots," (p.E542K, E545K, H1047R and H1047L) account for 80% of LMs. Although non-hotspot mutations in *PIK3CA* are prevalent in human sporadic cancers, they have never been reported in LMs. This data is sparse and the complete spectrum of *PIK3CA* mutations in isolated LMs has yet to be determined. We used single molecule molecular inversion probes (smMIPs) to sequence *PIK3CA* in 287 samples across 88 individuals with isolated LMs. This technique permits detection of low-level alternate allele fractions (AAF) of genetic variants with barcoding. The data was analyzed using the mosaic variant caller PISCES. As a secondary method for targeted confirmation, we used droplet digital PCR (ddPCR). *PIK3CA* genotypes and AAFs were also correlated with clinically relevant variables. Fifty-four of the 88 individuals had pathogenic or likely pathogenic *PIK3CA* variants with sufficient sequencing coverage, yielding a diagnostic rate of 61% by smMIPs. Of all the mutations detected in *PIK3CA*, 93% of the variants are in 1 of the 4 hotspots. We also identified 3 non-hotspot mutations (p.G106V, Q546K, and F937S), 1 of which has never been reported, and 2 of which have been previously reported in both cancer and overgrowth syndromes. Variants were generally detected in resected LM tissue and not in blood or skin overlying the lesion, and AAFs ranged from 1 to 22%. Preliminary genotype-phenotype comparisons suggest a correlation between less severe LMs and non-hotspot mutations. Full gene sequencing using smMIPs to detect low-level AAFs in tandem with sensitive analysis algorithms is an effective method for uncovering mutations across the *PIK3CA*-LM spectrum. However, since the overwhelming majority of detected variants are in hotspots, a sensitive, allele-specific method like ddPCR may be a more efficient first line assay. Identification of the underlying genetic etiology of isolated LMs is valuable as enthusiasm for medical therapies like sirolimus and other PI3K-AKT-mTOR inhibitors increases.

1251W

Personalized medicine in rare diseases and cancer: A case report of a lasting response in a young teenage patient with Proteus syndrome and secondary ovarian cancer. C. Leoni¹, G. Gullo², N. Resta³, A. Fagotti⁴, R. Onesimo¹, B. Schwartz⁵, J. Kazakin⁵, C. Ranieri⁵, G. Scambia⁴, G. Zampino¹. 1) Center for Rare Diseases and Birth Defects, Institute of Pediatrics, Department of Woman and Child Health, Fondazione Policlinico Universitario A. Gemelli IRCCS, Università Cattolica del Sacro Cuore, Rome, Italy; 2) Medical Oncology Department, St Vincent's University Hospital, Dublin, Republic of Ireland; 3) Division of Medical Genetics, Department of Biomedical Sciences and Human Oncology (DIMO) University of Bari "Aldo Moro", Bari, Italy; 4) Division of Gynecologic Oncology, Department of Woman and Child Health, Fondazione Policlinico Universitario A. Gemelli IRCCS, Università Cattolica del Sacro Cuore, Rome, Italy; 5) ArQule, Inc., Burlington, Massachusetts, United States of America.

Proteus syndrome (PS) is an ultra-rare disease characterized by progressive, disproportionate, segmental overgrowth caused by a somatic gain of function mutation p.Glu17Lys in the oncogene *AKT1*. The disease has high morbidity and mortality rates as well as increased risk of cancer development. Below is a case report of a 15 year old girl with a molecular diagnosis of PS and a relapsed inoperable *AKT1*(p.Glu17Lys) mutant low grade serous ovarian cancer (LGSOC) treated with the AKT inhibitor miransertib on a compassionate basis. During the first six years of life, the patient underwent multiple surgical amputations due to progressive overgrowth of hands and feet. At the age of 12 she underwent a bilateral salpingo-oophorectomy and hysterectomy due to LGSOC (International Federation of Gynecology and Obstetrics, FIGO stage IA). Targeted deep sequencing assay of PI3K/AKT pathway genes and comprehensive cancer panel (409 genes) showed the only oncogenic driver of LGSOC was the single mosaic gain of function mutation in *AKT1* gene (p.Glu17Lys, 15% mutant allele frequency). Twenty months after surgery, CT detected a recurrent solid mass close to the bladder and a partial portal vein thrombosis. Considering significant contraindications to standard therapeutic options, neither cytoreductive surgery nor palliative cytotoxic chemotherapy was recommended. After approval by local EC and baseline biochemical and imaging staging, the patient started treatment with miransertib at 100 mg/daily, 7 days on and 7 days off every 14 days. After 16 weeks of treatment CT showed a partial response (PR), according to RECIST 1.1 criteria and a complete resolution of the portal thrombosis. The 12-month CT scan confirmed the ongoing PR (~50% from baseline). During treatment a progressive reduction in bone and connective tissue overgrowth, especially in the hands, was observed; moreover, the patient reported a progressive improvement in joint mobility (hands, spine and knees) and, therefore, in her quality of life. No major side effects were recorded during 12 months of therapy. To our knowledge, this is the first documented case of PS showing a relapsed *AKT1*(p.Glu17Lys) mutant LGSOC, in which treatment with an AKT inhibitor led to an objective reduction of overgrowth lesions and sustained tumor response. Clinical and molecular data give a unique insight into the pathogenesis of this neoplasm in patients with PS and demonstrate that personalized medicine is feasible in ultra-rare diseases.

1252T

SUZ12 related overgrowth. W.T. Gibson^{1,2}, S.S. Cyrus¹, N. Tkachenko³, K. Avela⁴, B.H.Y. Chung⁵, H.M. Luk⁶, S. Choufani⁷, R. Weksberg⁷. 1) Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada; 2) Provincial Medical Genetics Program, Children's and Women's Health Centre of British Columbia, Vancouver, BC Canada; 3) Centro Hospitalar do Porto, Portugal; 4) Helsinki University Hospital, Department of Clinical Genetics, Finland; 5) Department of Paediatrics & Adolescent Medicine, University of Hong Kong, China; 6) Clinical Genetic Service, Department of Health, Hong Kong, China; 7) Genetics & Genome Biology Program, Hospital for Sick Children, Toronto, Canada.

The Polycomb Repressive Complex type 2 (PRC2), is a chromatin regulator, which mono-, di- or tri-methylates lysine 27 on Histone 3 (H3K27)- a repressive methylation mark. Three components of PRC2 have been implicated in rare overgrowth syndromes: Enhancer of Zeste2 (*EZH2*), Embryonic Ectoderm Development (*EED*) and Suppressor of Zeste 12 (*SUZ12*). In 2011, coding mutations in *EZH2* were found to be causative of Weaver Syndrome. Weaver Syndrome features overgrowth (tall stature and macrocephaly), intellectual disability, early bone maturation, characteristic facial features and susceptibility to malignancies. Subsequently, people with a phenotype similar to Weaver Syndrome were found to have *de novo* coding mutations in *EED*. This led to the recognition of the Cohen-Gibson Syndrome. In 2017, the first case of an individual with a rare coding variant in *SUZ12* and an overgrowth phenotype was described. Here we present 4 individuals (3 families – from Portugal, Finland and Hong Kong) with rare coding variants in *SUZ12*, the largest cohort to date. A c.1807T>C; p.(Phe603Leu) variant was transmitted from father-to-son; the c.1715_1716insCA; p.(Leu572fsPheTer11) and c.1878delT; p.(Phe626LeufsTer7) variants were identified in individual probands. Emerging features of *SUZ12*-overgrowth include prenatal overgrowth, postnatal tall stature, macrocephaly, large hands and feet and early bone maturation. Intellectual disability is variable and may be absent. None of the cases reported to date have developed haematological or solid malignancies. Notably, constitutional pathogenic variants in *NSD1* cause Sotos Syndrome, and the mutations that disrupt *NSD1*'s methylation of H3K36 result in a "downstream" DNA methylation signature in patients with Sotos syndrome. Since coding variants affecting PRC2 members also perturb its H3K27 methylation, these mutations may also generate a characteristic DNA methylation signature. Furthermore, rare variation in PRC2 complex members is collectively common among individuals not known to have rare genetic syndromes (as described within the ExAC and gnomAD populations). Given the multiple clinical features manifest by patients with syndrome-causing mutations in PRC2 members, and the population-wide prevalence of rare coding variants in *EZH2*, *EED*, *SUZ12* and other epigenetic regulators, it is plausible that rare coding variants in epigenetic regulators may also confer risk for common, complex diseases in the general population.

1253F

Identification and validation of variants of uncertain significance in novel candidate genes for 46, XY disorders/differences of sex development (DSD). A. Parivesh, H. Barseghyan, M. Almalvez, M. Bramble, E. Délot, E. Vilain. Center for Genetic Medicine Research, Children's National Medical Center, Washington, DC.

Background: Disorders/differences of Sex Development (DSD) have an estimated frequency of 0.5% of live births encompassing a variety of urogenital abnormalities ranging from mild hypospadias to sex reversal. Exome sequencing was performed on a subset of DSD cases with 46, XY karyotype identifying the underlying cause in ~35% of cases. To investigate the relevance of variants of unknown significance (VUS) in the unresolved cases, we utilized a mouse model in which the presence of a Y chromosome from the *poschiavini* strain (*Y^{pos}*) on a C57BL/6J (B6) background results in XY undervirilization and sex reversal, a phenotype characteristic to human 46, XY DSD cases. We measured RNA expression of the genes carrying VUS in 46 XY human cases and assessed differences of expression between the gonads of B6-*Y^{pos}* and under-virilized B6-*Y^{pos}* males at E11.5. This identified a list of 15 new candidate genes for XY DSD. **Objective:** To validate the novel candidate genes for 46, XY DSD by assessing the possible effects of novel mutations on protein function. **Methods:** Among the candidate genes identified by this study, *Adamts16* is expressed in embryonic gonadal tissue under the control of sex-determining gene *Sox9*. *ADAMTS16* belongs to the ADAMTS family of integral membrane and secreted glycoproteins that have matrix metalloproteinase functions and have several known proteolytic substrates. It is known to bind and cleave $\alpha 2$ macroglobulin (A2MG). *A2MG* and *ADAMTS16* in wild-type and two mutant forms were cloned into bacterial vectors for protein expression. *in vitro* binding/cleavage assays using the purified proteins were used to assess the mechanisms by which the novel *ADAMTS16* mutations affect binding to and/or cleavage of A2MG. **Results:** Mouse studies and exome sequencing identified 15 novel candidate genes potentially involved in 46, XY DSD pathogenesis. We used *in silico* prediction algorithm and showed that the two mutant forms of *ADAMTS16* (p.Arg100Trp; p.Phe469Val) should affect propeptide cleavage and peptidase activity respectively. We have set up an *in vitro* model that allows us to test this hypothesis and provide functional validation for some of the novel candidate genes that harbor VUS.

1254W

IDH1 and IDH2 mutations in a cohort of patients with Ollier disease and Maffucci syndrome. E.S. Partan^{1,2}, R. Martin², S. Robbins^{1,2}, L. Cirulli³, D. Valje^{2,4}, N. Sobreira^{2,4}. 1) Predoctoral Training Program in Human Genetics, Johns Hopkins School of Medicine, Baltimore, MD; 2) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD; 3) Duke University School of Medicine, Duke University, Durham, NC; 4) Department of Pediatrics, Johns Hopkins School of Medicine, Baltimore, MD.

IDH1 and IDH2 catalyze the oxidative decarboxylation of isocitrate to α -ketoglutarate (α KG). Variants in the active sites of IDH1 (p.R132) and IDH2 (p.R140/R172) result in the neomorphic ability of the enzyme to instead convert α KG to 2-hydroxyglutarate. Germline heterozygous variants in *IDH2* cause D-2-hydroxyglutaric aciduria type 2 (D2HGA). Somatic *IDH1* and *IDH2* variants have been observed in a wide variety of cancers. These variants have also been identified in 80% of the enchondromas, chondrosarcomas, and vascular anomalies in individuals with Ollier disease (OD) and Maffucci syndrome (MS). OD and MS are related conditions characterized by enchondromas with a 30% risk of chondrosarcoma transformation; individuals with MS also have vascular anomalies. We performed Sanger sequencing and/or whole exome sequencing on germline DNA from 25 probands with OD or MS and on tumor DNA from 16 samples from 11 probands. We identified an *IDH1*-p.R132C variant in a spindle cell hemangioma sample from a patient with MS; this variant was not present in two enchondromas from the same proband. The same p.R132C variant was also found in an enchondroma and chondrosarcoma of two probands with OD, while another proband with OD had a p.R132H variant. We also identified a proband with OD who had a homozygous *IDH1*-p.M91T (gnomAD 8.121e-6) variant in a chondrosarcoma and a heterozygous *IDH2*-p.I152T (gnomAD 0) variant in an enchondroma. Another proband with metaphyseal enchondromatosis with D2HGA had a heterozygous germline *IDH1*-p.R132C and a heterozygous somatic *IDH2*-p.T435M (gnomAD 0.003) variant in two enchondroma samples, the latter of which was not found in his blood. Finally, two other probands with enchondromatosis each had a germline, rare, heterozygous *IDH2* variant inherited from an unaffected parent (p.D225N and p.T435M). In summary, we identified *IDH1* or *IDH2* variants in the tumor samples of six probands, three of which were not in the previously identified positions, and germline variants in three probands, two of which were in novel positions. The identification of rare *IDH1* and *IDH2* variants not previously described in these disorders, the occurrence of germline variants, and the occurrence of multiple variants in some probands are novel findings in this group of disorders. Further investigation of the function of these different germline and somatic variants in isolation or in combination may improve our understanding of the pathophysiology of these disorders.

1255T

A familial novel whole *CDKN1C* gene deletion in siblings with Beckwith-Wiedemann syndrome and Dandy-Walker malformation. D. Aljeaid, R.J. Hopkin. Human Genetics, Cincinnati Children Hospital Medical Center, CINCINNATI, OH., OH.

Beckwith-Wiedemann syndrome (BWS) is a human genomic imprinting disorder with estimated prevalence of 1:10,340. It affects males and females equally. Originally described by Beckwith and Wiedemann (in 1953 and 1964) with hallmark features of omphalocele, macroglossia and macrosomia. The phenotypic features of BWS are highly variable that include hemihyperplasia, visceromegaly and increased risk of embryonal tumors. BWS is caused by genetic and epigenetic changes within two imprinting domains on chromosome 11p15 resulting in excess expression of growth genes. We present a family with two affected siblings (male and female) exhibiting clinical features of BWS and Dandy-Walker malformation due to a novel whole gene deletion of *CDKN1C*. Both patients were identified with omphalocele on prenatal ultrasound, additional findings of polyhydramnios, placentomegaly, macroglossia and DWM with enlarged cisterna magna and cerebellar vermis hypoplasia were detected by fetal MRI. On exam, they both had ear creases and pits. Family history significant for a brother who died at 2 months of age with sudden infant death syndrome, 6-year-old sister with umbilical hernia and 1-year-old brother with ear pits. Both siblings had normal chromosome microarray, IC1 and IC2 methylation testing and *CDKN1C* sequencing. Deletion of the entire coding sequence of *CDKN1C* gene was identified by deletion/duplication analysis. Maternal blood analysis was negative for the deletion, raising the suspicion of germline mosaicism. Paternal and siblings' testing are pending at this time. It's been reported that up to 15% of cases of BWS are familial and commonly due to *CDKN1C* mutations (maternally inherited with 50% recurrence risk), chromosome 11p15 abnormalities and genetic alterations with IC1. Pathogenic variants (mainly loss of function) in *CDKN1C* gene accounts for ~5% of sporadic cases and 40% of familial cases. Abnormalities of central nervous system have rarely been reported in BWS, such as Chiari malformation, hydrocephalus, schizencephaly and posterior fossa abnormalities including DWM, vermis hypoplasia and Blake's pouch cyst. Associated molecular defects reported involved the centromeric imprinted domain on 11p15.5 that includes IC2 and *CDKN1C*. It's been suggested that these genes are implicated in midline development of multiple organs including the brain. Clinical findings of this family provide additional support of CNS involvement in BWS and illustrate the phenotypic spectrum of BWS.

1256F

A de novo gain-of-function mutation in MYCN causes a novel megalencephaly syndrome. K. Kato^{1,2}, F. Miya^{3,4}, N. Hamada⁵, Y. Negishi¹, Y. Narumi-Kishimoto⁶, H. Ozawa⁶, H. Ito⁶, I. Hori¹, A. Hattori¹, N. Okamoto⁷, M. Kato⁸, T. Tsunoda^{3,4}, Y. Kanemura⁹, K. Kosaki¹⁰, Y. Takahashi¹, K. Nagata⁸, S. Saitoh¹. 1) Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan; 2) Nagoya City University Graduate School of Medical Sciences.; 3) Tokyo Medical and Dental University, Tokyo, Japan; 4) RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 5) Aichi Human Service Center, Kasugai, Japan; 6) Shimada Ryoiku Center Hachioji, Tokyo, Japan; 7) Osaka Women's and Children's Hospital, Osaka, Japan; 8) Showa University School of Medicine, Tokyo, Japan; 9) Osaka National Hospital, National Hospital Organization, Osaka, Japan; 10) Keio University School of Medicine, Tokyo, Japan.

Background: In this study, we aimed to identify the gene abnormality responsible for pathogenicity in an individual with an undiagnosed neurodevelopmental disorder with megalencephaly, ventriculomegaly, hypoplastic corpus callosum, intellectual disability, postaxial polydactyly and neuroblastoma. We then explored the underlying molecular mechanism with *in vitro* and *in vivo* experiments. **Methods:** Trio-based whole-exome sequencing was performed to identify disease-causing gene mutation. Biochemical and cell biological analyses were carried out to elucidate the pathophysiological significance of the identified gene mutation. **Results:** We identified a *de novo* heterozygous missense mutation (NM_005378.5; c.173C>T; p.Thr58Met) in the MYCN gene, at the Thr58 phosphorylation site essential for ubiquitination and subsequent MYCN degradation. The missense mutation was not listed in any public databases (e.g. ExAC) and Thr58 residue is evolutionally conserved. Mycn in whole-tissue extracts of mouse brains at various developmental stages showed high expression level at E13.5, then gradually decreased to its lowest level by postnatal day. The mutant MYCN (MYCN-T58M) was non-phosphorylatable at Thr58 and subsequently accumulated in cells and appeared to induce CCND1 and CCND2 expression in neuronal progenitor and stem cells *in vitro*. The mutation also promoted neuronal cell proliferation in the intermediate and subventricular zones during corticogenesis in mouse embryos. **Conclusions:** We identified a *de novo* c.173C>T mutation in MYCN. The temporal expression pattern of Mycn protein in the mouse brain reflected the situation in humans, and suggests Mycn plays an important role during corticogenesis. MYCN-T58M leads to stabilization and accumulation of the MYCN protein, leading to prolonged CCND1 and CCND2 expression. This may promote neurogenesis in the developing cerebral cortex, leading to megalencephaly. While loss-of-function mutations in MYCN are known to cause Feingold syndrome, this is the first report of a germline gain-of-function mutation in MYCN causing a novel megalencephaly syndrome similar to, but distinct from, CCND2-related megalencephaly-polymicrogyria-polydactyly-hydrocephalus syndrome. The data obtained here provides new insight into the critical role of MYCN in brain development, as well as the consequences of MYCN defects.

1257W

Biallelic mutations in EXOC3L2 cause a novel syndrome that affects the brain, kidney and blood. A. Shalata¹, S. Lauhasurayotin^{2,3}, Z. Leibovitz⁴, S. Dhanraj^{2,3}, H. Moshiri², H. Bakry⁵, H. Li^{2,3}, A. Arad⁶, Y. Hadid⁷, M. Mahroum¹, D. Herbert⁸, S. Egenburg⁹, J. Bajar⁸, S. Haddad⁸, M. Shohat⁸, S. Scherer⁸, S. Tzur^{9,10}, Y. Dror^{6,11}. 1) Simon Winter Inst. For Human Genetics Bnai-Zion Medical Center, Israel; 2) Genetics and Genome Biology Program, The Hospital for Sick Children, Toronto, Canada; 3) Marrow Failure and Myelodysplasia Program, Division of Hematology/ Oncology, Department of Pediatrics, The Hospital for Sick Children, Toronto, Canada; 4) Obstetrics-Gynecology Ultrasound Unit, Bnai-Zion Medical Center and Rappaport Faculty of Medicine, The Technion, Haifa, Israel; 5) Department of Pathology, Bnai Zion Medical Center, Haifa, Israel.; 6) Division of Nephrology, Department of Paediatrics, The Hospital for Sick Children.; 7) Ultrasound unit-Gyn-Obst Department- Baruch Padeh Peoria Hospital, Israel; 8) Sheba Cancer Research Center, Sackler School of Medicine, Tel Aviva University, Maccabi Genetic Institute, Israel; 9) Laboratory of Molecular Medicine, Rappaport Faculty of Medicine and Research Institute, Technion - Israel Institute of Technology, Haifa, Israel; 10) Genomic Research Department, Emedgene Technologies, Tel Aviv, Israel; 11) Institute of Medical Sciences, University of Toronto, Toronto, Canada.

Background Dandy-Walker malformation features include *agenesis/hypoplasia* of the cerebellar *vermis*, cystic dilatation of the *fourth ventricle* and enlargement of posterior fossa. Although Dandy-Walker malformation is relatively common and several genes were linked to the syndrome, the genetic cause in the majority of cases is unknown. Ciliary ... **Objective:** to identify the mutated gene responsible for Dandy-Walker malformation and kidney disease in two unrelated families. **Methods:** Medical assessment, sonographic, MRI imaging and pathological studies followed by CMA-chromosomal microarray analysis and WES-whole exome sequence were performed in three fetuses in the first family and in one patient in the second unrelated family. **Results:** We report four subjects from two unrelated families with homozygous mutations in the Exocyst Complex Component 3-Like-2 gene (*EXOC3L2*). *EXOC3L2* functions in trafficking of post-Golgi vesicles to the plasma membrane. In the first family a missense mutation in a highly conserved amino acid, p.Leu41Gln, was found in three fetuses; all had severe forms of Dandy-Walker malformation that was detectable by prenatal ultrasonography and confirmed by autopsy. In the second family the affected child carried a nonsense mutation, p.Arg72* and no detected protein. He had peritrigonal and cerebellar white matter abnormalities with enlargement of the ventricular trigones, developmental delay, pituitary hypoplasia, severe renal dysplasia and bone marrow failure. **Conclusion:** We propose that biallelic *EXOC3L2* mutations lead to a novel syndrome that affects hindbrain development, kidney and possibly the bone marrow.

1258T

Association of hydrocephalus and renal dysplasia with a homozygous *DLG5* frameshift variant in an alternatively spliced exon. F. Vogel¹, Z. Yüksel¹, A. Alhashem², K. Kampe¹, K.K. Kandaswamy¹, M. Werber¹, A.M. Bertoli-Avella¹, A. Rolfs^{1,3}, P. Bauer¹. 1) Centogene AG, Rostock, Germany; 2) Prince Sultan Military Medical City, Pediatrics, Riyadh, Saudi Arabia; 3) Albrecht-Kossel-Institute for Neuroregeneration, Medical University Rostock, Rostock, Germany.

The *DLG5* gene encodes the 1919-residue DLG5 protein which is composed of numerous functional domains, and has been implicated in a variety of cellular structures and processes. Homozygous knockout of the murine homologue *Dlg5* results in fully penetrant but variably severe hydrocephalus and renal cysts. Certain *DLG5* haplotypes have been suggested to predispose to inflammatory bowel disease, but a *DLG5*-linked monogenic phenotype has not previously been reported. We investigated a pediatric female patient diagnosed with obstructive hydrocephalus, renal dysplasia, atrial septal defect type II, and cleft lip and palate. Bilateral ventriculomegaly and echogenic kidneys had been noticed at gestational week 22. The parents are first cousins and there is a history of intra-uterine fetal death due to similar congenital abnormalities. Whole exome sequencing (WES) of the index case in a routine diagnostic setting did not reveal potentially pathogenic variants in known disease genes. An extended analysis of the WES data in a research frame identified the homozygous *DLG5* frameshift variant c.3081_3106del (p.Arg1027fs) in *DLG5*. Sanger sequencing confirmed this finding, and revealed heterozygosity in both healthy parents. We were therefore very surprised to note that data from the Exome Aggregation Consortium (ExAC) strongly suggest that heterozygous presence of a truncating *DLG5* variant is not tolerated (probability of loss-of-function intolerance = 1.00). A more detailed analysis, however, revealed that our variant and the resulting pre-terminal stop codon localize to the alternatively spliced in-frame exon 15. The shorter of the corresponding isoforms is therefore predicted to escape nonsense-mediated decay, and this, in turn, implies that certain functions of DLG5 may be retained upon presence of the c.3081_3106del variant. Our clinical-genetic findings, together with those on the murine *Dlg5* knockout, characterize *DLG5* as a novel recessive disease gene in humans. Theoretical considerations suggest that heterozygous truncating variants in exons other than exon 15 may be associated with a potentially distinct phenotype. Further studies are necessary to test this hypothesis.

1259F

De novo structural and single-nucleotide variants (SNVs) underlying associations between birth defects and childhood cancers: An assessment from the Genetic Overlap Between Anomalies and Cancer in Kids (GOBACK) Study. H. Li¹, P. Mishra¹, S. Sisoudiya², M.E. Scheurer^{3,4}, R.A. Gibbs^{1,2}, P.J. Lupo^{3,4}, S.E. Plon^{2,3,4}, A. Sabo¹. 1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 2) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 3) Department of Pediatrics, Baylor College of Medicine, Houston, TX; 4) Texas Children's Cancer Center, Texas Children's Hospital, Houston, TX.

As validated in our population-based study of 10 million live births, one risk factor for childhood cancer is being born with congenital anomaly. However, little is known about genetic architectures underlying this association, especially the role of structural variants (SVs), which are likely to be important. Here, we evaluated methods for identifying SVs, as well as SNVs, and pinpointed candidate variants underlying the congenital anomaly-cancer (CAC) phenotypes. Methods: From GOBACK, we selected 18 trios with various CAC combinations and performed WGS (40X). SVs were detected via 4 callers: Lumpy, Delly, BreakDancer, and Manta, followed by merging and filtering via SURVIVOR. A consensus call was made by at least 2 callers. We kept *de novo* events that were not detected by any caller in both parents and eliminated variants identified in the 1K Genomes Project as well as an internal project of healthy controls. Rare *de novo* SNVs were identified via Platypus and were annotated with respect to genes and functions using multiple algorithms. Results: We identified 4 *de novo* deletions across the 18 probands. While 3 of those were intergenic, a 5kb *de novo* intragenic deletion was validated in a patient with acute lymphocytic leukemia (ALL) and multiple anomalies. This *de novo* mutation resulted in an exonic deletion of *USP9X*, which encodes a deubiquitinating enzyme. Notably, *USP9X* is among the top genes recurrently mutated somatically in pediatric tumors, including ALL, and *USP9X* loss-of-function mutations are known to cause a similar multiple-anomaly syndrome as observed in this kid. Additionally, we identified 12 *de novo* nonsynonymous SNVs that may alter the function of protein-coding genes, including several that have been reported in congenital anomalies or cancers, such as *FOXK2*. We also identified a likely functional intronic mutation that overlaps with regulatory histone markers within *TCF12* in a patient with diencephalic syndrome (DS) and pilocytic astrocytoma (PA). *TCF12* has been implicated in multiple cancers, including PA, which is known to relate to the pathogenesis of DS. Conclusion: The GOBACK study has resulted in discoveries of several variants, both SVs and SNVs, that may underlie CAC phenotypes. The identification of a 5kb *de novo* deletion in a patient illustrates the need for comprehensive evaluation of all types of variants, from SNVs to SVs. Validation studies on children with similar CAC and assessment of a full spectrum of SVs are underway.

1260W

Missense variants in *PLS3* in patients with X-linked congenital diaphragmatic hernia. F. Petit^{1,2}, B.R. Pober³, R.D. Clark⁴, P. Giampietro⁵, H.H. Ropers⁶, H. Hu⁶, A-S. Jourdain², F. Frenois², M-A. Delrue⁷, B. Gilbert-Dussardier⁸, L. Devisme^{2,9}, B. Keren¹⁰, S. Manouvrier-Hanu^{1,2}, P. Bhayani¹¹, M. Loscertales¹¹, P.K. Donahoe^{11,12}, F.A. High^{11,13}, M. Longoni^{11,12}. 1) Clinique de Génétique, CHU Lille, Lille, France; 2) EA7364 RADEME, Université de Lille, Lille, France; 3) Department of Pediatrics, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA; 4) Loma Linda School of Medicine, Loma Linda, CA, USA; 5) Drexel University School of Medicine, Philadelphia, PA, USA; 6) Max Planck Institute for Molecular Genetics, Berlin, Germany; 7) Service de Génétique, CHU Bordeaux, Bordeaux, France; 8) Service de Génétique, CHU Poitiers, EA3808, Université de Poitiers, Poitiers, France; 9) Institut de Pathologie, CHU Lille, Lille, France; 10) Département de Génétique, Hôpital Pitié Salpêtrière, CHU Paris, France; 11) Pediatric Surgical Research Laboratories, Massachusetts General Hospital, Boston, MA, USA; 12) Department of Surgery, Harvard Medical School, Boston, MA, USA; 13) Department of Surgery, Boston Children's Hospital, Boston, MA, USA.

Congenital diaphragmatic hernia (CDH) has an incidence of around 1/3000 births. The most frequent genetic causes of sporadic CDH are recurrent chromosome abnormalities and *de novo* autosomal mutations. However, several lines of evidence suggest the existence of X-linked loci for CDH. Linkage analysis in a three-generation family with X-linked CDH in 10 affected individuals pointed to a 3 Mb region on the X chromosome (LOD score > 3), comprising 14 protein-coding genes. Using exome sequencing, we identified a familial missense variant in *PLS3* within this interval. Additional missense variants were discovered in 2 unrelated families with 2 and 3 affected males, respectively. These *PLS3* variants alter highly conserved residues and co-segregate with a novel multiple-anomaly congenital condition, characterized by CDH with or without heart and anterior body wall defects in males, and mild dysmorphic features in some female carriers. *PLS3* is expressed almost ubiquitously in solid tissues with replicative potential and has been shown to be required for the maintenance of actin microfilament bundles. Loss-of-function mutations and deletions have been reported previously in families with X-linked osteoporosis and osteoporotic fractures (MIM# 300910). We hypothesize that distinct mechanisms are at play with CDH-associated *PLS3* variants, as they map to the actin-binding domain of the protein, and they are predicted by *in silico* structural modeling to specifically disrupt interaction with actin. Published functional evidence in yeast supports a deleterious effect of substitutions at one of these residues, a fully conserved tryptophan in the actin-binding site, in the Plastin 3 ortholog fimbrin. We also showed that Pls3 is expressed in the diaphragm and lung mesenchyme of E13.5 mouse embryos. The expression pattern and timing are compatible with a role in diaphragm development for Plastin 3. Based on human and functional evidence, we propose that *PLS3* is a candidate for syndromic X-linked CDH.

1261T

Mutation spectrum of the *KMT2D* gene for the Kabuki syndrome in Korea. Y. Yeon, Y. Gu, M. Kim, S. Cho, M. Seong, S. Park. Seoul National University Hospital, Seoul, Select a Country.

Kabuki syndrome (KS) is characterized by distinct dysmorphic facial features, mental retardation and congenital anomalies including short kidneys, skeletal and intestinal abnormalities, cardiac and immunological defects. The *KMT2D* and *KDM6A* are the only causative genes which are known to cause KS. Whole exome sequencing (WES) was performed on 12 suspected patients with Kabuki syndrome through the rare disease genetic diagnosis project in Korea. This study revealed that 75% of probands (12 patients) harbored causative mutations in *KMT2D* genes. 55.9 % (6 / 9) of all positive cases were small insertion and deletion (indel) mutation, and nonsense mutation by single nucleotide substitution account for 33.3% (3 / 9). There was one missense mutation. From 5 small indel mutations, four were occurred by one base deletion, but a splicing mutation by complex indel was found in one case. For negative cases, meticulous reevaluation based on WES can uncover other novel causative genes.

1262F

Dual diagnoses: The first 400 patients of the CAUSES Research Study.

A.M. Elliott¹, J. Mwenifumbo¹, C. du Souich¹, S. Adam¹, C. van Karnebeek², T.N. Nelson³, M. Couse¹, L. Armstrong¹, G. Aubertin¹, R. Salvarinova², S. Stockler², M.L. McKinnon¹, S. Lewis¹, L. Clarke¹, G. Horvath², E. Lopez¹, K. Selby², A. Lehman¹, J.M. Friedman¹, *The CAUSES Study*. 1) Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada; 2) Department of Pediatrics, University of British Columbia, Vancouver, British Columbia, Canada; 3) Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, British Columbia, Canada.

Background: Genome-wide (exome or whole genome) sequencing (GWS) is particularly useful in patients with rare disease and can identify multiple concomitant single gene disorders in one patient. CAUSES, a trio-based translational research study, provides GWS to 500 children with suspected genetic disorders. Once the candidate variant list is generated, there is a multidisciplinary team approach to the discussion of variants that includes bioinformaticians, clinical and laboratory geneticists, genetic counsellors, and the referring physician(s). Variants are classified on a research basis as definitely causal, probably causal, uncertain or negative, taking clinical findings and family history into consideration. Definite, probable and uncertain (but compelling candidate) disease-causing variants are validated by Sanger sequencing and interpreted by a clinical laboratory. **Methods:** We reviewed Sanger-confirmed variants in the first 400 families seen in the CAUSES study to identify patients with >1 candidate gene potentially explaining the phenotype(s). Patients with an incidental/secondary finding as their second causal variant were not included. **Results:** In total, 19 patients (4.75%) underwent Sanger validation of >1 gene. Combinations of research classifications for causality included: two patients - definite/definite; two - definite/probable; four - definite /uncertain (including one patient with three genes Sanger validated); five- probable/uncertain and five with both variants classified as uncertain. **Discussion:** We reviewed this cohort of 19 patients and determined features that are shared, distinct and unanticipated based on the diagnoses. As an illustrative example, a male patient presented with mild developmental delay, cleft of the soft palate, severe aortic valve stenosis and aortic arch hypoplasia, imperforate anus, sacral vertebral anomalies and 12th rib hypoplasia. Facial dysmorphism included hirsutism, bushy eyebrows, long palpebral fissures and a short nose. This patient was subsequently found to have variants in both *KMT2D* (Kabuki) and *CHD7* (CHARGE). He lacked some of the cardinal features of each disorder but many of his findings can be seen in both disorders. This analysis highlights genotype/phenotype correlations and the importance of determining all relevant diagnoses in order to optimize clinical treatment, care and genetic counselling.

1263W

Integrated sequence analysis approaches improve genomic diagnosis of birth defects. K. Meltz Steinberg^{1,2}, D. Baldrige², J. Wambach², D. Wegner², D. Spencer^{1,3}, M. Shinawi², F.S. Cole². 1) McDonnell Genome Institute, Washington University School of Medicine, St. Louis, MO; 2) Department of Pediatrics, Washington University School of Medicine, St. Louis, MO; 3) Department of Medicine, Washington University School of Medicine, St. Louis, MO.

Structural birth defects affect approximately 3% of US born infants and are a common cause of infant mortality. Although environmental factors can cause birth defects, a majority is caused by inherited or *de novo* genetic variants; however, the current diagnostic yield from genomic testing only ranges between 25-40%. To improve diagnostic success in children with structural birth defects, we integrated exome, whole genome, and transcriptome sequencing data from 56 affected trios. Our bioinformatic pipeline identifies and characterizes single nucleotide variants (SNVs), small insertions/deletions (indels), copy number (CNVs) and structural variants (SVs). We first confirmed the sensitivity of our approach by correctly identifying, in a blinded fashion, 100% of pathogenic variants in ten trios previously diagnosed by clinical exome sequencing. Secondly, we reanalyzed with our pipeline 12 trio exomes that were non-diagnostic and identified biallelic pathogenic variants in two patients. For example, in a child with short stature, small hands and feet, cone-shaped epiphyses of hands and feet, brachytelephalangy of digits, dolicocephaly, persistent sagittal suture, and ADHD, we identified a novel missense and a rare non-canonical splice variant in *PHYH*, a gene mutated in Refsum disease. We confirmed exon 3 skipping encoded by the non-canonical split site variant in approximately half the *PHYH* transcripts in the proband's peripheral blood transcriptome and elevated serum phytanic acid concentration that is biochemically diagnostic of Refsum disease. Thirdly, we developed a CNV/SV detection strategy that combines three paired end/split read callers with two read depth callers. We benchmarked our pipeline on Genome in a Bottle samples, NA12878 and NA24385. For deletions ranging in size from 1kb to 1Mb, sensitivity ranges from 91.2 to 95.7%. We then compared sequenced based CNV/SV calls in the probands with CytoScanHD chromosomal microarray variant calls. We validated 100% of clinically relevant deletions and duplications greater than 100kb. Additionally, for all deletions 50-100kb and greater than 100kb we have 73% and 93% sensitivity, respectively. On average per proband, we identify five coding deletions and ten coding duplications that overlap ClinGen likely pathogenic or pathogenic variants. We conclude that we have developed a robust, integrated approach that improves diagnostic success in children with birth defects.

1264T

Failure to detect germline pathogenic variants in mTOR pathway (PIK3CA-AKT-mTOR) does not exclude a clinical diagnosis. A case report. A.I. Sanchez, J.A Rojas. Pontificia Universidad Javeriana, Bogota, Colombia.

Introduction: mTOR pathway diseases encompasses well known neurological disorders characterized by altered cortical architecture, intractable epilepsy, megalencephaly, severe ID, and autistic spectrum disorder. Germline and somatic mutations in different genes of this pathway are responsible for Smith-Kingsmore syndrome (SKS), ganglioglioma, focal cortical dysplasia, tuberous sclerosis complex, *PIK3CA* gene related overgrowth spectrum disorders (PROS) and isolated hemimegalencephaly among others. **Case report:** 5 years old male with megalencephaly, global developmental delay, and seizures born at 38 weeks. Mom 15 yo G1P1 and father 19 yo. No consanguinity. C-section for breech presentation, birth weight 3,04 Kg. At age 8 months old found to have hypotonia, global developmental delay, speech delay. Started PT, OT and speech therapy with slow improvement of milestones. Followed by neurology/DDBP and physical medicine. Walked at 4 yo. At age 5 yo, has difficulties for walking, says only 5 words, no phrases. Half brother on dad's side with severe global developmental delay and epilepsy; 20 yo paternal aunt with intellectual disability and no verbal communication. PE: macrocephaly (HC: 56 cm) prominent forehead, epicanthal folds, wide nasal bridge, cupped ears, narrow arched palate. Metabolic labs were normal (lactate/pyruvate, plasma AA, urinary organic acids and ammonia). Normal brain MRI. Proton brain MR spectroscopy with variations in NAA/Cr and Cho/Cr ratios compatible with developmental delay. Microarray was normal. Trio whole exome sequencing (certified genetic laboratory) was normal and trio whole genome sequencing (certified genetic laboratory) was positive for two heterozygous VUS in *FRAS1* gene (c.4556G>A p.Arg1519Gln and c.7171G>A p.Gly2391Ser), no genotype/phenotype correlation. **Conclusions:** Differentials include SKS and isolated megalencephaly with intellectual disability due to possible brain *mTOR* or *PIK3CA* somatic mutations, respectively. Brain biopsy in this case is difficult. Paternal gonadal mosaicism has to be taken into account in this case as well as an incomplete penetrant disorder. A clinical discernible disorder in this patient is probably due to mTOR pathway dysregulation. This is of interest because of therapeutic approaches (mTOR inhibitors). Overlapping phenotypes due to malregulation in any component of mTOR pathway are based on the presence of germinal or somatic mutations in the same gene and in the degree of mosaicism.

1265F

Identification of potential regulatory variants causing Branchio-Oto-Renal syndrome in an extended pedigree using Whole Genome Sequencing (WGS). P.A. Sanchez-Lara¹, C.E. Schwartz², C. Bodea³, R.L. Maas³, J.M. Graham, Jr.¹. 1) Pediatrics, Cedar Sinai Medical Center/University of California Los Angeles, Los Angeles, CA; 2) Greenwood Genetics Center, Greenwood, SC; 3) Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA.

Branchio-oto-renal (BOR) syndrome is an autosomal dominant disorder characterized by deafness, branchial arch anomalies, ear malformations, and renal anomalies. It has a prevalence of 1/40,000 and accounts for 2% of profoundly deaf children. The genetics of BOR syndrome is complex. Approximately 40% of patients with BOR syndrome have a mutation in the *EYA1* gene on chromosome 8q13.3, with over 203 published mutations spanning all 16 exons. In addition to coding sequence variants, 20% of mutations in *EYA1* involve complex cytogenetic rearrangements resulting in deletions or duplications. However, no variants or mutations spanning *EYA1* regulatory regions have been reported in the medical literature. The second most common gene associated with BOR syndrome is *SIX5* on chromosome 19q13.3, accounting for 2.5% of cases, followed by *SIX1* on chromosome 14q23, which is implicated in 2% of cases (Smith 2009). Approximately 55% of BOR syndrome cases have an unidentified mutation despite meeting clinical diagnostic criteria. Taken together, figure suggests that other genes may be involved. We present a large 5-generation family with 11 living affected individuals that meet the clinical criteria for BOR syndrome. Clinical findings include: branchial cleft cysts, ear anomalies (microtia, ear tags, ear pits), hearing loss, epibulbar dermoids, renal anomalies, facial nerve palsy, gustatory lacrimation, and seizures. Sanger sequence analysis of the *EYA1* gene did not identify any mutation or potential candidate. A BAC array and two independent high-density SNP arrays (Affymetrix 6.0 and NimbleGen 2.1M platforms) that included coverage for the *EYA1* locus revealed no copy number variations. Data from the SNP Array was used for linkage and flagged a region between rs11785226 and rs7018254 on chromosome 8. *EYA1* is at chr8:72109668-72268979, just outside of the linkage region (chr8:53,680,808-72,017,177) [hg19]. We hypothesized that the phenotype in this family was due to a non-coding regulatory variant in the *EYA1* gene. Unfortunately it was not possible to test *EYA1* gene expression on a sample from this family given that the gene is not expressed in blood or skin cells. Through the NIDCR sponsored Facebase2 project, additional samples were obtained and Whole Genome Sequencing (WGS) was completed on several affected individuals in the pedigree. Potential candidate regulatory elements can be assessed via future *in vitro* functional assays.

1266W

Intra-familial variable expressivity of orofaciodigital syndrome type 1 in three generations of French Canadian family. *S.A. Basalom, L.R. Russell.* Medical Genetics, McGill University, Montreal, Quebec, Canada.

Background Orofaciodigital syndrome (OFD) type 1 (MIM 311200) is a rare, X-linked ciliopathy that is generally lethal in males. The clinical features are highly variable and include alterations in the face, oral cavity, limbs, kidney and brain. Intellectual disability has been reported in 46-58% of affected individuals. We report the clinical features, radiological findings and the *OFD1* variant in a woman, her mother and her child. **Clinical cases** The proband is a 31 year old French Canadian female who presented in childhood with mild intellectual disability and syndactyly of the hands and feet, the latter requiring surgery. Surgery for cleft palate was reportedly performed in infancy. At 29 years of age, she presented with new onset, unprovoked seizures. On MRI she was found to have agenesis of the corpus callosum with neuronal migration abnormalities including polymicrogyria of the frontal and parietal lobes, bilateral band heterotopia, subependymal heterotopia, mainly on the left, cerebellar vermis hypoplasia and abnormal cerebellar folia. The cerebral hemispheres were asymmetric, the left hemisphere being more hypoplastic than the right. The physical exam showed microcephaly, mild facial asymmetry, high arched palate and multiple gingival frenulae. We also noted clinodactyly and syndactyly of several fingers but no brachydactyly or polydactyly. After she became pregnant at 31 years, a fetal MRI at 31 6/7 weeks gestation showed a female fetus with bilateral frontal polymicrogyria, agenesis of corpus callosum with secondary colpocephaly, and pontocerebellar hypoplasia with extremely hypoplastic cerebellar vermis. The daughter was born at term and on physical examination had lingual hamartomas, bifid tongue and natal teeth. A postnatal MRI confirmed the prenatal findings. On physical examination, the infant's grandmother had syndactyly of the left hand, left double hallux and lingual hamartomas. **Results** A comprehensive brain malformation panel revealed a frameshift heterozygous, likely pathogenic variant in *OFD1*: c.1168_1171delAATT. **Conclusion** This family demonstrates well the intra-familial variable expressivity present in *OFD1*. As well, the late diagnosis of *OFD1* in the proband suggests that *OFD1* may be underdiagnosed. We recommend consideration of *OFD1* in the differential diagnosis of brain malformations in females, particularly when there is a wide spectrum of malformations.

1267T

Post-natal correction of a BBsome gene, *Bbs8*, rescues fertility in a mouse model of Bardet-Biedl syndrome. *M.R. Cring, J.E. Garrison, G. Kim, V.C. Sheffield.* University of Iowa, Iowa City, IA.

Bardet-Biedl syndrome (BBS) is a genetically heterogeneous, autosomal recessive disorder that affects primary cilia biogenesis and function. The cardinal features of BBS include retinal degeneration, obesity, polydactyly, kidney anomalies, intellectual disability, and male infertility. No treatments are currently available for this disease. There are 21 genes currently known that can independently cause BBS. Eight of these protein products together form a complex called the BBsome, which is important for trafficking proteins to the cell membrane and well as retrograde trafficking along the ciliary axoneme. Males with BBS are infertile as they predominately lack spermatozoa flagella. When flagella are present, they are typically shortened or dysmorphic, and TEM images show clear disorganization of microtubules and outer dense fibers. Otherwise, acrosome formation and nuclear condensation appears normal. Using a gene trap mouse model which allows for post-natal correction of a BBsome component, *BBS8*, we have shown that post-natal correction of *BBS8* can rescue fertility. Furthermore, we show that this correction is not time dependent, and can occur in older mice as well as young mice. TEM imaging shows that microtubule and outer dense fiber organization are rescued following post-natal correction of *BBS8*. Using a conditional knock-out mouse model for *BBS8*, which allows disruption of *BBS8* following tamoxifen administration, we show that flagella are still present in mice for at least 15 days following *BBS8* knock-out. This indicates that BBS proteins are expressed early during the process of spermatogenesis. Future work is aimed at determining how long flagella persist after loss of *BBS8*. We also intend to use CRISPR-Cas9 mediated homologous recombination in a mouse model to repair the most common BBS mutation and thereby address potential therapies for BBS patients.

1268F

Expanding the phenotype of CEP55-related disorders: A case report.

R. Mostafavi¹, A. Choudhri^{2,3}, E.K. Pivnick^{2,4}. 1) St. Jude Children's Research Hospital, Memphis, TN; 2) Le Bonheur Children's Hospital, Memphis, TN; 3) University of Tennessee Health Science Center, Memphis, TN; 4) UT Le Bonheur Pediatric Specialists, Memphis, TN.

CEP55 is a centrosomal protein with a key role in mitosis and cytokinesis. CEP55 also contributes to stabilization of the Akt/PI3K pathway to suppress Caspase-3 mediated apoptosis. In interphase, CEP55 localizes to the centrosome and is later recruited to the midbody to facilitate completion of cytokinesis after phosphorylation occurs at residues S425 and S428, and subsequently at S436 by Erk2, Cdk1, and Plk1 respectively. CEP55 protein lacking residues 355-464 cannot undergo phosphorylation, which prevents dissociation of CEP55 from the centrosome and recruitment to the midbody in cytokinesis. TSG101 and ALIX bind with CEP55 and colocalize at the midbody to enable abscission. Previous case reports of CEP55-related disorders present as lethal autosomal recessive ciliopathies. A homozygous nonsense variant (c.256C>T; p.R86*) was found in two fetuses with Meckel-like syndrome (hydranencephaly and encephalocele, cystic hygroma, renal cysts, and clubfeet). Also, a homozygous nonsense variant (c.1274C>A; p.S425*) was identified in 3 fetuses of another family with features of MARCH syndrome (multinucleated neurons, anhydramnios, renal dysplasia, cerebellar hypoplasia, and hydranencephaly). We present 3 sisters with microlissencephaly, dysmorphic features, and severe intellectual disability who are homozygous for a CEP55 splice site variant (c.1191+2T>C). Parents are consanguineous and are heterozygous for the CEP55 variant. Our cases, as well as the previous case reports, demonstrate the clinical heterogeneity of CEP55-related disorders. The previously reported cases were more severely affected than our siblings, who are currently living at ages 19, 14, and 3 years. However, the homozygous nonsense mutation at position 86 in the Meckel-like family is thought to result in lack of the TSG101- and ALIX-binding domain of CEP55, disrupting abscission. The homozygous nonsense mutation in the MARCH syndrome family resulted in truncation of CEP55 and lack of the 3 phosphorylated serine residues that are necessary for CEP55 dissociation from the centriole and localization to the midbody. The milder phenotype of our cases may be due to the loss of the phosphorylated residues and aberrant RNA splicing. RNA studies for these sisters are in process. Given the phenotypic variation of our cases and those previously reported, CEP55 should be included in workup of patients presenting with a suspected ciliopathy.

1269W

Whole exome sequence identified the deletion in 5' UTR or upstream intronic region of CREBBP in a patient with Rubinstein-Taybi syndrome.

Y. Enomoto¹, Y. Tsurusaki¹, Y. Kuroda², H. Murakami², K. Ida², M. Umegae², N. Harada², Y. Kimura¹, T. Naruto¹, K. Kurosawa^{1,2}. 1) Clinical Research Institute, Kanagawa Children's Medical Center, Yokohama, Kanagawa; 2) Division of Medical Genetics, Kanagawa Children's Medical Center, Yokohama, Kanagawa; 3) Clinical Laboratory, Kanagawa Children's Medical Center, Yokohama, Kanagawa.

Rubinstein-Taybi syndrome (RSTS) is an autosomal dominant disease characterized by broad thumbs and great toes, short stature, moderate to severe intellectual disability (ID) and a distinct facial appearance. Their common facial features are downslanted palpebral fissures, low hanging columella, high palate, grimacing smile, and talon cusp. RSTS arises in rate of about 1/100,000-125,000. The causative genes are CREBBP and EP300, which have similar functions as histone acetyltransferase that regulates transcription via chromatin remodeling. 50-70% of patients have mutation in CREBBP and 3% have mutation in EP300. Here we report a Japanese boy patient with RSTS. He has severe ID, broad great toes, hirsutism, exotropia, congenital heart defects (CoA, VSD), cryptorchidism. Pathogenic variant was not detected in CREBBP by panel-based next-generation sequencing (Haloplex (Agilent Technologies)). However, CNV analysis using whole exome sequence (WES) detected a *de novo* deletion in the region including ADCY9, which localizes in the proximity to CREBBP. The deletion breakpoints are unclear at present. However, WES data showed normal 2 copies alleles in at least coding region of exon1 in CREBBP. Thus, there is a possibility that the deletion in 5' UTR or upstream intronic region of CREBBP caused the disease in patient.

1270T

Alterations in *TANGO2* on the intact chromosome 22q11.2 allele as a possible cause of sudden death. D.M. McDonald-McGinn¹, B. Nowakowska^{2,4}, T.B. Crowley², M. Unolt^{1,3}, D.E. McGinn¹, T. Gambin², M. Geremes², M. Hestand⁴, R. Ahrens-Nicklas¹, E.H. Zackai¹, S. Hopkins⁵, V. Vetter⁶, S. McCormack¹, M. Falk¹, J. Vermeesch⁴. 1) Human Genetics, Children's Hospital of Philadelphia and UPENN School of Medicine, Philadelphia, PA; 2) Instytut Matki Dziecka, Warsaw, Poland; 3) Ospedale Bambino Gesù and La Sapienza University, Rome, Italy; 4) KU Leuven, Leuven, Belgium; 5) Neurology, Children's Hospital of Philadelphia and UPENN School of Medicine, Philadelphia, PA; 6) Cardiology, Children's Hospital of Philadelphia and UPENN School of Medicine, Philadelphia, PA; 7) Endocrinology, Children's Hospital of Philadelphia and UPENN School of Medicine, Philadelphia, PA.

Background: 22q11.2DS is the most common CNV in humans with an incidence of ~1/2000 live births. Highly variable phenotypic expressivity is a hallmark of the condition as no clinical feature is fully penetrant. We previously described allelic variations of genes within the 22q11.2 region of the non-deleted chromosome unmasking autosomal recessive conditions to explain such variability. Here we report the first patients with both 22q11.2DS and deletions/mutations of the *TANGO2* gene on the other allele. Homozygous and compound heterozygous variants in *TANGO2* have been identified in patients with recurrent metabolic encephalomyopathic crises associated with rhabdomyolysis, cardiac arrhythmias, and neurodegeneration. Thus, this novel observation is critical in providing surveillance/intervention (perhaps lifesaving) for this subset of vulnerable patients. **Methods:** Sequencing was performed using Illumina HiSeq after exome capture with the Sure Select All Human V6 design. ANNOVAR was used to annotate relevant information for gene names, predicted variant pathogenicity, reference allele frequencies, metadata from external resources, and then added to the VCF. A HMZDeFinder algorithm was used to search for small, hemizygous deletions within 22q11.2 and compared with WES from 164 samples sequenced on the same platform and processed using the same pipeline as controls. **Results:** Two putative hemizygous deletions encompassing 9 exons of *TANGO2*, were identified in two unrelated patients, including one with a vascular ring, hearing loss, hypotonia, microcephaly, brisk reflexes, episodic dystonia, hypocalcemia, hypothyroidism, hypoglycemia and a SUDC at age 5 years. *TANGO2* truncating mutations were then found in 5 additional patients of our 1422 followed with 22q11.2DS. Deletions and mutations were validated by Sanger sequencing and parental origin was estimated. **Conclusions:** We hypothesize that the deletion including *TANGO2* is responsible for the SUDC in our 5 year old patient and that the deletion and mutations in *TANGO2* are responsible for the atypical features in our remaining 5 patients. Overall, deletions and mutations in *TANGO2* are common and based on our data may occur in as many as 1:200 patients with 22q11.2DS. Thus, investigation of the other allele in all patients with 22q11.2DS appears to be warranted to ensure identification of autosomal recessive conditions that could alter the overall phenotype, approach to genetic counseling and medical management.

1271F

Multiple *de novo* genes variants: A frequent issue in homozygous admixed Puerto Rican (Hispanic) population. A.S. Cornier^{1,2,3}, F. Velez^{4,5}, S. Carlo^{1,6}, E. Albino⁷. 1) Genetic Section, San Jorge Children's Hospital, Santurce, PR; 2) Department of Public Health, Ponce Health Sciences University, Ponce, PR; 3) Department of Pediatrics, Universidad Central del Caribe School of Medicine, Bayamon, PR; 4) Department of Pediatrics, San Juan City Hospital, San Juan, PR; 5) Genetics Division, Stanford University, Palo Alto, CA; 6) Department of Biochemistry, Ponce Health Sciences University, Ponce, PR; 7) School of Health Professions, Medical Science Campus, University of Puerto Rico.

Populations with a high degree of homozygosity and admixture may produce individuals with several pathogenic dominant variations within different genes. This poses a challenge in determining which one(s) or how much of the respective proteins alterations is/are responsible for the phenotype. We are presenting the case of a three year old female, initially evaluated at 12 months in the Genetics clinic due to hypotonia and developmental delay. The child is the product of a non-consanguineous marriage. She was born to a 26-year-old female G2P2 with an unremarkable pregnancy. BW: 7lbs 13oz and BL: 19 ½ inches. Developmentally, she rolled over at age 7 months, sat at 9 months, and crawled at 15 months of age. At age 18 months, she was not walking nor had any language. Physical examination was significant for frontal bossing, mild hypertelorism, long eyelashes and moderate to severe peripheral and central hypotonia. Initial laboratory work up demonstrated marked ketosis and lactic/pyruvate aciduria with normal amino acids analysis, as well as normal carnitine, acylcarnitines and ammonia profiles. Chromosomal microarray studies showed a 426 kb gain of 1p31.3 arr [hg19]1p31.3 (61,699,735-62,125,907) x3. Whole exome sequencing studies documented five novel *de novo* variants, including an SCN8A heterozygous variant c.1084G.T (p.Ala362Ser), a TRPV4 heterozygous variant (c.69T.G p.Ser23Arg); a ZBTB18 heterozygous variant (c.1140G.T p.Met380Ile), a RERE heterozygous (variant, pathogenic?) c.2483C.G (p.Pro828Arg), and heterozygous pathogenic variant on MARS2, c.171dupA(p.Thr573Asnfs*49). Similar variants in four of these genes have been determined to have an autosomal dominant mode of inheritance, and are responsible for developmental delay, intellectual disabilities and hypotonia. We have characterized the genotype/phenotype associated to these mutations in this patient by establishing the molecular pathophysiology of the variants and its implications in our patient's phenotype. Multiple gene mutations/variations including more than one genetic disorder are not uncommon in the Puerto Rican population. The availability of comprehensive genetic variation data from cohorts of Hispanic patients have the potential to greatly improve the identification of disease-associated genetic variants in densely populated small geographical areas such as Puerto Rico. This provides insight into the complexities of admixture and homozygosity within the context of clinical phenotypes.

1272W

Using deep phenotyping to refine the association of location of XPD mutations with function.

G.W. Nelson¹, M.A. Levoska^{2,3}, J.M. Pugh², P.F. Hannon⁴, D. Tamura², E.R. Heller⁵, S.G. Kahn², M. Scheibye-Knudsen⁶, J.J. DiGiovanna², K.H. Kraemer². 1) Advanced Biomedical Computing Center, Frederick National Lab, NCI, Frederick, MD, MD; 2) Laboratory of Cancer Biology and Genetics, NCI, Bethesda, MD; 3) Medical Research Scholars Program, NCI, Bethesda, MD; 4) Michigan State University College of Human Medicine, East Lansing, MI; 5) Department of Dermatology, University of Pennsylvania, Philadelphia, PA; 6) DNA Repair Section, National Institute on Aging, Baltimore, MD.

XPD is a helicase component of the TFIIH complex, involved in both DNA repair and transcription. Different variants in XPD cause radically different phenotypes, reflecting the dual function of the protein and the complex, with repair-affecting variants causing xeroderma pigmentosum (XP) while transcription-affecting variants cause trichothiodystrophy (TTD). XP is characterized by extreme sensitivity to sun damage and sun-induced cancers. The characteristic symptom of TTD is brittle, sulfur-deficient hair, but subjects have a broad range of disorders: skeletal, developmental, immune, and beyond. In addition there is an intermediate phenotype, sharing milder aspects of the two diseases. Both XP and TTD have a range of severity, suggesting a multidimensional gradation of phenotypes. We sought to obtain a more complete picture of the range of phenotypes, and better predictors of the severity of disease, by deep phenotyping 67 subjects with two inherited mutations in XPD. We used data on 176 phenotypes, including characteristic symptoms (e.g. degree of sun sensitivity, and hair structure), a wide range of additional disease phenotype (hearing loss, skeletal abnormalities, measures of developmental delay) and standard clinical measurements. Previous studies have mapped mutations associated with XP and TTD onto the 3D structure of XPD, but there is not a simple dichotomy between these diseases. Our deep phenotyping data gives a continuous and multidimensional measure of phenotypic variation, with the primary axis distinguishing XP and TTD. Also, assigning individual variants to a specific disease is confounded by most subjects being compound heterozygotes, but PCA separates the effect of the individual's two mutations, as long as one of the mutations occurs twice or more. We present the 51 mutations present in our subjects, mapped onto the structures of XPD and TFIIH determined by x ray crystallography and freeze fracture EM. Coloring the residues by the principal components of the phenotypes reveals a continuum of functional information about each mutation site. A novel observation is that mutations associated with an intermediate phenotype occur preferentially on protrusions of XPD, possibly indicating a role in TFIIH complex formation. A limitation is that some mutations will significantly alter protein structure, and some mutations fall on the tail of the TTD protein, whose structure is undetermined.

1273T

TAB2 deletion syndrome in a female with cardiac abnormality, short stature and normal intelligence.

M.C. van Rijn¹, M.F.C.M. Knapen², M.W. Wessels¹, F.S. Jehee¹, L.J.C.M. van Zutven¹. 1) Dept. Clinical genetics, Erasmus MC, Rotterdam, Zuid-Holland, Netherlands; 2) Dept. of Obstetrics and Prenatal Medicine, Erasmus MC Rotterdam, Zuid-Holland, Netherlands.

A 32 years old Romanian patient was referred for genetic counselling during her first pregnancy because of a suspected syndromic condition. She was operated twice before the age of 1 year for obstruction of the oesophagus or stomach, medical details were no longer available. At the age of 6 years she was diagnosed with a perimembraneous VSD and mitral valve and tricuspid valve prolaps with moderate insufficiency. Psychomotor development was normal. The father of our patient was reported to have similar cardiac problems and physical features as well as a normal intelligence. Physical examination showed a disproportionate short stature (-2,7 SDS) with a relatively long trunk and short extremities, and a small head circumference (-2,5 SD). She had some dysmorphic features including, long face, small eyes with short palpebral fissures, a beaked nose with a broad tip of the nose, small mouth with a high palate, retrognathia, small jaws and dental crowding. The hands were small with short fingers, feet were short (size 2-3) and broad with relatively short toes and a sandal gap. A large scar was visible on the abdomen after juvenile gastric surgery. There was hyperlaxity of some of the joints. As some of the features resembled Feingold syndrome, DNA analyses was performed of the MYCN gene, no pathogenic variant was found. SNP array analysis (0.15 Mb resolution) revealed a loss of 2.7 Mb in band 6q24.3q25.1, containing 19 genes including the TAB2 gene (OMIM 605101). Ultrasound was regularly performed during pregnancy revealing no abnormalities in the fetus. A healthy baby boy was born. DNA analysis showed he was not a carrier of the maternal TAB2 deletion. About 40 TAB2 deletion syndrome patients were described in the medical literature so far (Cheng 2017). Overlapping features are facial dysmorphisms, short stature, cardiac abnormalities and connective tissue abnormalities. Most patients show developmental delay and/or intellectual disability. In addition to this patient with normal intelligence, only one family is described with several affected persons without intellectual disability. Gastric or esophageal abnormalities requiring surgery in childhood were not reported so far. .

1274F

Early onset developmental delay and epilepsy in pediatric patients with WDR45 variants. H. Wang, H. Chen, Y. Qian, Y. Lu, X. Dong, B. Wu, W. Zhou. Children's Hospital of Fudan University, Shanghai, China.

Background: Developmental delay (DD) is a neurological disorder presented with defects in gross motor, fine motor, language and cognition. WDR45 is one of the disease-causing genes of DD. Previously, WDR45 de novo mutations were reported in certain adult and pediatric patients due to iron accumulation. **Clinical report:** We reported five pediatric female patients with DD and epilepsy. Their ages were below 3 years at the first consultation and precision diagnosis were difficult based on clinical information and phenotype. **Methods:** Children with DD and/or epilepsy presenting to the molecular diagnostic center of Children's Hospital of Fudan University were enrolled between May 2016 and May 2017. Whole-exome sequencing (WES) was conducted for patients and their parents. We described the phenotypes of these patients carrying WDR45 variants. Furthermore, we overexpressed the candidate variants in HeLa cells to evaluate their effect on autophagy and detected the changes using Western blot and immunofluorescence staining with confocal microscopy. **Results:** In 623 enrolled pediatric patients (274 females; 487 below 6 years), five WDR45 de novo mutations were detected, including c.19C>T (p.Arg7*), c.401G>C (p.Arg134Pro), c.503G>A (p.Gly168Glu), c.700C>T (p.Arg234*), and c.912delT (p.Ala305Leufs*25). All the five patients with WDR45 variants exhibited DD and epilepsy. The missense mutations of p.Arg134Pro and p.Gly168Glu showed accumulation of LC3-containing autophagic structures and abnormally enlarged cell volume compared with the control HeLa cells. Western blotting revealed a significant increase in LC3II/GAPDH. **Conclusion:** The identification of WDR45 mutations provided further evidence that WES played an important role in the diagnosis of neurological disorders with common phenotypes, and WDR45 is associated with neurological disorders and not very rare in Chinese female pediatric patients with DD and/or epilepsy. The diagnosis of patients with WDR45 mutations enabled the precise genetic counseling of parents for subsequent children.

1275W

Whole exome sequencing in a patient with congenital developmental abnormalities reveals a variant within PDGFA, a gene of uncertain significance as likely causal. A. Gupta, E. Conboy, P. Pichurin, E.W. Klee. Mayo Clinic, Rochester, MN.

Statement of purpose: We report a newborn female with multiple congenital anomalies including cloacal anomalies (abnormality of the vagina, uterus didelphys, unilateral renal agenesis, anal atresia, imperforate anus), cerebellar hypoplasia, atrial septal defect, patent ductus arteriosus and bilateral talipes equinovarus. After birth she was noted to have urinary retention, abdominal distention, unilateral facial palsy and subjectively hypertelorism. She had a healthy full sibling and no affected family members. **Methods:** Genomic DNA was extracted from whole blood of the proband and parents. Whole Exome Sequencing (WES) trio test was carried out through GeneDx, the XomeDx test which included evaluation of variants that are identified to be de novo, compound heterozygous, homozygous, heterozygous and X-linked in addition to relevant analysis based on the family structure and reported phenotype.

Results: WES revealed a heterozygous de novo missense Variant of Uncertain Significance- c.443G>T p.R148L within *PDGFA*, designated a Gene of Uncertain Significance with potential relationship to the phenotype of the patient.

Discussion: This gene encodes a member of the protein family comprised of both platelet-derived growth factors (PDGF) and vascular endothelial growth factors (VEGF) and binds and activates PDGF receptor tyrosine kinases, which play a role in a wide range of developmental processes. Although no pathogenic variants in *PDGFA* have been reported in association with a specific human disease to date, abnormal PDGF signaling has been implicated in a range of diseases such as cancer, atherosclerosis, restenosis, pulmonary hypertension, and retinal diseases, as well as in fibrotic diseases, including pulmonary fibrosis, liver cirrhosis, scleroderma, glomerulosclerosis, and cardiac fibrosis. Homozygous *Pdgfa*-null mice have been shown to die at the prenatal stage or postnatal stage before age 6 weeks and develop lung emphysema secondary to failure of alveolar septation, reduced numbers of oligodendrocytes and a dysmyelinating phenotype, abnormal Leydig cell development and spermatogenic arrest, and abnormalities and reduction in the size of most other organs. The evidence from animal knockout models of *PDGFA* and role of PDGF signaling in various developmental pathways seem to suggest the variant within *PDGFA* to be potentially responsible to the patient's phenotype of congenital developmental malformations.

1276T

Diagnostic dilemmas in the next-generation sequencing era: Novel Mendelian phenotype, blended phenotype or phenotypic expansion. *M. Muriello¹, G. McCarrick¹, C. Smith¹, H. Levy^{1,2}, D. Valle¹, H. Dietz¹, N. Sobreira¹.* 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD; 2) Johns Hopkins University, School of Medicine, Baltimore, MD.

Whole exome sequencing (WES) has been the main tool for the discovery of novel disease genes related to rare Mendelian phenotypes in clinic and research. Mostly by using WES, the number of genes with phenotype-causing variants described in OMIM went from 2,346 in 2009 to 3,917 in June 7th 2018. Still, approximately 80% of the 19,000 known genes are not related to a disease phenotype and 3,352 Mendelian or suspected to be Mendelian phenotypes do not have a known molecular basis (OMIM, June 7th, 2018). So, it is not surprising that more than 50% of the clinical WES are non-diagnostic. But, most unsolved WES uncover multiple candidate variants in known disease genes that could be related to a phenotypic expansion or to a blended phenotype. Chong et al. (2015) described that among the 617 novel discoveries identified by the Centers for Mendelian Genomics, 363 were related to new phenotypes and 198 were related to phenotypic expansions. Posey et al. (2017) described that 4.9% of 2076 cases submitted to WES where blended phenotypes. Here we describe 4 unrelated probands in whom we performed WES and did not identify the causative variant(s) in a known disease gene. Each proband was an isolated case with a unique multiple congenital anomalies phenotype without a clinical diagnosis. Briefly, Case 1 is a 10 year old male with gingival hyperplasia, dysmorphic features, and dysplastic phalanges; Case 2 is a 12 year old male with hypoplastic left heart syndrome, developmental delay and eosinophilic esophagitis; Case 3 is a 14 year old female with aortic root dilation, scoliosis, Chiari malformation, and radioulnar synostosis; and Case 4 is a 43 year old female with microcephaly, Pierre-Robin sequence, facial dysmorphisms, and congenital ocular defects. As part of the Baylor-Hopkins Center for Mendelian Genomics, we performed WES of the proband and their parents (father of case 1 was not available for WES). Using PhenoDB analysis tool, we identified the best candidate variants in each proband but no causative variant(s) in a known disease gene related to their phenotype was identified. In all of them, a small list of candidate variants in novel and known disease genes was created and the possible novel disease genes were submitted to GeneMatcher, PhenomeCentral, MyGene2 and *matchbox*. In each proband we excluded, to the best of our ability, the possibility of phenotypic expansion and blended phenotype and suggest that they each, have a novel rare Mendelian phenotype.

1277F

Effective use of off-target reads in evaluation of copy number variation from aligned exome data: Validation study in 325 undiagnosed patients. *H. Suzuki¹, T. Uehara¹, H. Yoshihashi², H. Fukushima³, R. Sumazaki³, Y. Yamaguchi¹, T. Takenouchi^{1,4}, R. Kosaki⁵, K. Kosaki¹.* 1) Center for Medical Genetics, Keio University School of Medicine, Tokyo, Japan; 2) Department of Medical Genetics, Tokyo Metropolitan Children's Medical Center, Tokyo, Japan; 3) Department of Child Health, Faculty of Medicine, University of Tsukuba, Ibaraki, Japan; 4) Department of Pediatrics, Keio University School of Medicine, Tokyo, Japan; 5) Divisions of Medical Genetics, National Center for Child Health and Development, Tokyo, Japan.

[Back ground] Next generation sequencing is used to detect single nucleotide variants and short insertions and deletions on a genome-wide scale. Currently, large copy number variations [CNV] are screened by microarray. There have been several attempts to detect such large CNV from aligned short reads from exome sequencing by evaluating read-depth information in regions targeted by the bait (i.e. on-target region) that mainly captures the coding regions. Those programs basically neglect off-target regions of the genome. Recently, a revolutionary algorithm EXCAVATOR2 which exploits read-depth information of off-targeted regions concurrently with on-targeted regions was proposed using simulated genome data. Here we validated performance of EXCAVATOR2 using real patient data. [Materials and Methods] Patients with multisystem disorders who were suspected to have genetic diseases were ascertained through the Japanese genome project on rare and undiagnosed diseases. These patients were analyzed both by 1) comparative genomic hybridization, 2) standard exome analysis followed by EXCAVATOR2. Two methods of analysis were performed by two researchers who were blinded for each other's results. In both analysis, the threshold was set as 1 Mb for detection of both gains and losses. Results from comparative genomic hybridization was considered as gold standard. Potential CNVs detected by EXCAVATOR2 only within the off-target regions were discarded from further analysis. [Results] Initial analysis of a cohort of 93 patients with array CGH detected 7 patients with CNV (1.3-9.7 Mb). Analysis of the same cohort with EXCAVATOR2 detected all these 7 patients with "proven" CNVs. Hence, sensitivity and specificity of EXCAVATOR2 compared with array CGH was 100% and 100%, respectively. In another cohort of 232 patients with standard exome analysis, 19 patients had CNVs (1.3-14.7 Mb) which detected further analyzed by EXCAVATOR2. All these 19 patients were confirmed by using array CGH. [Discussion] We demonstrated analytic validity of EXCAVATOR2 in clinical settings involving 325 undiagnosed patients. Effectiveness of the software EXCAVATOR2 that use both off-target and on-target reads was unequivocal in evaluation of copy number variation from aligned exome data. Hence, exploitation of exome data allows detection of both base-level mutations and mega base-level structural changes in a single assay rather than tiered approach with microarray and NGS.

1278W

Singleton exome strategies in under-represented populations. *E. Ullah¹, B. Guan¹, CE. Prada^{2,3}, RB. Hufnagle¹.* 1) Ophthalmic Genetics and Visual Function Branch, National Eye Institute, Bethesda, MD; 2) UC Department of Pediatrics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 3) Hospital Internacional de Colombia, Santander, Colombia.

Accurate molecular diagnostics are critical to care and management for patients with Mendelian genetic disorders. Recent advances in sequencing technologies have both revolutionized the pace of new gene discovery and increased the probability of genetic diagnosis. Whole-exome sequencing (WES) is reported to identify the disease-causing mutations in approximately 25-40 % of cases of monogenic disease. Here, we use WES as the first approach to genetic diagnosis in previously untested patients from an under-represented population in healthy population databases (Colombia, South America). We recruited 128 singletons with a variety of genetic disorders. About 50% cases had multisystem presentation while in the remaining cases the genetic disorder affected mainly neuromuscular (12%), muscular (7%), neural (6%), metabolic (5%), cardiac (5%), skeletal (4%), immune (4%), eye (2%), renal (2%), reproductive (2%) and dermatologic (1%) tissues. Family history was known for 12% of cases whereas 88% were sporadic. We employed a case vs control approach where one patient served as a case and the remaining as controls (1:N-1 strategy), and standard variant filtering strategies. This analysis solved 41 % of the cases in total, identifying a previously reported mutation or a rare coding variant matching the gene-associated phenotype and inheritance pattern. One reported truncating allele NM_001195798.1:c.11G>A, NP_000518.1:p.W4* in the *LDLR* gene was identified in three unrelated patients with familial hypercholesterolemia, indicating a likely founder allele. Subsequent analysis of solved cases revealed that positive family history and phenotypes predominantly affecting a single tissue were more likely to be solved by this approach. Additionally, this cohort permits for curation of common variants in the Colombian population for further studies. Semi-automation of variant prioritization process and calculation of population variant frequencies is expected to further increase the efficiency and reproducibility of this approach. We conclude that WES as the first approach to genetic diagnosis and use of a large cohort for case-control analysis is effective for rapid identification of disease-causing mutations. This approach can also be applied to cohorts where control data is not available, and identify common founder alleles, improving the cost-effectiveness of molecular diagnostics in these settings.

1279T

Treatment of the FXTAS and FXPOI mouse model with the antioxidant tempol. *I. Gazy¹, M.S. Lange², E. Musselwhite³, X. Zhao¹, G. Hoffman², J. Johnson³, K. Usdin¹.* 1) NIDDK, National Institutes of Health, Bethesda, MD, 20892, USA; 2) Department of Biology, Morgan State University, Baltimore, MD 21251, USA; 3) Department of Obstetrics and Gynecology, University of Colorado-Denver, Aurora, CO 80045, USA.

The human *FMR1* gene contains a CGG repeat tract in the 5'untranslated region. Premutation (PM) alleles have 55-200 CGG repeats and carriers of such PM are at risk of developing two distinct Fragile X-related disorders: FX-associated tremor/ataxia syndrome (FXTAS) and FX-associated primary ovarian insufficiency (FXPOI). FXTAS is a late-onset neurodegenerative disorder characterized by intention tremor, cerebellar ataxia, parkinsonism and cognitive decline. While FXTAS is more prevalent in male PM carriers than female carriers, females are also at risk of developing the ovarian dysfunction disorder FXPOI. Previous work on human carriers as well as on a PM knock-in mouse model that recapitulates many of the pathologies observed in PM carriers, have implicated mitochondrial dysfunction as well as oxidative damage in the pathogenesis of FXTAS and FXPOI. We envision that oxidative damage can potentially contribute to the pathology of these disorders in 2 different ways. First, oxidative stress can be deleterious to the normal function of the brain and ovaries by damaging the biomolecules necessary for normal cell function. Additionally, oxidative damage is thought to promote CGG repeat expansions and this could potentially contribute to pathology by enhancing the toxic effects of CGG containing transcripts. Antioxidants have been proposed to be generally beneficial in cases of premature ovarian failure or ovarian damage unrelated to the FX PM. Furthermore, antioxidants are currently being tested as treatments for neurodegenerative disorders as well as for Fragile X syndrome. We, therefore, decided to test the effect of tempol, an antioxidant that crosses the blood brain barrier, on repeat instability, brain pathology as well as ovarian function in PM mice. Ultimately, this work should help assess whether antioxidants have therapeutic potential for these disorders.

1280F

Chronic atrial and intestinal dysrhythmia syndrome: Expanding the phenotype of *SGOL1* gene mutations. C. Baquero Montoya^{1,13}, C. Ortiz Piedrahita¹, R. De Vivero Camacho², DP. Sanchez Hernández¹, A. Becerra Romero^{3,4}, MM. Contreras Ramirez¹, JA. Rivera Echeverry¹, M. Corrales³, G. Lenis³, A. Guerra Solarte³, J. Buitrago³, H. Cala³, G. Osorio⁵, JC. Pérez^{6,13}, SC. Mesa Restrepo¹, AI. Henao¹, AC. Sierra Montoya¹, P. Ruiz Navas¹, A. Santamaria Escobar¹, AM. Correa Varela¹, F. Restrepo¹, S. Molina Restrepo³, SI. Alzate Vanegas³, PA. Cañaveral Londoño¹, OL. Rincón¹⁰, MA. Cano Romero¹⁰, AM. Bedoya¹¹, MC. Gil Rodríguez¹², CM. Alvarez⁵, B. Aristizabal¹⁰. 1) Department of Paediatrics, Hospital Pablo Tobón Uribe, Medellín, Colombia; 2) Department of Paediatrics, Hospital Infantil Napoleón Franco Pareja, Cartagena, Colombia; 3) Trasplant Surgery, Hospital San Vicente Fundación, Rionegro, Colombia; 4) Grupo de Inmunología Celular e Inmunogenética, Universidad de Antioquia, Medellín, Colombia; 5) Grupo de Gastrohepatología, Facultad de Medicina, Universidad de Antioquia, Medellín, Colombia; 6) Department of Pathology, Hospital Pablo Tobón Uribe, Medellín, Colombia; 7) Link Diagnostico Digital, Medellín, Colombia; 8) Department of Nutrition, Hospital Pablo Tobón Uribe, Medellín, Colombia; 9) Psychology, Hospital Pablo Tobón Uribe, Medellín, Colombia; 10) Molecular Biology, Laboratory, Hospital Pablo Tobón Uribe, Medellín, Colombia; 11) Laboratory, Hospital Pablo Tobón Uribe, Medellín, Colombia; 12) Department of Pharmacology- Physiology, Medical School, University of Zaragoza, Spain; 13) Dinamica IPS, Medellín, Colombia.

Introduction: A novel cohesinopathy caused by a founder mutation in *SGOL1* termed Chronic Atrial and Intestinal Dysrhythmia Syndrome (CAID) was described in 16 French Canadians and 1 Swede. This disorder is characterized by cardiac and gastrointestinal involvement. So far no neuropsychiatric symptoms have been reported, which is remarkable taking into account that the brain is the organ most sensitive to cohesin disruption. The population genetic analysis of this disorder determined that the rare disease-causing haplotype was of northern European origin. Here, we report on the first cases of CAID syndrome due to c.67A>G mutation in *SGOL1* in Latin America, which showed the combination of sick sinus syndrome and chronic intestinal pseudo-obstruction plus neurological involvement. **Materials and Methods:** Patient 1, is a 31 years old Colombian male, who since the age of 13 began with vomiting, abdominal distention and intestinal dysmotility leading to intestinal failure. At the age of 28 years old, he was diagnosed with sick sinus syndrome. Basic imagenologic studies including resonance enterography as well as intestinal biopsy with electron microscopy were carried out. Subsequently, Holter test and molecular analysis were performed. Patient 2, is a 13 years old Colombian boy, who a year ago presents with abdominal pain, vomiting, intestinal pseudo obstruction, bradycardia and behaviour impairment. Imagenological studies including brain MRI, intestinal biopsy and ecocardiography were performed. Finally, whole exome sequencing was done. **Results:** Both initial imagenological studies demonstrated delay in intestinal transit with loop dilatation; the resonance enterography suggested the presence of chronic intestinal pseudo occlusion syndrome. In patient 1, Holter test showed cardiac sinus dysfunction and in patient 2, ecocardiography displayed systolic dysfunction of left ventricle. The intestinal microscopic findings included fibrosis of muscular layer. Sanger sequencing and whole exome sequencing demonstrated a homozygous mutation in *SGOL1* (c.67A>G; p.Lys23Glu). **Conclusion:** The description of these cases confirmed the presence of patients with CAID syndrome in Latin America, probably a second cluster. The finding of neuropsychiatric involvement could expand the phenotype, what would be expected knowing that central nervous system is affected in all cohesinopathies previously described; or might be a variant of the clinical presentation in Latin America. .

1281W

Addressing missing heritability in rare genetic diseases. M. Tarailo-Graovac^{1,2}, B.I. Drögemöller^{3,4}, P.A. Richmond⁵, A.B.P van Kuilenburg⁶, G. Horvath^{4,7}, C.J Ross^{3,4}, C.D. van Karnebeek^{5,6}, W.W. Wasserman^{3,5}. 1) Departments of Biochemistry, Molecular Biology and Medical Genetics, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada; 2) Alberta Children's Hospital Research Institute, University of Calgary, Calgary, AB, Canada; 3) Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver, BC, Canada; 4) BC Children's Hospital Research Institute, Vancouver, BC, Canada; 5) Departments of Medical Genetics and Pediatrics, Centre for Molecular Medicine and Therapeutics, Vancouver, BC, Canada; 6) Departments of Clinical Chemistry, Pediatrics and Clinical Genetics; Laboratory Genetic Metabolic Diseases, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands; 7) Department of Pediatrics, Division of Biochemical Diseases, University of British Columbia.

Rare genetic diseases have low prevalence individually, but collectively are common. Accurate diagnosis is crucial as it enables counselling; appropriate/ tailored clinical management as well as identification of targets for treatment. However, diagnosing rare diseases poses a challenge, and despite extensive, expensive clinical tests, low diagnostic rates, misdiagnoses, and delays to appropriate care are not uncommon. In recent years, with the advent of high throughput sequencing (HTPS), we have witnessed an important step forward in diagnosis of rare disorders (from ~10% using traditional to ~40% using genome-wide genetic testing). Simultaneous profiling of a subset of (i.e. using gene panels) or the majority of protein-coding regions (i.e. using whole exome sequencing, WES) in an individual has reduced the dependency of diagnosis on traditional approaches, and has allowed hypothesis-guided prioritization based on both exome and phenome information. Nonetheless, even with those advances establishing a diagnosis is still a challenge and in most phenotypic categories more than 50% of rare disease patients remains undiagnosed. A plausible explanation for the missing heritability in rare suspected genetic diseases may be in the preferred technology used so far. Exome sequencing approaches capture only a small (<2%) portion of the genome and are very limited in the capacity to detect structural variants (SVs), copy number variants (CNVs), insertions, translocations, inversions, tandem repeat expansions, as well as non-coding and deep-intronic variants. Here we report on our cohort of rare disease patients who underwent deep clinical phenotyping but remained undiagnosed despite extensive tests including exome sequencing. Our expanded genome-wide analyses revealed the following source of missing heritability: (1) non-uniform capture of targeted exomes; (2) spectrum of pathogenic structural variation; (3) somatic mosaicism; (4) phenocopy (i.e. known well-defined clinical phenotype caused by novel, unsuspected gene); and (5) complex composite effects of multi-gene defects. Our data shows that whole genome sequencing (WGS) substantially improves diagnostic/discovery potential of HTPS, and advocates the importance of utilizing the WGS for its potential to detect "all" classes of genetic variation in a single test.

1282T

The prevalence of suspected pathogenic mutations for early and adult onset dominant disorders in a healthy Israeli population. S. Tzur^{1,2}, N. Ruhrman Shahar³, L. Bazak⁴, T. Talmy¹, N. Mizrahi¹, S. Kovo¹, R. Attali¹, H. Ostrer⁴, L. Basel-Salmon^{3,5,6,7}. 1) Emedgene Technologies, Tel Aviv, Israel; 2) Laboratory of Molecular Medicine, Rambam Health Care Campus, Haifa, Israel; 3) Raphael Recanati Genetic Institute, Rabin Medical Center Beilinson Campus, Petah Tikva, Israel; 4) Department of Pathology, Albert Einstein College of Medicine, Bronx, NY, USA; 5) Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel; 6) Pediatric Genetics Unit, Schneider Children's Medical Center of Israel, Petah Tikva, Israel; 7) Felsenstein Medical Research Center, Rabin Medical Center, Petah Tikva, Israel.

Introduction Evaluating the frequency of suspected pathogenic mutations for early and adult onset dominant disorders among healthy individuals has two main implications: (1) for adult onset disorders, we can learn about the prevalence of such variants in asymptomatic individuals and about disease penetrance; (2) for early onset severe dominant disorders, such analyses of healthy adults could contribute to better understanding of the limitations in clinical predictive genomics in the context of prenatal genome sequencing. **Methods** We analyzed the exomes of 166 apparently healthy subjects, while applying stringent criteria for pathogenicity including only rare, known pathogenic or loss of function variants in all genes implicated in dominant conditions. For each variant, we assessed risk and predicted effect based on the available literature and bioinformatic tools. **Results** For diseases of adult onset, 20 suspected variants were found in 18 individuals (11%), predominantly in genes associated with cardiac, metabolic, and oncologic conditions. For early onset diseases, upon first analysis, 7 individuals (4%) presented suspected pathogenic mutations. However, interpretation of these variants proved challenging since in large part we identified mutations in genes that are: more likely to cause a dominant disease exclusively with missense variants, cause more severe disease when found in recessive forms, cause relatively mild disorders, have old ClinVar submissions, or have a relatively high allele count in gnomAD. Finally, after deep investigation, none of the mutations was considered to be unambiguously pathogenic. **Conclusions** For adult onset dominant diseases, we found a high prevalence (11%) of suspected pathogenic mutations with potential clinical implications. We may recommend a follow up study for these individuals including medical assessment, genetic counseling and surveillance. For early onset dominant diseases, preliminary analysis revealed suspected findings in 4% of the cohort. However, a deep investigation revealed that none were unambiguously pathogenic or harmful. These results have serious ramifications pertaining to prenatal diagnostics, since the initial analysis wrongly demonstrated a high rate of individuals with early onset disorders. This may prompt errant clinical decision making in prenatal diagnosis if not performed with caution. This study emphasizes the complexity and precautionary measures involved in predictive genomics.

1283F

Biallelic variants in CBL in brothers with intellectual disability and Noonan-like syndrome. G. Gotway^{1,2}, C. Hornbuckle². 1) University of Texas Southwestern Medical Center, Dallas, TX; 2) Children's Medical Center Dallas.

Two full brothers 18 months apart in age presented with developmental delays and hyperpigmented macules and moles. The older brother (A) walked at 18 months and did not speak until after age 2; the younger brother (B) did not sit until a year of age, walked after age 2 and did not speak until age 4. Both were in life skills classes in school and cannot live independently as adults. Patient A had bilateral retinal detachment at age 15 and was diagnosed 3 months later with a grade 1 cerebellar pilocytic astrocytoma. At that time, he was noted to have several café-au-lait (CAL) macules and numerous darker pigmented macules and moles. Patient B had similar lesions, and their mother and several of her relatives carried a clinical diagnosis of neurofibromatosis. *NF1* sequencing was unremarkable. A sequencing panel for Noonan syndrome revealed two variants in the *CBL* gene in both brothers; one (c.1099C>A; p.Gln367Lys) was rated likely pathogenic and the other (c.1129A>G; p.Thr377Ala) was rated a variant of uncertain significance. The LP variant was found in the mother and maternal great grandmother, both of whom have CAL macules but no moles. The mother had learning disabilities but lives independently, and has a lipomatous tumor on her hip. The maternal grandmother, an obligate carrier, has a diagnosis of Coat's disease and CAL macules. The mother has a brother and maternal half-brother in special education who were not tested. The father was found to carry the VUS. He also had learning problems in school, but does not have a history of CAL macules; he is adopted, and does not know his family history. Both variants occur in exon 8, which contains the linker domain of the CBL protein and is a hotspot for pathogenic changes. Both variants are absent from control databases (gnomad, ESP, 1000 genomes). The Glu367 position has been found to be altered in two other patients, one with a Noonan-syndrome like phenotype and one with chronic myelomonocytic leukemia. The Thr377 position has not been reported to be altered in any other germline cases, but the same alteration was seen in a seminoma, and a different missense at this site was reported in a study of somatic mutations associated with hematopoietic clonal expansion. The more severe neurologic impairment and skin findings in the brothers raises the possibility that this is a case of biallelic pathogenic variants, one likely being a milder hypomorphic allele.

1284W

Genotype-phenotype correlation of mutations in *PIGN* causing Fryns syndrome, MCAHS and neurological phenotypes. U. Kini¹, A. Tardivo¹, A. Pagnamenta², J. Taylor², T. Kinoshita³, Y. Murakami³, DDD Study. 1) Oxford Centre for Genomic Medicine, Oxford University Hospitals NHS FT, Oxford, Oxfordshire, United Kingdom; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford; 3) abumoto Department of Intractable Disease Research, Research Institute for Microbial Diseases, Osaka University, Osaka, Japan.

About 30 genes encoding the GPI anchor pathway are responsible for the post-translational modification of approximately 150 proteins within the cell. While genes in the early part of the pathway have been mostly reported to cause a pure neurological phenotype (intellectual disability +/- seizures), genes in the later part can have additional findings of congenital malformations including Hirschsprung disease, diaphragmatic hernia, genito-urinary abnormalities and cleft palate. Recessive mutations in *PIGN* have been reported to cause intellectual disability, seizures, Fryns syndrome and MCAHS (multiple congenital abnormalities, hypotonia, seizures). No genotype-phenotype correlation has been established as yet. **Methods:** We studied 38 patients (10 new and 28 from the literature) with recessive mutations in *PIGN*. The 10 new patients were ascertained by exome sequencing either in a clinical setting or via the DDD study. FACS analysis and expression studies were carried out to prove pathogenicity in appropriate cases. **Results:** *PIGN* consists of 2 pfam domains (phosphodiesterase and PigN). The PigN domain includes most of the 14 transmembrane (TM) domains except for the first and the last TM domains. 11/38 (29%) were reported to have lethal Fryns syndrome and showed biallelic severe loss of function (LOF) mutations (intragenic deletions, frameshifts and stop mutations). All mutations were present in the phosphodiesterase or PigN domains; the exception was one patient with homozygous missense mutations in the TM domain. 16/38 (42%) patients had a MCAHS phenotype; all patients had at least one mild missense mutation and at least one of the mutations was in the PigN domain. The remaining 11/38 (29%) had a pure neurological phenotype. Both mild missense and severe LOF mutations were reported in these cases but were restricted to the phosphodiesterase domain. **Conclusion:** Biallelic severe LOF mutations lying in any domain of *PIGN* results in lethal Fryns syndrome. The presence of at least one mutation in the PigN domain results in the MCAHS phenotype, while mutations restricted to the phosphodiesterase domain cause a pure neurological phenotype.

1285T

When two is worse than one: Machine-learning-based identification of digenic disease genes based on network and genomic data. S. Mukherjee^{1,4}, J.A. Capra^{1,3,4,5}, J. Meiler^{2,4,5,6}. 1) Department of Biological Sciences, Vanderbilt University, Nashville, TN; 2) Department of Chemistry, Vanderbilt University, Nashville, TN; 3) Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN; 4) Center for Structural Biology, Vanderbilt University, Nashville, TN; 5) Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN; 6) Department of Pharmacology, Vanderbilt University School of Medicine, Nashville, TN.

Identifying the genetic causes of rare diseases can significantly enhance treatment options, and the integration of whole exome and genome sequencing into clinical diagnosis provides a comprehensive perspective on a patient's mutations. However, identifying genetic variants that contribute to disease among thousands of mutations in a patient's genome remains challenging. Recognizing this, the NIH has established the Undiagnosed Diseases Network (UDN) with the goal of diagnosing patients with rare, complex, and unresolved diseases by combining genome sequencing and genetic analyses with the latest tools for identifying the effects of individual variants and their potential contributions to rare disease. Despite substantial efforts, more than half of UDN cases remain unsolved. Synthetic lethality studies in model systems have highlighted the importance of gene-gene interactions to the genetics of diverse phenotypes. Furthermore, hundreds of human digenic diseases are known, where variants in a pair of genes act synergistically to cause disease phenotypes, even when mutations in the genes individually are benign. However, there is no high-throughput method for identification of gene pairs with the potential to cause digenic disease when mutated in patients. We hypothesized that digenic gene pairs could be distinguished from non-digenic pairs based on their functional, network, and genomic properties, such as shared biochemical pathways, shared phenotypes, coexpression, network proximity, and haploinsufficiency. We trained a random forest (RF) machine learning classifier using a positive set of clinically curated gene pairs associated with digenic diseases from the Digenic Diseases dAtabase (DIDA). Using the attributes described above as features, we evaluated the ability of the classifier to distinguish known digenic gene pairs from negative sets of gene pairs in which simultaneous mutations were not associated with disease. The RF classifier was able to accurately distinguish the digenic disease pairs from the negative sets; it achieved an average area under the curve (AUC) of 0.9 in the Receiver Operating Characteristic (ROC) curve and 0.5 in the Precision-Recall (PR) curve in 10-fold cross validation. Our work demonstrates that digenic gene pairs can be accurately identified by their functional and network attributes. We are now using this pipeline to find potentially causal digenic gene pairs in patients with unsolved rare diseases from the UDN.

1286F

MCM3 and MCM7: New causative genes underlying Meier-Gorlin syndrome. K.M. Knapp¹, F.L. Harms², C.W. Ockeloen³, E. Bongers³, A.P. Jackson⁴, N. Pachter⁵, J. Denecke⁶, K. Kutsche², L.S. Bicknell¹. 1) Department of Pathology, Dunedin School of Medicine, University of Otago, Dunedin, New Zealand; 2) Institute of Human Genetics, University Medical Center Hamburg-Eppendorf, Hamburg, Germany; 3) Department of Human Genetics, Radboud University Medical Center, Nijmegen, Netherlands; 4) MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK; 5) Genetic Services of Western Australia, King Edward Memorial Hospital, Subiaco, Western Australia, Australia; 6) Department of Pediatrics, University Medical Center Hamburg-Eppendorf, Hamburg, Germany.

Meier-Gorlin syndrome (MGORS) is a rare genetic disorder characterized by primordial dwarfism, microtia, and patellar aplasia/hypoplasia, which is caused by impaired DNA replication initiation. Eight causative genes encoding essential components of the pre-replication (pre-RC) and pre-initiation complexes (pre-IC) have been identified to date. By performing Chromium whole-genome sequencing on a MGORS cohort we have identified biallelic variants in two novel disease genes not previously associated with MGORS: *MCM3* and *MCM7*. These genes encode subunits of the replicative helicase complex MCM2-7, which represents a component of the pre-RC complex. One individual is homozygous for a missense variant in *MCM3*, and structural modelling indicates that substitution of this highly conserved residue would disrupt interactions with ORC2 and CDC6. For *MCM7*, we have identified two unrelated individuals with biallelic pathogenic variants. Both individuals are compound heterozygous for a different truncating allele (predicted to lead to nonsense-mediate mRNA decay) *in trans* with a different missense variant, which based on the essential nature of MCM subunits, is likely a hypomorphic allele. The affected residues are strongly conserved, with one residue being located within the core of MCM7 and the other within a region identified as required for MCM3 and ATP binding. Based on these structural implications, we hypothesise that the *MCM3* and *MCM7* variants identified disrupt pre-RC chromatin loading, reducing proliferative potential and thereby cause Meier-Gorlin syndrome. There is also some phenotypic variation amongst these individuals, and their clinical features will be discussed in context with the previously described individual with biallelic *MCM5* variants, as well as the wider mutation-positive MGORS cohort.

1287W

CysteinyI-tRNA synthetase (CARS) mutations cause a multi-system, recessive disease phenotype that includes tissues dependent on cysteine. M.E. Kuo^{1,2}, W. Introne³, M.C. Malicdan³, W.A. Gahl³, A. Antonellis^{1,4,5}. 1) Cellular and Molecular Biology Program, University of Michigan, Ann Arbor, MI; 2) Medical Scientist Training Program, University of Michigan, Ann Arbor, MI; 3) NIH, Undiagnosed Diseases Program and Office of the Clinical Director, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 4) Department of Human Genetics, University of Michigan, Ann Arbor, MI; 5) Department of Neurology, University of Michigan Medical School, Ann Arbor, MI.

Aminoacyl-tRNA synthetases (ARSs) are a class of highly specialized, essential enzymes responsible for performing a crucial step of protein translation: charging tRNA molecules with cognate amino acids. Consistent with the essential function and ubiquitous expression of these enzymes, mutations in 32 of the 37 ARS-encoding loci have been implicated in severe, early-onset recessive phenotypes involving many different tissues. Current evidence based on genetic and functional studies suggest a loss-of-function mechanism; however, our understanding of the allelic and locus heterogeneity of ARS-related disease is incomplete. To date, cysteinyI-tRNA synthetase (*CARS*) variants have not been implicated in any human disease phenotype. In collaboration with the NIH Undiagnosed Diseases Program, we identified a patient with recessive microcephaly, developmental delay, and brittle hair and nails who is homozygous for p.Ser688Glnfs*2 *CARS*. Based on the recessive phenotype and the phenotypic overlap between this patient and patients with recessive diseases associated with other ARS enzymes, we hypothesized that p.Ser688Glnfs*2 *CARS* causes a loss-of-function effect that is responsible for the phenotype of our patient. Interestingly, *CARS* encodes two protein isoforms that harbor different C-terminal amino-acid sequences, and the p.Ser688Gln*2 *CARS* variant is predicted to truncate the protein upstream of both C-termini. Here, we present unpublished clinical, genetic, and functional studies that support the pathogenicity of p.Ser688Glnfs*2 *CARS*. Protein expression studies on patient fibroblasts revealed that a single, truncated protein isoform is expressed and stable. An *in vivo* yeast complementation assay using the human *CARS* gene revealed that p.Ser688Glnfs*2 *CARS* causes severely reduced yeast cell growth compared to wild-type *CARS*, which is consistent with a loss-of-function allele and the recessive phenotype. In sum, our findings for the first time implicate a loss-of-function *CARS* allele in human inherited disease; this notion is supported by the identification of two additional patients with *CARS* mutations by a collaborating group. Interestingly, *CARS* is the first ARS associated with brittle hair and nails, which likely reflects the high cysteine content in human keratins. As such, this study improves our understanding of the clinical and locus heterogeneity of ARS-related human disease phenotypes.

1288T

Severe PI3Kinase overgrowth syndrome treated with the AKT inhibitor miransertib. A.D. Irvine^{1,2,3}, O. Kubassova⁵, M. Hinton⁵, B. Schwartz⁴, J. Kazakin⁴, V. Dvorakova^{1,3}. 1) Dermatology, Our Lady's Children's Hospital, Dublin, Leinster, Ireland; 2) Clinical Medicine, Trinity College Dublin, Dublin 2; 3) National Childrens Research Centre, OLCHC, Dublin 12, IRELAND; 4) ArQule, Inc., Burlington, MA, USA; 5) IAG, Image Analysis Group, Philadelphia, USA; 6) IAG, Image Analysis Group, London, UK.

Somatic activating mutations in phosphatidylinositol 3-kinase (PI3K) and serine-threonine protein kinase (AKT1) have been shown to cause abnormal segmental overgrowth phenotypes. Miransertib (ARQ 092) is a novel highly selective oral AKT-inhibitor with a proven in-vitro efficacy. Its potential as a therapeutic option for overgrowth patients who have exhausted conventional treatment methods was investigated through the assessment of the longitudinal MRI data acquired from one patient between 12/2015 and 04/2018 with the treatment started on 06/2016. A term baby girl had a large posterior thoracic wall lipoma and large feet at birth. Progressive lipomatous overgrowth and leg length discrepancy were treated with repeated surgical procedures during her first ten years. Progressive replacement of normal tissues with fatty overgrowth resulted in severe subcutaneous, intra-abdominal and paraspinal involvement causing obstructive uropathy with bladder displacement and neuropathic bladder dysfunction, impaired mobility and dysfunctional sitting and lying posture. At 13yrs the patient was wheelchair bound and a suprapubic catheter was required to relieve obstructive uropathy. Sirolimus was commenced age 15 but discontinued after 12 months due to disease progression. The patient was unable to lie flat due to respiratory compromise caused by extreme diaphragm elevation by intraabdominal lipomatous overgrowth with flank pain caused by massive lipomatous distension. Surgical remedy was deemed impossible due to the infiltrative nature of the fatty overgrowth incasing vital structures in the patient's retroperitoneum. Cellular studies indicated high levels of AKT activation in affected fibroblasts. ARQ 092 was commenced on compassionate basis at a dose of 30 mg OD. The patient experienced improvement in her respiratory compromise and flank pain and was able to lie flat at night. Radiologist analysis of serial volumetric DIXON MRI analysis defined regions of interest on slices of the in-phase images to define areas of fatty overgrowth (using Dynamika software as a platform for imaging data analysis). Calculated volumes of fatty overgrowth declined by approximately 15% after treatment commenced. The patient has now completed 22 months of therapy with clinically stable disease and clear radiological improvement. The drug was well tolerated with no significant toxicities other than hyperlipidaemia comparable to lipid profile derangement previously noted on sirolimus therapy.

1289F

Mutations in mTOR and MAPK pathways cause heterogeneous lymphatic anomalies. D. Li¹, M.E. March¹, T.W. Wenger², C. Seiler³, A. Gutierrez-Uzquiza¹, C. Kao¹, E. Pinto⁴, E.J. Bhoji¹, J. Strausbaugh⁵, J.B. Belasco^{5,6}, Y. Dor^{4,5}, H. Hakonarson^{1,5,7}. 1) Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Division of Craniofacial Medicine, Seattle Children's Hospital, Seattle, WA; 3) Zebrafish Core Facility, The Children's Hospital of Philadelphia Research Institute, Philadelphia, PA; 4) Center for Lymphatic Imaging and Interventions, The Children's Hospital of Philadelphia, Philadelphia, PA; 5) Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, PA; 6) Division of Oncology, The Children's Hospital of Philadelphia, Philadelphia, PA; 7) Divisions of Human Genetics and Pulmonary Medicine, The Children's Hospital of Philadelphia, Philadelphia, PA.

Lymphatic anomalies are thought to arise through defects that occur during lymphatic development, which include a variety of diagnoses: lymphangiectasia, central conducting lymphatic anomaly (CCLA), generalized lymphatic anomaly (GLA), congenital chylothorax, and lymphedema. Although recent studies have demonstrated benefit of sirolimus in the treatment of GLA, the absence of clear clinical distinctions between these entities, due to their rarity and lack of accepted standardized nomenclature, has hampered development of innovative therapies. We report exome sequencing results in 35 independent patients, including a four-generation pedigree structure consistent with autosomal dominant inheritance. Seven pathogenic variants are identified in seven families in four genes encoding components of RAS/MAPK signaling (*ARAF*, *BRAF*, *KRAS*, and *PTPN11*) and a receptor tyrosine kinase (*EPHB4*). Splicing mutation in *EPHB4* (c.2334+1G>C), leading to use a cryptic splice donor with retention of 12-bp intron sequence resulting in reduced phosphorylation and loss of function effect, was identified in the four-generation pedigree with proband diagnosed with CCLA. The family history is notable for edema of low extremities, and variable chylous effusion. Zebrafish *ephb4a* morpholino resulted in lymphatic vessel misbranching and developmental deformities effectively rescued by mTOR and MEK inhibitors. We also identified a recurrent somatic gain-of-function *ARAF* mutation (c.604T>C:p.S214P) in a 12-year-old male with advanced anomalous lymphatic disease unresponsive to sirolimus therapy and in another unrelated adult patient. Endothelial cell studies showed the mutation altered actin skeleton and cadherin organization, which were fully reversed by inhibition of MEK signaling. Functional relevance of the mutation was also validated by recreating a lymphatic phenotype in a zebrafish model, with rescue of the anomalous phenotype using a MEK inhibitor. Subsequent therapy of the lead proband with a MEK inhibitor led to dramatic clinical improvement, with remodeling of the patient's lymphatic system with resolution of the lymphatic edema, marked improvement in his pulmonary function tests, cessation of supplemental oxygen requirements, and near normalization of daily activities. Taken together, our work demonstrates the realization of precision medicine approach enabled through a novel genetic discovery in a patient with lymphatic anomaly of a previously unknown etiology.

1290W

Characterization of lymphatic anomaly-associated mutations in a cell model system. *M.E. March¹, D. Li¹, A. Gutierrez Uzquiza¹, L. Matsuoka¹, R. Pellegino da Silva¹, J. Garifallou¹, C. Kaminski¹, J. Smiler¹, C. Kao¹, H. Hakonarson^{1,2,3}.* 1) Center for Applied Genomics; Children's Hospital of Philadelphia; 2) Department of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA; 3) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

Central conducting lymphatic anomaly is a rare, congenital, and progressive disorder in development of lymphatic vessels and architecture. Multiple patients diagnosed with lymphatic anomaly have been analyzed through whole exome sequencing at the Center for Applied Genomics. Although treatment for lymphatic anomaly is typically performed with inhibitors of the mTOR pathway, we have identified mutations in three genes that are canonically associated with signaling through the ERK pathway. Mutations have been found in a fourth gene with existing but less well established connections to ERK signaling. Functional characterization of the discovered mutants has begun with expression of each wild type or mutant protein in an endothelial cell line. Overexpression of one mutant (in the *ARAF* gene) clearly alters cell morphology, intracellular actin structures, and localization of the vascular endothelial cell adhesion molecule (VE-CAM), such that cells form less complete cell-cell junctions. Perturbations in cell structure and morphology are reversed upon treatment of the cells with inhibitors of ERK signaling. Characterization of the other mutants is currently underway, and may provide evidence for alterations to treatment of lymphatic anomaly patients in which ERK-related mutations have been observed.

1291T

Conservative management of pseudoanodontia of two cases of GAPO syndrome and Cleidocranial dysplasia. *M.I. Mostafa, M.A. AbdelRahmen, M.A. Abdelkader, M.I. Mehrez.* National Research Centre, Dokki, Cairo, Egypt.

GAPO syndrome (OMIM # 230740) is a very rare genetic disorder. GAPO is an acronym for syndrome's main manifestations: (G) Growth retardation, (A) Alopecia, (P) Pseudoanodontia (clinical absence of teeth due to a failure in eruption) and (O) Optic atrophy. Autosomal recessive mode of inheritance has been suggested in several reports. Cleidocranial dysplasia (CCD; OMIM #119600), is an autosomal dominant skeletal dysplasia affecting the formation of both bone and teeth. The main dental characteristic features are supernumerary teeth and Pseudoanodontia. In this report, conservative management of pseudoanodontia of two cases of GAPO Syndrome and Cleidocranial dysplasia is suggested. The dental management of 19-year old male suffered from Cleidocranial dysplasia depended on preservation of the unerupted teeth by orthodontic traction after their surgical exposure. In spite of the long duration of treatment, the patient and operators were satisfied with the outcome. Dental management of GAPO syndrome is also a challenging procedure. Pseudoanodontia resulting from the unerupted primary and permanent teeth causes an increase in bone ridge volume in a buccolingual direction. This condition hinders the installation of dentures. Surgical removal of teeth, even if partial, seems too complicated. Unlike Cleidocranial dysplasia, surgical exposure of the impacted teeth followed by orthodontic traction is not recommended because teeth are ankylosed. In the presented case, conventional complete denture was suggested. The outcome after more than three years of follow up was unexpectedly excellent. Both the parents and the patient were satisfied especially regarding retention, function and esthetics. Management of pseudoanodontia is still challenging due to lack of researches on such rare cases. Each case should be managed independently according to its own clinical and radiographic presentation combined with results of previously reported cases. The best way of management could be the easiest one.

1292F

Both hetero- and homozygote IDCA syndrome-associated G-protein β -subunit 5 knockout mice demonstrate heart rate perturbation and cardiac dysfunction. P. De Nittis¹, A. Sarre², J. Chrast¹, C.D. Koopman^{3,4}, J. Bakkers^{3,4}, W.F. Simonds⁵, T. Pedrazzini⁶, A. Raymond¹. 1) Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland; 2) Cardiovascular Assessment Facility, University of Lausanne, Lausanne, Switzerland; 3) Department of Medical Physiology, Division of Heart and Lungs, University Medical Center Utrecht, Utrecht, The Netherlands; 4) Hubrecht Institute-KNAW, University Medical Centre Utrecht, Utrecht, The Netherlands; 5) Metabolic Diseases Branch/NIDDK, National Institutes of Health, Bethesda, MD, USA; 6) Experimental Cardiology Unit, Department of Cardiovascular Medicine, University of Lausanne Medical School, Lausanne, Switzerland.

Cardiac electrophysiology is controlled by the sympathetic and parasympathetic branches of the autonomic nervous system. Parasympathetic modulation of cardiac output is primarily mediated by acetylcholine via M₂-muscarinic receptors (M₂R). The activation of M₂R triggers G-protein-dependent switching of G-protein-gated inwardly rectifying K⁺ channels (GIRK), hyperpolarizing the cell membrane. The kinetic of M₂R-GIRK signaling is negatively regulated by "regulator of G-protein signaling" (RGS) proteins. In particular, the heterotrimeric G-protein β -subunit 5 (GNB5)-RGS6 complex is involved in cardiac GIRK deactivation. In human, inactivating mutations in *GNB5* are causative of the autosomal recessive Intellectual Developmental Disorder with Cardiac Arrhythmia (IDCA) syndrome (MIM:#617173) with impaired cardiac sinus conductance as one cardinal symptom. *Gnb5*-null mice recapitulated many of the corresponding human disease phenotypes such as learning deficiencies, hyperactivity, impaired motor coordination and perturbed vision. Here, we used *Gnb5* knock-out mice to assess *Gnb5*-mediated signaling in the autonomic control of the heart. We monitored the electrophysiological properties of the heart in homozygous (KO, n=8), heterozygous (HET, n=8) and wild-type (WT, n=7) young-adult male mice using *in vivo* ECG (electrocardiography) telemetry. Cardiac function was assessed by two-dimensional and M-mode echocardiography. KO mice were smaller than WT and HET mice ($p < 0.001$), and had a smaller heart ($p < 0.05$), but exhibited a better cardiac function, as judged by an increase in fractional shortening and ejection fraction ($p < 0.01$ as compared to WT mice). Surprisingly, HET mice also exhibited better function when compared to WT littermates. Moreover, baseline ECG parameters were not different between groups. However, KO mice treated with carbachol (CCh), a parasympathomimetic, presented significant CCh-induced bradycardia. In contrast, the anti-parasympathetic drug atropine had the same effect in KO, HET and WT mice. Altogether, our data demonstrate that *Gnb5* KO mice exhibit altered heart rate responses to pharmacological manipulations that were consistent with impaired parasympathetic activity, a result further corroborated by studies in zebrafish larvae, knocked-out for both *gnb5a* and *gnb5b* paralogs. We conclude that even a partial reduction in *Gnb5* function can perturb normal heart activity, and that *Gnb5* signaling is essential for its parasympathetic control.

1293W

A Japanese patient with a novel missense mutation affecting the same amino acid as the recurrent *PACS1* mutation. N. Miyake¹, S. Ozasa², H. Mabe², S. Kimura³, N. Matsumoto¹. 1) Yokohama City University, Yokohama, Japan; 2) Kumamoto University, Kumamoto, Japan; 3) Children's Sleep & Development Medical Research Center, Hyogo Prefectural Central Rehabilitation Hospital, Kobe, Japan.

PACS1 (phosphofurin acidic cluster sorting protein 1, MIM#607492) encodes *PACS1* protein, which is a trans-Golgi-membrane traffic regulator. *PACS1* is highly expressed in human fetal brain and widely in various tissues. A single unique missense mutation in *PACS1* (NM_018026.3:c.607C>T, p.Arg203Trp) is known to cause Schuurs-Hoeijmakers syndrome (MIM#615009), characterized by intellectual disability (severely affected language development), characteristic face (arched eyebrows, widely-spaced eyes, wide mouth with downturned corners, and thin upper vermillion), oral aversion, seizures, and brain structural abnormalities. To date, at least 31 patients with *PACS1* mutation have been reported. Interestingly, all had the identical missense mutation (c.607C>T, p.Arg203Trp) occurring *de novo*. By trio-based whole exome sequencing, we analyzed a girl showing developmental delay, short stature, growth impairment, hypotonia, microcephaly and characteristic dysmorphic face, and identified a different *PACS1* mutation at the same amino residue (c.608G>A, p.Arg203Gln) only in the proband. We confirmed this variant occurred *de novo*. This amino acid is located very close to the functionally important CK-binding motif (Arg196 to Tyr200), implying that Arg203 is functionally important. In addition, our patient showed similar clinical features to those observed in the patients with the recurrent mutation, further supporting the prediction that the novel variant (c.608G>A, p.Arg203Gln) is also pathogenic. A previous report suggests that the recurrent *PACS1* mutation is likely to have dominant-negative or gain-of-function effects. Generally hotspot mutations strongly suggest "gain-of-function". In addition, at least five variants leading *PACS1* truncation have been registered in public control database, suggesting that these loss-of-function *PACS1* variants may result in no phenotype or different phenotype effects from those seen in Schuurs-Hoeijmakers syndrome. Since information on the mechanism by which *PACS1* mutation causes disease is still limited, further analysis of additional patients is required.

1294T

Study of genetic defects in patients with malformations affecting digits.

A. Rai¹, P. Srivastava¹, P. Sharma¹, N. Gupta², R.D. Puri³, S.J. Patil⁴, P. Ranganath⁵, K. Gowrishankar⁶, A. Dutta⁷, J.P. Soni⁸, S. Danda⁹, K. Mandal¹, S.R. Phadke¹. 1) Sanjay Gandhi Postgraduate Institute of Medical Sciences, Lucknow, India; 2) Division of Genetics, Department of Pediatrics, All India Institute of Medical Sciences, New Delhi, India; 3) Institute of Medical Genetics & Genomics, Sir Ganga Ram Hospital, New Delhi, India; 4) Department of Medical Genetics, Narayana Multispecialty Hospital, Bangalore, India; 5) Department of Medical Genetics, Nizam's Institute of Medical Genetics, Hyderabad, Telangana, India; 6) Kanchi Kamakoti CHILDS Trust Hospital, Chennai, India; 7) North Bengal Medical College & Hospital Sushrutanagar, Siliguri, India; 8) Dr. S.N. Medical College, Jodhpur, India; 9) Christian Medical College & Hospital, Vellore, India.

Malformations affecting digits are the most frequent congenital malformation in human occurring in about one in 1000 neonates. Polydactyly and oligodactyly indicates increase or decrease in the number of digits. Other anomalies affecting digits are camptodactyly, clinodactyly, brachydactyly and arachnodactyly. The anomalies of digits can be isolated or syndromic. Hundred cases with malformations affecting digits were studied for genetic etiologies using Sanger sequencing, cytogenetic microarray and exome sequencing. After detailed phenotyping cases with monogenic disorders with known causative genes, Sanger sequencing has been done. Cases of unknown etiology were subjected to chromosomal microarray and exome sequencing. For genes of monogenic disorders having more than 10 exons exome sequencing has been done. Sanger sequencing identified causative variants in 13 cases (9 were known and 4 were novel) out of 27 monogenic malformation cases affecting digits. These include *ECEL1*, *FGFR2*, *ROR2*, *PITX1*, *TP63* and *SOST1*, *NOG1* genes. Chromosomal microarray identified copy number variations in two cases out of 19 cases; 7q21.2-q22.1 deletion (7 MB) in a patient with split hand-split foot malformation and sensorineural hearing loss and 13q31.1 deletion (1.4 MB) in a patient with short 5th metacarpal and ptosis, genu valgum. By homozygosity mapping, two candidate genes were identified in a patient with Arthrogryposis Renal Dysfunction and Cholestasis and in a patient with Ellis-van Creveld syndrome which was further subjected to exome sequencing due to its large genes. Samples for exome sequencing were outsourced which were sequenced using Nextera Rapid Capture Exome from Illumina and raw data were analysed for disease causing variations. Exome sequencing identified causative variants in 23 cases (10 were known and 13 were novel) out of total 59 cases. This study showing the efficacy of these three techniques in evaluating the genetic cause of limb malformation affecting digits. The mutation data was used for providing genetic counseling to the families. The mutation data was used for providing genetic counseling to the families. Exome sequencing data of 26 cases in whom the causative variants were not identified is being analysed for variants in novel genes.

1295F

Mutation spectrum and phenotypic features of Asian patients with clinical suspicion of Noonan syndrome. E.C. Tan, A.L. Koh, E.S. Tan, M.S. Brett, A.H.M. Lai, S.S. Jamuar, I. Ng. KK Women's & Children's Hospital, Singapore, Singapore.

The aims of this study were to identify the underlying mutations and characterize the phenotypic features of patients with clinical suspicion of Noonan syndrome from our population. The 17 subjects of Southeast Asian descent were identified from the hospital's Genetics clinics after assessment by the attending clinical geneticists. Next generation sequencing was performed with a custom-designed targeted gene panel using the HaloPlex™ Target Enrichment System and the MiSeq System. Complete data for all 17 patients was available for only eight features, of which short stature was the most common (71%), followed by pectus deformity (47%). Heterozygous missense mutations were found in 13 patients: eight in *PTPN11*, three in *SOS1*, and one each in *RIT1* and *KRAS*. All are previously reported mutations that have been identified in other patients with Noonan syndrome. Seven of these mutations were *de novo* while two were inherited. The inheritance of the identified mutations (all *PTPN11*) for the remaining four subjects was unknown as parental samples were not available. The two inherited mutations (one each for *PTPN11* and *SOS1*) were from parents who had some features of Noonan syndrome. Eleven of the 13 patients with identified mutations had low-set ears while 10 had short stature. Nine had cardiac defects, with pulmonary stenosis found in five patients. Four of the eight patients with *PTPN11* mutations had atrial septal defect. Only two had pulmonary stenosis which is reported to be common for *PTPN11* mutation carriers, while another two had hypertrophic cardiomyopathy, which is negatively associated with *PTPN11* mutations. Our study provides the mutation and phenotypic spectrum of Noonan syndrome from a new population group. Our diagnostic yield was 76.5% for a panel of 14 genes, similar to what others reported. While the mutation spectrum in our patients appeared to be similar to populations elsewhere, indicating that there are no population-specific hotspots; the clinical manifestations seemed to be heterogeneous. Our patients with the same mutations or affected gene did not have similar clinical phenotypes with each other or with what had been reported in other studies. Although the four patients who did not have a mutation identified had fewer clinical features, they might have mutations in genes that were not in the panel.

1296W

Mouse model of Costello syndrome: Impaired hepatic energy homeostasis and resistance to high-fat diet induced obesity in mice with a heterozygous *Hras* G12S mutation. D. Oba^{1,6}, S. Inoue¹, S. Miyagawa-Tomita², Y. Nakashima³, T. Niihori¹, S. Yamaguchi⁴, Y. Matsubara^{1,5}, Y. Aoki¹. 1) Medical Genetics, Tohoku university school of medicine, Sendai, Japan; 2) Veterinary Technology, Yamazaki gakuin University, Tokyo, Japan; 3) Pediatrics, Seirei Hamamatsu General Hospital, Shizuoka, Japan; 4) Pediatrics, Shimane University, Faculty of Medicine, Shimane, Japan; 5) National Center for Child Health and Development, Tokyo, Japan; 6) Medical Genetics, Saitama Children's Medical Center, Saitama, Japan.

Costello syndrome is a "RASopathy" that is characterized by hypertrophic cardiomyopathy, dysmorphic facial appearance, growth retardation, hypoglycemia and predisposition to malignancies. Our group identified that germline *HRAS* mutations cause Costello syndrome in 2005. Overview of *HRAS* mutations reported in Costello syndrome patients reveals that more than 80% of patients with Costello syndrome have a *HRAS* G12S mutation. Clinical manifestations, including hypoketotic hypoglycemia, increased resting energy expenditure and severe growth deficiency, suggest metabolic impairment in patients with Costello syndrome. Mouse models of Costello syndrome with a *HRAS* G12V mutation have been reported. However, mice with the most frequent G12S mutation have not been generated. In this study, to explore the mechanisms of energy reprogramming by the oncogenic *HRAS* mutation *in vivo*, we generated new knock-in mice expressing a heterozygous *Hras* G12S mutation (*Hras*^{G12S/+}). *Hras*^{G12S/+} exhibited dysmorphic facial features, including a round head and a shortened nasal bridge, cardiomegaly with cardiomyocyte hypertrophy and kidney anomalies. *Hras*^{G12S/+} mice showed resistance to high-fat diet induced obesity with microvesicular hepatic steatosis. Under fasted conditions, hypoketosis and elevated blood levels of long-chain fatty acyl-carnitines were observed. In addition, decreased hepatic gene expressions associated with mitochondrial fatty acid oxidation was observed in *Hras*^{G12S/+} mice compared with wild type (*Hras*^{+/+}) mice. We also examined an intraperitoneal glucose tolerance test (IPGTT) and gene expressions related to glucose, organic acid and glutamine metabolism in liver tissues. These findings suggest the altered energy homeostasis, including impaired mitochondrial fatty acid oxidation, in *Hras*^{G12S/+} mice. Further analysis showed elevated fecal lipid levels in *Hras*^{G12S/+} mice compared with *Hras*^{+/+} mice under fed a high-fat diet, suggesting impaired lipid absorption in *Hras*^{G12S/+} mice. Our findings suggest that altered energy homeostasis as well as fat malabsorption would be associated with the lean phenotype in *Hras*^{G12S/+} mice during high-fat diet feeding..

1297T

Mutation spectrum and clinical manifestations in low-mosaic patients with Tuberous Sclerosis Complex (TSC). K. Giannikou¹, K. Lasseter¹, J.M. Grevelink², M.E. Tyburczy³, K.A. Dies³, J. Glass⁴, L. Hamieh¹, A. Thorne⁵, A. Nag⁶, D. Krueger⁷, S. Ruoss², D.N. Franz², E.A. Thiele⁸, M. Sahin³, D.J. Kwiatkowski¹. 1) Division of Pulmonary and Critical Care Medicine and of Genetics, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, United States of America; 2) Boston Dermatology and Laser Center, Massachusetts General Hospital, Boston, United States of America; 3) Department of Neurology, Boston Children's Hospital and Harvard Medical School, Massachusetts, Boston, United States of America; 4) Department of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, United States of America; 5) Center for Cancer Genome Discovery, Dana Farber Cancer Institute, Boston, Massachusetts, United States of America; 6) Department of Pediatrics, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, United States of America; 7) Pulmonary and Critical Care Medicine, Stanford University, California, United States of America; 8) Department of Neurology, Massachusetts General Hospital, Boston and Harvard Medical School, Massachusetts, United States of America.

Background: Tuberous Sclerosis Complex (TSC) is due to mutations in *TSC1/TSC2*. Mosaicism is well-known in TSC, and has been associated with milder symptoms. The prevalence of mosaicism in different organ systems and tissues in TSC has not been studied in detail. **Aim:** To define the mutation prevalence in multiple tissues from TSC patients with low allele frequency mosaicism, and to correlate the degree of mosaicism in various tissues/cell types with clinical features. **Material and methods:** 34 TSC patients (12 male, 22 female; 3 familial cases; median age 25 yrs) who contacted our lab for genetic analysis, and had <10% mutant allele frequency [MAF] in *TSC1/TSC2* in DNA from blood, saliva, or/and skin lesions are reported. 176 different DNA samples from these 34 patients were analyzed, including normal skin, skin lesions (e.g. angiofibroma, ungual fibroma), cultured cells (melanocytes, keratinocytes, fibroblasts), blood, saliva, urine and semen. Mosaic mutations were identified by targeted MPS (Illumina) performed on >1 skin lesion per patient and use of custom code in Unix/Python/Matlab with mean target coverage of 660 reads. All mutations identified were validated by amplicon NGS in all patient samples. **Results:** The mosaic patients all met diagnostic criteria for TSC. Facial angiofibroma (86%) and renal angiomyolipoma (61%) were common, while autism/intellectual disability (21%) and LAM (8%) were uncommon in comparison to other TSC series (e.g. full heterozygous pediatric cases from Dabora *et al.*, AJHG 2001). Three patients (9%) had *TSC1* germline mutations and 31 (91%) had *TSC2* mutations; 8 of them had a second *TSC2* somatic hit seen only in skin lesions. Saliva and normal skin DNA MAFs were generally similar to blood DNA. However, skin lesion MAFs were significantly (>3-fold) higher than blood DNA for 22 of 23 cases, with 6 showing zero MAF in blood DNA. Two women with MAF 0.2% and 5.9% in blood DNA had a child with TSC. Three mildly affected men with MAF 0%, 1.4%, and 6.9% in blood had no evidence of the mutation in two different semen samples analyzed, while two men had a *TSC2* mutation in semen (MAF 2.6%-16.7%) with blood MAFs of 2.3% and 3.9%, respectively. **Conclusions:** TSC due to low-level (<10% MAF) mosaicism has a distinctive clinical appearance. Analysis of TSC skin lesions as well as multiple samples enhances mutation detection and can clarify first vs. second hit mutations. Transmission risk can be assessed through analysis of sperm DNA in mosaic TSC men.

1298F

Phenotypic and mutational spectrum of twenty-six Egyptian patients with Sjögren-Larsson syndrome: Identification of novel and founder *ALDH3A2* gene mutations. M.S. Abdel-Hamid¹, M.Y. Issa², G.M.H. Abdel-Salam², S.F. Abdel-Ghaffar¹, M.S. Zaki². 1) Medical Molecular Genetics, National Research Centre, Giza, Cairo, Egypt; 2) Clinical Genetics, National Research Centre, Giza, Cairo, Egypt.

Sjögren-Larsson syndrome (SLS) is a rare neurocutaneous recessive disorder characterized by congenital ichthyosis, spastic diplegia and intellectual disability. It is an inborn error of lipid metabolism caused by recessive mutations in the *ALDH3A2* gene encoding the fatty aldehyde dehydrogenase that play a pivotal role in metabolism of long-chain aliphatic aldehydes and alcohols. In this report, we describe the clinical, neuroradiological and molecular findings of 26 patients derived from 19 unrelated Egyptian families with SLS. All patient shared the typical clinical characteristics of SLS including spasticity, ichthyosis and intellectual disability. Brain MRI showed deep white matter affection as demyelination in all patients varied in severity and found to correlate well with the severity of the presentation. Mutational analysis of the *ALDH3A2* gene revealed 11 distinct pathogenic mutations including 7 previously unreported ones. The mutations identified were 6 missense, 2 splice site, 1 nonsense, 1 inframe deletion and a large intragenic deletion of exon 1. Three mutations (p.S365L, p.R9* and p.G400R) were recurrent in our patients with frequencies ranging from 16% to 26%. Interestingly, patients carrying the two mutations p.S365L and p.R9* shared similar haplotypes suggesting possible founder effects in our population. In conclusion, [MP1] we presented a large cohort of patients with the characteristic clinical and brain imaging findings of SLS but with variable severity. We also identified many novel and founder *ALDH3A2* mutations thus expanding the mutational spectrum of the disorder. .

1299W

Identification of new sequence variants in rare-genetic disease-causing genes using high-throughput DNA sequencing. M.C. Ergoren^{1,8}, E. Manara², K. Terali³, S. Paolacci², P. Enrico⁴, B. Sanlidag⁵, E. Dirik⁶, G. Mocan⁶, S. Temel⁷, M. Bertelli⁸. 1) Medical Biology, Near East University, Faculty of Medicine, Nicosia, Cyprus; 2) MAGI Euregio, Bolzano, Italy; 3) Medical Biochemistry, Near East University, Faculty of Medicine, Nicosia, Cyprus; 4) MAGI'S LAB, Rovereto, Italy; 5) Pediatrics, Near East University, Faculty of Medicine, Nicosia, Cyprus; 6) Medical Pathology, Near East University, Faculty of Medicine, Nicosia, Cyprus; 7) Medica Genetics, Uludag University, Faculty of Medicine, Bursa, Turkey; 8) Experimental Health Sciences Research Center (DESAM), Near East University, Nicosia, Cyprus.

The Human Genome Project was accomplished in draft form in 2001. In the last two decades, approximately 50% of genes causing 7,000 rare monogenic disorders, which are often difficult to diagnose based only on symptoms, have been identified. More than 50% of individuals with rare genetic disorders are yet to be diagnosed and treated in order to improve their quality of life. Generally, suggestive clinical features can be used to distinguish one condition from another; however, in some cases these clinical features overlap with a number of other genetic conditions. In our current study, the recent high-throughput genomic technology, which involves whole-exome sequencing (WES), was used to detect rare-genetic disease-causing variants and proved itself as a successful strategy for gene identification. Here, WES was used to sequence trios as well as larger pedigrees. An in-house bioinformatics pipeline was also employed to select pathogenic variants. Accordingly, we identified novel genetic variants each of which causes neonatal progeroid (Wiedemann–Rautenstrauch) syndrome, Treachers–Collins syndrome, or Joubert syndrome. The translational effects of these variants were predicted by *in silico* molecular modelling. Our study highlights the importance of discovering rare-disease-causing variants both in clinical diagnosis and in improved understanding of the underlying molecular mechanisms.

1300T

Short rib polydactyly with Joubert syndrome associated with a novel KIAA0753 mutation: Expanding the phenotype of skeletal ciliopathies. S. Albanyan¹, A. Hammarsjö^{2,3}, G. Grigelioniene^{2,3}, S. Blaser¹, K. Millar¹, S. Telesca⁴, D. Chitayat^{1,4}. 1) The Hospital for Sick Children, Toronto, ON, Canada; 2) Department of Molecular Medicine and Surgery, Center for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden; 3) Department of Molecular Medicine and Surgery, Center for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden and 3) Clinical Genetics, Karolinska University Hospital, Stockholm, Sweden; 4) Dept. of Obstetrics and Gynecology, The Prenatal Diagnosis and Medical Genetics Program; University of Toronto, Toronto, Ontario, Canada.

Skeletal ciliopathies are a rare group of diseases with significant genetic heterogeneity and phenotypic overlap. The main features of skeletal ciliopathies include thoracic hypoplasia, short tubular bones with metaphyseal dysplasia, trident pelvis and polydactyly. Extra skeletal manifestations include abnormalities of the brain, retina, genitalia, kidney, and liver. To date, at least 30 genes have been identified in association with skeletal ciliopathies. Mutations in KIAA0753 have recently been reported in association with, orofaciocigital syndrome (OFD), Joubert syndrome (JBTS) with growth hormone deficiency, short rib thoracic dysplasia in combination with Joubert syndrome (JBTS). We report a biallelic, pathogenic variant in KIAA0753 in female siblings born to a consanguineous couple. Both had short-rib polydactyly and molar tooth brain sign. Both the first and second sibling died shortly after birth of respiratory failure. Physical exam on the first child revealed macrocephaly, wide fontanelles, frontal bossing, micrognathia, low-set ears, small nose with a broad nasal bridge, short and bowed limbs, and brachydactyly. Fetal radiography showed narrow thorax, trident ilia with spikes at the sacrosciatic notches and short bowed humeri, femora, tibiae, radii and ulnae. ArrayCGH was normal female and showed loss of heterozygosity of many segments. The second sibling was conceived naturally two years later and ultrasound done at 20 weeks gestation showed mild ventriculomegaly, abnormal 4th ventricle, abnormal cerebellar vermis and short tubular bones and a bell shaped chest. Fetal MRI at 22 weeks gestation showed mild ventriculomegaly and molar tooth abnormality. Examination after delivery showed prominent forehead, short limbs and right postaxial polydactyly. Skeletal survey showed short ribs with metaphyseal flaring, handlebar clavicles, hypoplastic lower pelvis and polydactyly. Brain MRI showed multiple congenital abnormalities in the posterior fossa, and a "molar tooth" appearance of the midbrain, consistent with Joubert syndrome. In summary, our results alongside clinical data from 8 previously reported patients indicate that KIAA0753-related ciliopathy syndrome constitutes a phenotypic continuum from lethal skeletal dysplasia, with severe thoracic hypoplasia, bowed limbs, and JBTS features at the severe end, to non-lethal metaphyseal skeletal dysplasia with mild thoracic hypoplasia, JBTS and OFD at the mildest phenotypic end of the spectrum.

1301F

Characterization of an emerging X-linked trichothiodystrophy due to a novel truncating variant in RNF113A. D.T. Belefard¹, A. Abu-El-Haija¹, P.M. Martin², S. Rego¹, A. Huang³, G. Capodanno³, J.T. Shieh^{1,2}, J. Van Ziffle², N. Risch², A. Slavotinek¹, B. Mendelsohn¹. 1) Division of Medical Genetics, University of California, San Francisco, San Francisco, CA; 2) Institute for Human Genetics, University of California, San Francisco, San Francisco, CA; 3) Division of Pediatric Endocrinology, University of California, San Francisco, San Francisco, CA.

Trichothiodystrophy (TTD) is an emerging group of heterogeneous conditions characterized by short stature, photosensitivity, seizures, immune deficiency, endocrinopathies, intellectual disability (ID), and brittle, sulfur-deficient hair with "tiger-tail" banding pattern on polarizing light microscopy. Corbett *et al.* 2015 described the first X-linked TTD caused by a nonsense mutation in *RNF113A*. We now contribute to the phenotypic spectrum of this rare, X-linked TTD by providing a detailed characterization of a ten year old boy with non-photosensitive TTD due to a previously-unreported truncating variant in *RNF113A*. The patient presented in status epilepticus with a history of refractory seizures, transient neonatal hyperinsulinism, failure to thrive with G-tube dependence, duodenal strictures, hypergonadotropic hypogonadism, and global developmental delay. OFC, weight, and height were 2 SD below the mean for age. He had facial differences and 2-3 toe syndactyly. His hair was sparse and brittle and his skin showed diffuse cutis marmorata without ichthyosis. He also had bilateral undescended testes and a small phallus with elevated gonadotropins and was hypogammaglobulinemic and lymphopenic with multiple infections. Clinical whole exome sequencing was performed as a duo (proband and mother) as part of the Prenatal and Pediatric Genomic Sequencing Program (P3EGS) at UCSF and showed a maternally inherited hemizygous truncating variant in *RNF113A*, c.903_910delGACAGACCA (p.Gln302fs*12). Brain MRI showed diffuse dysmorphology. Hair banding pattern on polarized light microscopy was nondiagnostic. Maternal X-inactivation studies showed skewed X-inactivation with a ratio of 100:0. Ours is the first report of a non-familial X-linked TTD caused by a novel truncating *RNF113A* variant. A single prior report described two male first cousins with microcephaly, ID, sparse brittle hair, facial differences, seizures, frequent infections, brain dysmorphology, and variable endocrinopathies including severe growth failure, central adrenal insufficiency, diabetes insipidus, and gonadal dysgenesis with microphallus. While the duodenal strictures seen in our patient were not reported, other structural GI differences were described. Interestingly, our patient's mother had short stature and 100% skewed X-inactivation as seen in other obligate female carriers. However, further studies are ongoing and are essential in characterizing the full phenotypic spectrum of this condition.

1302W

Craniofrontonasal syndrome caused by introduction of a novel uATG in the 5'UTR of *EFNB1*. V.L. Romanelli Tavares¹, E. Kague^{1,2}, C.M. Musso¹, T.G.P. Alegria³, R.S. Freitas⁴, D. Bertola^{1,5}, S.R.F. Twigg⁶, M.R. Passo-Bueno¹.

1) Centro de Pesquisa sobre o Genoma Humano e Células-Tronco, Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, SP, Brazil; 2) Department of Physiology, Pharmacology and Neuroscience, University of Bristol, Bristol, UK; 3) Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, São Paulo, Brazil; 4) Centro de Atendimento Integral ao Fissurado Lábio Palatal (CAIF), Curitiba, Paraná, Brazil; 5) Instituto da Criança do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, São Paulo, São Paulo, Brazil; 6) Clinical Genetics Group, MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK.

Craniofrontonasal syndrome (CFNS) is an X-linked disorder caused by *EFNB1* mutations, in which females are more severely affected than males. Severe male phenotypes are associated with mosaicism, supporting cellular interference for sex bias in this disease. Although many variants have been found in the coding region of *EFNB1*, just two pathogenic variants have been identified in a same nucleotide in 5' UTR, disrupting the stop codon of an upstream open reading frame (uORF). uORFs are known to be part of a wide range of post-transcriptional regulation processes and only recently has their association with human diseases come to light. In the present study, we analyzed *EFNB1* in a female patient with typical features of CFNS. We identified a variant, located at c.-411, creating a new upstream ATG (uATG) in the 5'UTR of *EFNB1*, which is predicted to alter an existing uORF. Dual-luciferase reporter assays showed significant reduction in protein translation, but no difference in the mRNA levels. Our study demonstrates, for the first time, the regulatory impact of uATG creation on *EFNB1* levels and suggests that this region should be target in molecular diagnosis of CFNS cases without pathogenic variants in the coding and splice sites regions of *EFNB1*.

1303T

Argininosuccinic aciduria in Korean population: Its low prevalence among the Urea cycle disorders. G. Kim¹, D. Kim², Y. Kim², B. Lee^{1,2}, H. Yoo^{1,2}.

1) Medical Genetics & Genomics, Asan Medical Center, Seoul, South Korea; 2) Department of Pediatrics, Asan Medical Center, University of Ulsan College of Medicine.

In Korea, more than 90% of newborns are screened for inborn metabolic disorders by tandem mass spectrometry as part of the newborn screening program. Urea cycle disorders (UCDs) are included in this screening. Argininosuccinic aciduria (ASA) may be suspected in newborns showing the elevation of citrulline in their screening. ASA is the is caused by an argininosuccinate lyase deficiency and is biochemically characterized by elevation of citrulline, argininosuccinic acid and arginine deficiency. Clinically, patients with ASA exhibit hyperammonemia and other characteristic features including hepatic fibrosis, hypertension, neurocognitive deficiencies, and trichorrhexis nodosa. In the current study, we retrospectively reviewed the clinical findings, biochemical profiles, and genotypic characteristics of five Korean patients with ASA, who showed typical phenotypes and biochemical findings of the disease. Molecular analysis of these patients revealed six novel ASL mutations. Of note, these five patients were the all patients identified in Korea over the two-decade period. ASA is the second most common inherited UCD in the USA and European countries (17%), whereas ASA was as rare as 5.1–5.6 % in Japan and 6.3% in Korea, making it the fourth most common UCD in these East Asian countries. Although there were no epidemiological surveys of UCDs, ASA is expected to be extremely rare in other East Asian countries, including China and Taiwan. In contrast, ASA was the most common UCD in Malaysia. These findings suggest that there are the ethnic differences in the prevalence rates of each UCD among various populations.

1304F

Farber disease (acid ceramidase deficiency): Patient demographics from an ongoing natural history study. A. Solyom¹, C.R. Ferreira², P. Tanpai-boon³, J. Mitchell³, N. Guelbert⁴, N. Mungan⁵, F.D. Bulut⁶, B. Magnusson⁶, E. Sundberg⁶, N. Arslan⁷, B. Makay⁸, R. Puri⁸, S. Bijarnia⁸, L. Selim⁸, I. Gamal el Din⁹, M. DiRocco¹⁰, S. Ozen¹¹, E.D. Batu¹¹, M. Torcoletti¹², D. Tetzl¹³, A. Kimura¹³, P. Harmatz¹⁴. 1) Enzyvant, Basel, Switzerland; 2) Children's National Medical Center, Washington, DC, USA; 3) Montreal Children's Hospital, Montreal, Canada; 4) Children's Hospital of Cordoba, Cordoba, Argentina; 5) Cukurova University Hospital, Adana, Turkey; 6) Karolinska University Hospital, Stockholm, Sweden; 7) Dokuz Eylul University Hospital, Izmir, Turkey; 8) Sir Ganga Ram Hospital, Delhi, India; 9) Cairo University Children's Hospital, Cairo, Egypt; 10) Istituto Giannina Gaslini, Genoa, Italy; 11) Hacettepe University Hospital, Ankara, Turkey; 12) University of Milan, Milan, Italy; 13) Enzyvant, Boston, USA; 14) UCSF Children's Hospital Oakland, Oakland, USA.

Introduction: Farber disease is a rare lysosomal storage disease caused by mutations in both alleles of the *ASA1* gene. The resulting deficiency of the lysosomal enzyme acid ceramidase causes accumulation of the pro-inflammatory and pro-apoptotic sphingolipid ceramide. The resulting broad spectrum of symptoms and severity may delay diagnosis or lead to misdiagnosis. The study described here is the first comprehensive and systematic clinical study of the natural history of Farber disease and is currently ongoing. **Materials and methods:** The Observational and Cross-Sectional Cohort Study of the Natural History and Phenotypic Spectrum of Farber Disease (NCT03233841) was designed to collect data on demographics, clinical presentation and phenotype, and diagnostic history including enzyme activity and *ASA1* mutations when available, using retrospective and prospective data on living patients diagnosed with Farber disease who have or have not undergone hematopoietic stem cell transplantation (HSCT), as well as medical history data on deceased patients. In addition, prospective clinical assessments will be used to assess the changes in symptoms over time. **Results:** Demographic data is available on 24 patients, including 12 living and 12 who are deceased. The ratio of females to males is 10/14. The average age of the living patients is 11.7 years, with ages ranging between 2 and 28 years. 9 living patients (75%) are below the age of 18 years, with an average age of 6 years. 8/12 (67%) of the living patients and 10/12 (83%) of the deceased patients were born in the last 10 years. Countries of birth of the patients able to be enrolled so far include Afghanistan, Argentina, Canada, Egypt, India, Iraq, Italy, Mexico, Sweden, Syria, Turkey, and the USA. **Conclusions:** From November 2017 through May 2018, 24 patients representing the full phenotypic spectrum of Farber disease, from rapidly progressive (severe), to slowly progressive (attenuated) have been enrolled from 12 centers in 7 countries. 75% of the living patients are under 18 years of age and 83% of the deceased patients were born in the last decade. Additional centers are expected to be enrolling patients over the coming months. The systematic analysis of this unique set of data will hopefully expand and deepen the understanding of acid ceramidase deficiency, as well as contribute to the information available for the evaluation of future potential therapies.

1305W

Our genematcher data sharing experience: 10 days on average to confirm the pathogenicity of a candidate gene. C. Thauvin-Robinet^{1,2,3,4}, A.L. Bruehl^{1,2}, A. Vitobello^{1,2}, F. Tran Mau-Them^{1,2}, S. Nambot^{1,2,3}, V. Quéré^{1,2}, P. Kuentz¹, J. Thevenon^{1,3}, M. Assoum¹, S. Moutton^{1,3}, N. Houcinat^{1,3,4}, N. Jean-Marçais^{1,3}, M. Lefebvre^{1,2}, A.L. Mosca-Boidron^{1,2}, P. Callier^{1,2}, C. Philippe^{1,2}, L. Faivre^{1,3}. 1) UMR1231 GAD, Inserm - Université Bourgogne-Franche Comté, Dijon, France; 2) Unit for Innovation and Genomic Diagnosis of Rare Diseases, FHU-TRANSLAD, Dijon University Hospital, Dijon, France; 3) Centre of Reference for Rare Diseases: Development disorders and malformation syndromes, Genetics Department, FHU-TRANSLAD, Dijon University Hospital, Dijon, France; 4) Centre of Reference for Rare Diseases: Development disorders, Genetics Department, FHU-TRANSLAD, Dijon University Hospital, Dijon, France.

The number of rare diseases is assessed between 6,000 and 7,000, with more than 200 diseases newly described every year, assigned to known gene, novel genes, or without molecular basis. Therefore, the total number of rare genetic disease was predicted to be around 15,000. Whole-exome or -genome sequencing (WES/WGS) appears highly powerful to identify the molecular basis of heterogeneous conditions such as intellectual disability and/or multiple congenital anomalies (ID/MCA), particularly resistant to older approaches because of frequent *de novo* mutations. However, a large majority of results remain so non-conclusive, linked to the difficulty in recruiting an adequate number of affected individuals with ultra-rare conditions, the heterogeneous clinical spectrum, the uncertain functional impacts of variants and the small number of known genes involved in diseases (~3,500 genes). Data-sharing aims to facilitate recurrence in ultra-rare conditions by identifying additional patients carrying variants in the same gene, in order to draw definitive conclusions on their implication in the disease. Several web-based tools have been developed during the last years and grouped in MatchMaker Exchange. In this study, we report the experience of a regional center that has implemented WES as a first-line diagnostic test since 2013, and attached to a research lab devoted to gene identification, in using GeneMatcher during the last 2.5 years to share candidate genes that are not yet known as responsible of human disease. We have shared 70 candidate genes identified by WES performed in individuals affected by ID/MCA. We evaluated the ability to determine the involvement of these genes and the necessary timeframe: 60/70 genes (85%) were matched to at least one other mutated individual. Data-sharing permitted to confirm the implication of 24/70 genes (39%) in ID/MCA and to rule out 6% of genes. The waiting period between submission and the first match varied, with an overall median of 4 hours. When a match occurred, the median response time between the first email to contact a submitter and the response was estimated at 31 hours. The rapid identification of these new genes remained essential for clinical characterization, genetic counselling and for translation to the diagnostic field. The introduction of friendly genematching tools, such as GeneMatcher, appeared therefore an essential step for the rapid identification of new genes responsible for ID/MCA.

1306T

Two novel FOXC2 gene mutations result in autosomal domain hereditary distichiasis in Chinese families. L. Jiang, D. Luo, R. Zheng. University of Electronic Science and Technology of China, Chengdu, Sichuan, China.

Distichiasis presents as double rows of eyelashes arising from aberrant differentiation of the meibomian glands of the eyelids, and it may be sporadic or hereditary. *FOXC2* gene mutations in hereditary distichiasis are rarely reported. Here, we examined two unrelated Chinese families with hereditary distichiasis but without lymphedema or other features of LD syndrome. The *FOXC2* gene was amplified and sequenced in all family members. Clinical examinations showed distichiasis, lower eyelid ectropion, congenital ptosis and photophobia in all affected individuals. Sequence analysis revealed two different frameshift mutations respectively, one is c.936 dup CGCCGCC and the other is c.937dupGCCGCT, in the coding region of the *FOXC2* gene. Although the two novel mutations inset different number of bases in different position, caused the same protein truncation (p.Tyr313CysfsX 464). These novel mutations are the first report of a *FOXC2* mutation in hereditary distichiasis in the Chinese population. This novel frameshift mutation causing distichiasis not only expands the germline mutation spectrum of the *FOXC2* gene in the Chinese population, but also increases the understanding of the phenotypic and genotypic correlations of *FOXC2*, and may potentially lead to improved genetic counseling and specific treatment for families with *FOXC2* in the future.

1307F

Retrospective examination of the clinical utility of clinical whole genome sequencing in the Illumina iHope program. A. Malhotra¹, C.M. Brown¹, J. McEachern¹, K. Robinson¹, D. Henry², F. Sirchia³, J.C. Ward⁴, D.R. Bentley¹, J.W. Belmont¹, D.L. Perry¹, R.J. Taft¹. 1) Illumina, Inc, San Diego, CA; 2) Pediatrics, University of California, San Francisco, CA; 3) Città della Salute e della Scienza University Hospital, Medical Genetics Unit, Turin, 10126, Italy; 4) Division of Genetics, Department of Pediatrics at UTHSC, Memphis, TN.

Clinical whole genome sequencing (cWGS) can detect multiple variants types including single nucleotide variants, small insertions and deletions, copy number variants, and mitochondrial variants in a single test. This approach has only been recently implemented for clinical use therefore data evaluating the impact of cWGS on patient care is limited. To investigate the clinical utility of cWGS, surveys were administered to the ordering providers for 77 probands enrolled in the Illumina iHope program, a philanthropic initiative that aims to provide cWGS to patients with rare and undiagnosed disorders who have limited access to genomic testing. The ordering providers were asked questions related to change in management as a result of the cWGS results. To date, 59 responses have been received, which includes 36 probands with at least one reported variant with potential clinical implications. Of these 36 cases with a genetic finding, 29 (80%) reported that cWGS provided a new diagnosis which was not previously suspected. In addition, the survey results indicated that cWGS resulted in a change in medical management for 24 (66%) cases. The most frequently reported change in management was the implementation of condition-specific supportive interventions, for example, additional testing and therapies including speech, physical and occupational therapy, renal imaging, hearing tests, and echocardiograms. Additionally, in one case, cWGS allowed the proband to subsequently be placed on a kidney transplant list, while in another case, monitoring for Wilms tumor through periodic ultrasounds has been implemented for a child diagnosed with Bohring-Opitz syndrome. Testing also had implications for additional family members in some cases, for example a maternally inherited X-linked likely pathogenic variant resulted in a recommendation to screen maternal family members. Of the 36 responses where a variant was detected, suspected diagnoses were ruled out in five probands, and four additional responses specifically noted that the results of cWGS resulted in avoidance of future unnecessary testing since the molecular diagnosis explained the proband's phenotype. In summary, the results of this study support that cWGS may be beneficial for individuals with diverse clinical presentations.

1308W

Homozygosity stretches around homozygous mutations for autosomal recessive disorders in patients from non – consanguineous families from India: Inheritance by descent is common. S.R. Phadke, P. Srivastava, P. Sharma, A. Rai, S. Masih. Dept of Medical Genetics, Sanjay Gandhi PGIMS, Lucknow, Uttar Pradesh, India.

In India the custom of marriages amongst caste groups has been followed for ages. We have seen that for many rare autosomal recessive (AR) disorders, the affected individuals are homozygous for rare disease causing pathogenic variations, suggesting effects of inbreeding. Hence we analysed the exome sequencing data of cases with pathogenic/likely pathogenic sequence variations for autosomal recessive disorders for the homozygous regions around the causative sequence variations. All 24 cases with AR disorders from consanguineous families were homozygous for the disease causing variations (12 out of 24 being novel variations) and had large (Average – 77.2 Mb, Range - 5 Mb to 271 Mb) stretches of homozygosity around the disease causing pathogenic or likely pathogenic variations. For AR disorders from non-consanguineous families, the disease causing variations were in homozygous form in 13 (9 being novel) out of 19 cases and 6 were compound heterozygous. In the cases with homozygous pathogenic variations from non-consanguineous families; there were stretches of homozygosity around the causative sequence variations (Average - 27.9 Mb, 0.6 Mb to 188 Mb). We also reviewed our data of SNP microarray of cases from 50 consanguineous and 50 nonconsanguineous families. In cases born to consanguineous parents the sizes of Regions of Homozygosity (ROH) regions were 28 Mb to 770 Mb (average 251.44 Mb i.e. 8.75 % of genome per case (range 0.97% to 26.78% of total genome). Though third degree consanguinity is the commonest, many families have consanguineous marriages in previous generations as well leading to homozygosity of more than 10% of the genome. In cases born to non-consanguineous parents, the sizes of runs of homozygosity regions varied from 3 Mb to 49 Mb (0.10 % to 1.7% of total genome; average 0.74%). Twenty six cases had runs of homozygosity stretching more than 5 Mb while 24 cases did not have any region of homozygosity more than 5 Mb. Though 52% cases had ROH regions more than 5 Mb only 8 of 26 had more than 1% genome with ROH (upto 1.7%). Long stretches of homozygosity around homozygous rare pathogenic variants supports the notion that the system of marriages between closed groups (castes) has many founder mutations and using the strategy of homozygosity by descent even in non-consanguineous families can be fruitful in identifying novel pathogenic variations and novel genes. .

1309T

Novel structural variants originating in *F8* non-coding regions explain previously unresolved cases of severe hemophilia A. M.M. Wheeler¹, J.M. Johnsen^{2,3}, G.F. Pierce⁴, C. Watson⁵, B.A. Konkle^{2,6}, D.A. Nickerson¹, NHLBI Trans-Omics for Precision Medicine. 1) University of Washington, Seattle, WA; 2) Bloodworks NW, Seattle, WA, USA; 3) University of Washington, Department of Medicine, Seattle, WA, USA; 4) National Hemophilia Foundation, New York; 5) American Thrombosis and Hemostasis Network, Chicago, IL; 6) Division of Hematology, University of Washington, Seattle, WA, USA.

Hemophilia A is an X-linked bleeding disorder resulting from deficiency in coagulation factor VIII. Numerous genetic variants (>2000) affecting the *F8* gene have been implicated as causative of hemophilia A. These include single nucleotide variants (SNVs), indels and structural variants (SVs) such as copy number variants and large intra-chromosomal inversions. For the vast majority of patients, causative variants are identified using targeted sequencing of *F8* coding regions or the use of methods which detect known SVs (e.g. inverse shifting PCR, long-range PCR, MLPA). However, these approaches fail to explain 1-3% of hemophilia A cases. In this study, we specifically performed analyses using whole genome sequence (WGS) data from 38 cases of hemophilia A that remained unexplained after exhausting available clinical testing methods. These cases were present in the My Life, Our Future hemophilia study recently sequenced by the NHLBI TOPMed program, and comprised mild (N=22), moderate (N=5) or severe (N=11) factor VIII level deficiencies. WGS analyses revealed several mild and moderate cases harbored *F8* deep intronic variants recently associated with hemophilia (N=8) and/or likely causal variants in the functionally-related *VWF* gene (N=11). In 6 severe hemophilia cases, analyses detected *F8* SVs strongly supported by WGS evidence (i.e. supported by multiple types of read evidence and multiple reads). Two deletions in intron 6 were detected in a single hemophilia case, a finding which suggests intron 6 may contain regulatory elements critical for *F8* expression. Moreover, three distinct large inversions were detected in five other cases: one case with a 720Kb inversion with breakpoints in *F8* intron 6 and *SPRY3* intron 1, one case with a 20Mb inversion with breakpoints in *F8* intron 1 and *INTS6L* intron 8, and three cases with a 7.4Kb inversion with breakpoints in *F8* intron 25 and the *SMIM9* intron 1. These SV events have not been previously reported in hemophilia and were also not present in the larger, sequenced My Life, Our Future dataset (N=2182), indicating these SVs are novel and likely causative of severe hemophilia A. Based on these analyses, our prediction is that additional deleterious SVs remain to be discovered in unexplained cases of severe hemophilia A. This work demonstrates comprehensive analyses that include detection of SVs and variants in non-coding region should be broadly applied to unexplained cases of Mendelian conditions.

1310F

Illumina Clinical Services Laboratory contributions to the ClinVar and ClinGen knowledge bases. S. Schmidt, A. Malhotra, A.J. Coffey, N.J. Burns, R. Rajkumar, K. Bluske, A. Chandrasekhar, A.R. Clause, F. Mullen, J. AVECILLA, M. Bennett, M. Rajan, D.R. Bentley, D.L. Perry, R.J. Taft. Illumina, Inc, San Diego, CA.

The human genetics community is increasingly reliant on public, centralized and vetted databases of information. ClinVar is a valuable resource of variants and their classifications, derived from clinical testing or evidence in the literature. In 2016, the Illumina Clinical Services Laboratory (ICSL) launched an ongoing program to provide systematic ClinVar donations and content updates. In 2017, ICSL launched a Gene Curation Programme that uses the ClinGen Gene Curation Framework to build a resource of curated gene-disease associations (GDAs). ClinVar submissions to date have been derived from variants interrogated as part of the Illumina TruGenome™ Predisposition Screen, a physician-ordered clinical whole genome sequencing (cWGS) test intended for generally healthy adults. Variants are interpreted using a classification scheme based on ACMG guidelines, but with an additional classification category, variant of unknown significance-suspicious (VUS-S). In addition to manual curation, the variant curation workflow utilizes an autocategorization algorithm that incorporates information about allele frequency, disease prevalence and penetrance estimates, and inheritance mode. Recently, 122,000 ICSL ClinVar entries were updated, including 40,000 classification changes. In addition, >10,000 new entries were submitted. A further submission of >50,000 variants derived from the predisposition screen is planned in 2018. Gene curation is a prerequisite for variant interpretation. ICSL's gene curation process includes investigation of disease mechanism and production of evidence summaries that describe both gene function and key phenotypic features of the associated disease. The current pipeline contains 194 genes including 147 GDAs for the ACMG 59 genes. To evaluate concordance, 16 GDAs previously curated by ClinGen were completed by ICSL. All curations were concordant with the clinical validity awarded. Forty-one GDAs previously curated by the BabySeq Project were also curated. Of the 41, 28 (68%) were fully concordant. Of the 13 discordant GDAs, ten classifications from the BabySeq Project were upgraded, with more evidence added to strengthen the association. ICSL's in-depth and rigorous curation methodology aims to provide high quality and accurate evidence-based classifications. The contributions made by ICSL will therefore help maximize the benefit of ClinVar and the ClinGen Knowledge Base for all users.

1311W

Using UK Biobank to assess the pathogenicity, penetrance and expressivity of monogenic disease variants. M.N. Weedon, B. West, M. Tuke, S. Jones, A.R. Wood, T. Laver, R.N. Beaumont, J. Tyrrell, A.T. Hattersley, T.M. Frayling, C.F. Wright. University of Exeter, Exeter, United Kingdom.

The UK Biobank (UKB) presents an excellent opportunity to assess the pathogenicity and penetrance of clinically important monogenic variants. However, this opportunity is thought to be limited because the genotyping quality of rare variants in the UKB is uncertain. Here we describe a systematic method for evaluating the analytical validity of rare variant genotypes from the UKB arrays, investigate the relationship between data quality and minor allele frequency (MAF), and evaluate the association of a subset of clinically-interesting, well-genotyped variants with relevant phenotypes in UKB. We evaluated the analytical validity of 4688 putatively clinically relevant variants using cluster intensity plots and classified 29% as being high quality. Minor allele frequency was an extremely good predictor of the likelihood of a variant being genotyped well by the UKB genotyping arrays: the false positive rate (FPR) was ~7% at MAF>0.00005, but FPR~60% at MAF>0.00001. All variants at MAF<0.000005 as being low quality. We then focused on variants with MAF>0.00001 that are good quality and have been reported as a cause of a monogenic disease in ClinVar or where variants in genes are known to cause two specific monogenic diseases: Maturity-Onset Diabetes of the Young (MODY) and developmental disorders. We assessed the frequency of these variants in the population and the association of the variants with 408 clinically relevant traits available in UKB. We identified several rare variants classified as pathogenic with substantially reduced penetrance or variable expressivity. For example, the P415A *PER3* variant that has been reported to cause familial advanced sleep phase syndrome is present at 0.5% frequency in the population and associated with only an 8 min ($P=2 \times 10^{-4}$) advance in sleep timing. We also observed novel associations with relevant traits for heterozygous carriers of some rare recessive conditions, e.g. heterozygous carriers of the R799W *ERCC4* variant that causes Xeroderma pigmentosum have a 0.3SD ($P=2 \times 10^{-8}$) increased ease of sunburning. Within our two disease subsets, we were able to refine the penetrance estimate for the R114W *HNF4A* variant in diabetes (only ~10% by age 40yrs), and refute the previous disease-association of *RNF135* indevelopmental disorders. In conclusion, this study shows the power of the UK Biobank to define the effect of rare variants, which will be much enhanced in future when genome-wide sequencing data become available.

1312T

Explaining the unexplained with genetics: Patients with bronchiectasis have a higher frequency of pathogenic variants linked to the disease. J. Bastarache, G. Bernard, J. Denny, L. Bastarache. Vanderbilt University Medical Center, Department of Medicine, Nashville, TN., USA.

Background Bronchiectasis can be caused by Mendelian diseases such as cystic fibrosis (CF) & primary ciliary dyskinesia (PCD), but most affected patients are not diagnosed with a genetic condition. We leveraged clinical & genetic data for 35,842 de-identified subjects to test the **hypothesis** that patients with unexplained bronchiectasis have a higher frequency of pathogenic alleles in genes linked to bronchiectasis compared to unaffected controls.

Methods We identified cases with bronchiectasis without a diagnosis of CF or PCD by manual review. Using ClinVar & OMIM we found 38 pathogenic variants that cause CF, PCD, or other diseases linked to bronchiectasis for which we had extant Exome BeadChip genotype data. We used regression analysis to compared the frequency of these variants among bronchiectasis cases (n=240) & controls (n=31,689) using a dominant model. We used a PheWAS to find additional phenotypes associated with bronchiectasis. **Results** 23 of 240 bronchiectasis cases (9.6%) carried a pathogenic allele, significantly more than unaffected controls. One individual was homozygous for a pathogenic *CFTR* mutation (p.Leu206Trp) & had a clinical syndrome consistent with CF. Two individuals were heterozygous for mutations in both *CFTR* & *SCNN1G*. The remaining 20 individuals were heterozygous a single pathogenic alleles in *CFTR*, a PCD gene, or *SCNN1G*. Variants in *SCNN1G* were not significantly enriched on their own, consistent with the previously described digenic inheritance with *CFTR*. PheWAS of bronchiectasis cases revealed associations with CF- and PCD-related phenotypes, such as hypovolemia (p=1e-11; OR=3.3) & hearing loss (p=2e-8; OR=2.45). **Conclusions** Cases with clinically unexplained bronchiectasis are more likely to carry pathogenic alleles linked to bronchiectasis, and more likely to have additional phenotypes characteristic of CF or PCD. These results suggest our bronchiectasis cohort is enriched for patients with undiagnosed Mendelian disease. Given the limited variants on our genotyping platform, we cannot establish how many symptomatic heterozygotes carry a second pathogenic allele. With the advent of gene specific treatments for diseases like CF, identifying these patients in clinical practice may be important.

Associations results			
Pathogenic variant set	# of cases w/ variant	OR	P
Recessive only	16	2.9	6.00E-05
Dominant only (SCNN1G)	9	1.3	0.4
CF variants	10	3.2	4.00E-04
PCD variants	6	2.4	0.04
Bloom syndrome & Ataxia-telegiectasia	0	-	-
CFTR + SCNN1G variants	2	27	6.00E-05
ALL	23	2	3.00E-03

1313F

Noninvasive immunohistochemical diagnosis and novel MUC1 mutations causing autosomal dominant tubulo-interstitial kidney disease. M. Zivna¹, K. Kidd², A. Pristoupilova¹, V. Baresova¹, B. Blumenstiel³, M. Harden³, H. Hartmannova¹, K. Hodanova¹, V. Stranecky¹, A. Vrbacka¹, P. Vyletal¹, L. Piherova¹, J. Zivny⁴, H. Hulkova^{1,5}, V. Robins⁶, O. Viklicky⁶, A. Greka^{3,7,8}, A.J Bleyer^{1,2}, S. Kmoch^{1,2}. 1) Research Unit for Rare Diseases, Department of Pediatrics and Adolescent Medicine, First Faculty of Medicine, Charles University, Prague, Czech Republic; 2) Section on Nephrology, Wake Forest School of Medicine, Winston-Salem, NC, USA; 3) Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA; 4) Institute of Pathophysiology, First Faculty of Medicine, Charles University, Prague, Czech Republic; 5) Institute of Pathology, First Faculty of Medicine, Charles University in Prague and General University Hospital, Prague, Czech Republic; 6) Nephrology Department, Institute for Clinical and Experimental Medicine, Prague, Czech Republic; 7) Brigham and Women's Hospital, Boston, MA; 8) Harvard Medical School, Boston, MA.

Autosomal dominant tubulo-interstitial kidney disease (ADTKD) encompasses a group of genetic disorders characterized by renal tubular and interstitial abnormalities, leading to slow progressive loss of kidney function requiring dialysis and kidney transplantation. This condition has been identified in almost 1000 families world-wide. Genetic studies revealed causal mutations in *UMOD*, *REN*, *MUC1* and *SEC61A1* encoding uromodulin, renin, mucin-1 and translocon subunit SEC61A, respectively. ADTKD-*MUC1* is caused by a cytosine duplication within a seven-cytosine tract in the GC rich variable number of tandem repeats (VNTR) region of the *MUC1* gene. This duplication produces a frameshift during translation, resulting in a new protein, MUC1fs. Cytosine duplication in VNTR of *MUC1* is the most frequent cause of ADTKD. Sequencing of GC rich VNTR of *MUC1* is extremely difficult. We performed MUC1fs immunostaining on urinary cell smears and various tissues in a group of positive and negative controls as well as in individuals from 37 ADTKD families that were negative for mutations in known ADTKD genes. Novel analytic methods were used to identify *MUC1* frameshift mutations. The sensitivity and specificity for MUC1fs immunostaining of urinary cell smears were 94.2% and 88.6%. Further genetic testing on 17 families with positive MUC1fs immunostaining revealed 6 families with 5 novel *MUC1* frameshift mutations that all predict production of the identical MUC1fs protein. We developed a noninvasive immunohistochemical method to detect MUC1fs in urinary cell smears that may be useful in the future for diagnostic testing after further validation. Production of the MUC1fs protein may be central to the pathogenesis of ADTKD-*MUC1*.

1314W

Whole exome sequencing is essential for defining emerging phenotypes and providing diagnoses in underserved populations. *S. Rego, R. Gallagher, B. Mendelsohn, P-M. Martin, U. Hodoglugil, D. Belefard, M. Sabbadini, J. van Ziffle, N. Risch, P-Y. Kwok, A.M. Slavotinek.* University of California San Francisco, San Francisco, CA.

There are many barriers to health care, including genetic testing, in underserved and underrepresented minority (US/URM) patients, among them cost and lack of insurance coverage. However, genetic testing, including whole exome sequencing (WES), has great potential to provide diagnoses in US/URM patients and also contributes data that may help to better characterize variants that have not been previously described. Here we present three US/URM patients (as defined by type of insurance, self-identification with an ethnic/racial minority group, or residence in a medically underserved area) who had clinical WES as part of an NIH-funded project to study the utility of WES in US/URM populations. An 8 year-old female with intellectual disability (ID) presented with acute weakness and fatigue and creatine kinase levels suggestive of rhabdomyolysis. She had a history of an abnormal newborn screen with inconclusive follow-up testing. She had low T4 and no history of seizures. Her parents are consanguineous. WES showed homozygosity for p.Arg86* in *TANGO2*—a newly identified cause of rhabdomyolysis, ID, seizures, abnormal thyroid studies and life-threatening cardiac arrhythmias. A 17 year-old female with a history of mild ID, hypotonia, macrocephaly, and supraventricular tachycardia was found to be heterozygous for the recurrent variant, p.Gln406* in *MECP2*. This patient has a learning disability rather than classical Rett syndrome. X-inactivation studies were normal. A 10 year-old male born with cryptorchidism and micropenis had a clinical course including abnormal hair texture, neonatal hypoglycemia, refractory epilepsy, and failure to thrive. WES revealed a hemizygous variant, p.Gln302fs*12, in *RNF113A*. This is the third patient described with a variant in *RNF113A* and establishes it as the cause of an emerging trichothiodystrophy syndrome. Two of the three genes (*RNF113A* and *TANGO2*) described above were either absent or uncommon on commercial gene panels, highlighting the need for WES in the diagnostic process. A third patient's results support a broader understanding of the phenotype for patients with *MECP2* variants. We conclude that all three patients required WES for a diagnosis due to the variable and emerging nature of many genetic syndromes, and that access to WES was the limiting factor in establishing a diagnosis. Such evidence is essential for acquiring insurance coverage in disadvantaged populations. .

1315T

New strategies for analyzing exomes from patients with rare and unknown disorders. *K. Schmitz-Abe^{1,2,3}, P. Agrawal^{1,2,3}.* 1) The Manton Center for Orphan Disease Research, Boston Children's Hospital, Boston, MA, USA; 2) Division of Newborn Medicine, Boston Children's Hospital, Boston, MA, USA; 3) Harvard Medical School, Boston, MA, USA.

The Gene Discovery Core (GDC) of The Manton Center for Orphan Disease Research at Boston Children's Hospital (BCH) has been enrolling families, mostly trios, with rare and unknown disorders for the last seven years. Of those enrolled patients, whole exome sequencing (WES) has been performed on 1738 individuals from 721 families without an initial clinical diagnosis. We aim to determine the genetic and molecular basis on those by re-analyzing the WES data. Reportedly, WES is able to determine the cause of disease in a third to half of the patients with genetic rare diseases. One of the reasons for the relatively low yield may be that we are missing important variants in the non-coding regions or copy number variants not detected by WES. Another important reason is the high number of rare coding pathogenic variants in candidate genes per family with scant information about them in the literature. This makes linking the genetic variant with phenotype extremely challenging, especially when we try to determine the functional effects using cellular or animal modeling. To overcome this issue, a) we re-processed the entire WES data using our own in-house bioinformatics and analysis pipeline (Variant Explorer Pipeline); b) we use 55 databases with important genetic discoveries to rank and filter variants (Roadmap, OMIM, Clinvar, MGI, PolyPhen-2, SIFT, Encode project, Roadmap, gnomG, etc); c) we use artificial neural network to optimize (weight) our inputs. The model was trained using reliable results of families with clinical diagnosis; and d) we sub-grouped the families with similar phenotypes and interrogate the data for candidate genes. Genes carrying mutations in several non-overlapping phenotypes were likely to be non-pathogenic and were removed from the candidate gene list. Using this approach, we have reduced the number of potential candidate genes by 73%. We are currently reviewing the reanalyzed data to determine the true candidates that can go for the functional determination and Matchmaker stage. We highly recommended processing all WES data from patients using the same pipeline, use adequately all the important genetic information available online and compare the results within families with variable phenotype.

1316F

Novel gene discovery through an undiagnosed disease program in East Asian. C. Law, C. Lam. Department of Pathology, The University of Hong Kong, Hong Kong, China.

Introduction: We had encountered numerous undiagnosed diseases in Hong Kong over the past decades and the disease spectrum is heterogeneous. With the advance of next-generation sequencing (NGS) techniques, it is now within our reaches to solve these undiagnosed cases using Clinical Whole Exome Sequencing (CWES) and Clinical Whole Genome Sequencing (CWGS). In this regard, we set up the first Undiagnosed Diseases Program (UDP) in Hong Kong and aim to end the diagnostic odyssey for patients suffering from rare conditions. **Methodology:** Proband-only CWES/CWGS approach was performed after consent over 100 cases and the subsequent bioinformatics analysis was done using our in-house algorithm. The CWES/CWGS findings, together with the clinical, biochemical and radiological features of the patients were interpreted by Pathologists qualified in Genetics. **Results:** Majority of the cases were contributed by known disease genes. Of-noted, some "rare diseases" were actionable. In addition, we also discovered novel genes, for example, *AK9* in congenital myasthenic syndrome (CMS), *EBF3* in Moebius syndrome, *CDC25B* for a syndrome with cataract, dilated cardiomyopathy and multiple endocrinopathies (CDCME), and *TTC21A* for Joubert syndrome. **Conclusions:** We recommended a proband-only CWES/CWGS approach for patients with undiagnosed disease for more than 3 months. The clinical interpretation of CWES/CWGS is would require understanding in clinical medicine, laboratory medicine and genetics. Therefore, the interpretation and clinical reporting should be handled by experience and qualified persons in clinical genomics.

1317W

Genomic sequencing reveals previously unrecognized phenotypes in a high proportion of those with unanticipated monogenic disease risk.

M.S. Lebo^{1,2,3,4}, T.W. Yu^{2,5,6}, S. Fayer⁷, C. Blout⁷, J.B. Murry¹, T.S. Schwartz², C.A. Genett⁸, W.A. Betting⁷, T.K. Truong⁵, I.A. Holm^{2,5}, P.B. Agrawal^{2,5}, R.B. Parad^{2,8}, K. Machini^{1,2,4}, K.D. Christensen^{2,7}, J. Krier^{2,7}, J.L. Vassy^{2,7,9}, O. Ceyhan-Birsoy^{1,10}, A.H. Beggs^{2,5}, H.L. Rehm^{1,2,3,4,11}, R.C. Green^{1,2,3,7}. 1) Partners HealthCare Personalized Medicine, Cambridge, MA; 2) Harvard Medical School, Boston, MA; 3) Broad Institute of MIT and Harvard, Cambridge, MA; 4) Pathology, Brigham and Women's Hospital, Boston, MA; 5) Division of Genetics and Genomics, The Manton Center for Orphan Disease Research, Boston Children's Hospital, Boston, MA; 6) Department of Neurology, Boston Children's Hospital, Boston, MA; 7) Department of Medicine, Brigham and Women's Hospital, Boston, MA; 8) Department of Newborn Medicine, Brigham and Women's Hospital, Boston, MA; 9) VA Boston Healthcare System, Boston, MA; 10) Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY; 11) Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA.

Background Identification and reporting of unanticipated findings from genome or exome sequencing (GS) is increasing, among both those who present with a suspected genetic condition and as a predispositional screen in apparently healthy individuals. Population analyses of the ACMG list of 59 genes (ACMG59, recommended as a minimum for secondary findings) reveal pathogenic or likely pathogenic (P/LP) variants in 1-3% of individuals. However, the implications of searching for unanticipated variants in disease-associated genes beyond the ACMG59 are less well understood. **Methods** In the MedSeq and BabySeq Projects, we analyzed ~5000 disease-associated genes to identify monogenic disease risk variants (MDR, defined as P/LP variants in genes associated with dominantly inherited monogenic disease or bi-allelic P/LP variants in genes associated with recessively inherited monogenic disease). Detailed phenotyping and family history analyses were conducted among those identified with MDRs through EHR review, questionnaires, clinical visits, and follow-up testing. **Results** GS was done on 269 individuals (110 adults and 159 infants), including 50 adults and 31 infants with a prior medical indication. Unanticipated MDRs were found in 16/110 (14.5%) adults and 18/159 newborns (11.3%), with 1 adult and 4 newborns having a finding in the ACMG59. Within the adult cohort, 4 MDRs (*EYA4*, *PPOX*, *RDH5*, *HFE*) prompted discovery of clinical features suggestive of a previously unrecognized genetic condition. Among the newborn cohort, 1 MDR (*ANKRD11*) provided an unsuspected molecular diagnosis for an observed phenotype, and 4 (*ELN*, *BTB*, *G6PD*, *GLMN*) prompted discovery of clinical features suggestive of a previously unrecognized genetic condition, including 3 with highly effective interventions. **Discussion** In this, the largest sample to date in which P/LP variants were extensively curated for the entire medical exome, unanticipated MDRs were discovered in 11-15% of individuals. Phenotyping of participants with MDRs revealed previously unrecognized but related clinical features in 9/34 (26.5%) of those in whom MDRs were detected. The remainder of those with MDRs in whom clinical features of the related condition were not seen may benefit from enhanced surveillance or have medical resources expended needlessly if they never develop the condition. Replication of these observations in larger sample sizes may inform the clinical advantage of broad screening in GS to reveal subclinical phenotypes.

1318T

Mutations in the CFAP-coding genes lead to male infertility with multiple morphological abnormalities of the sperm flagella. *W. Li, F. Zhang.* Obstetrics and Gynecology Hospital, State Key Laboratory of Genetic Engineering, Fudan University, Shanghai 200011, China.

Background: Sperm motility is vital to human reproduction. Malformations of sperm flagella can cause male infertility. Men with multiple morphological abnormalities of the flagella (MMAF) have abnormal spermatozoa with absent, short, coiled, bent and/or irregular-caliber flagella, which impair sperm motility. Genetic factors are involved in pathogenesis of MMAF. **Main Questions:** The known MMAF-associated genes are only responsible for approximately half of human cases. Novel pathogenic mechanisms still remain to be elucidated in MMAF. **Experimental Design:** We conducted genetic analyses using whole-exome sequencing and genome-wide comparative genomic hybridization microarrays in a multi-center cohort of 80 Chinese men affected by MMAF. Furthermore, the genome editing technology of CRISPR-Cas9 was employed to generate knockout mouse models for further functional investigations on our identified novel genes for MMAF. **Main Results:** Biallelic mutations in the genes encoding cilia and flagella associated proteins (CFAPs) were identified in 15 (19% out of 80) cases with MMAF, including seven in *CFAP43*, two in *CFAP44*, four in *CFAP65* and two in *CFAP69*. These CFAP-coding genes are specifically or preferentially expressed in testis. Knockout mouse models of these four genes manifest short, coiled flagella and other phenotypes consistent with human MMAF. **Conclusions:** Our experimental observations on human subjects and mouse models strongly suggest that biallelic mutations in *CFAP43*, *CFAP44*, *CFAP65* and *CFAP69* can cause sperm flagellar abnormalities and impair sperm motility.

1319F

Expanded analysis of Lyso-Gb3 and Lyso-Gb3 analogs for the diagnosis of Fabry disease. *M. Schultz, B. Netzel, D. Gavrilov, D. Matern, D. Oglesbee, S. Tortorelli, P. Rinaldo, K. Raymond.* Biochemical Genetics Laboratory, Department of Laboratory Medicine and Pathology Mayo Clinic, Rochester, MN 55905.

Background: Fabry disease (FD) is an X-linked lysosomal disorder caused by alpha galactosidase deficiency. We investigated the biomarker globotriaosylsphingosine (Lyso-Gb3) and related species (Lyso-Gb3 analogs) in a cohort of Fabry patients tested at the Mayo Clinic Biochemical Genetics Laboratory. Biomarkers can address challenges in the diagnosis of affected females with normal enzyme activity and in the interpretation of novel genetic variants. **Methods:** Lyso-Gb3 and eight analogs (mass differences -28, -12, -2, +14, +16, +18, +34, and +50amu) were measured by LC MS/MS from the serum of 59 FD patients (9 female, 50 male, 18 pediatric), and 214 normal control subjects (86, 128, 48). Patients were confirmed as having FD through enzyme testing and either mutation status, urine ceramide trihexosides, clinical presentation, or family history. **Results:** The median Lyso-Gb3 level in FD affected males was 64.05 ng/ml (1.49-113.82; 1st-99th percentile) and in FD females was 11.05 ng/ml (0.81-80.63; 1st-99th percentile); reference population was 0.38 ng/ml (0.16-0.73 ng/ml; 1st-99th percentile). No significant difference was noted between male and female Lyso-Gb3 levels in the non-FD, reference population. During analysis, several peaks were noted in the chromatogram for each analog mass. A total of 23 analogs were measured, (28(2), -12(3), -2(2), +14(4), +16(4), +18(3), +34(2), +50(3)) and analyzed. 22/23 analogs were elevated in all FD specimens, regardless of sex. Notably, in six patients with the p.A143T variant the Lyso-Gb3 analog values were slightly elevated, but in a pattern distinct from affected Fabry males. Three patients with the p.N215S late onset mutation had an elevated Lyso-Gb3 level and a pattern of analogs consistent with FD. **Conclusions:** In our cohort, the Lyso-Gb3 concentration and analog ratios clearly distinguished affected males and females from controls. This study adds to a body of work suggesting the Lyso-Gb3 biomarker can aid in the evaluation of genetic variants of unknown significance, especially in females where other biochemical testing can be normal.

1320W

Identification of protein features associated with Mendelian disease

variants. S. Iqbal^{1,2}, E. Perez-Palma³, J. Jespersen^{4,5}, P. May⁶, H. Heyne^{1,2}, K. Lage^{1,4}, A. Palotie^{1,2,7}, J. Cottrell¹, F. Wagner¹, M. Daly^{1,2,7}, A. Campbell¹, D. Lal^{1,2,3,8,9}. 1) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA; 2) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 3) Cologne Center for Genomics, University of Cologne, Cologne, Germany; 4) Department of Surgery, Massachusetts General Hospital, Boston, MA, USA; 5) Department of Bio and Health Informatics, Technical University of Denmark, Lyngby, Denmark; 6) Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg; 7) Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland; 8) Epilepsy Center, Neurological Institute, Cleveland Clinic, Cleveland, USA; 9) Genomic Medicine Institute, Lerner Research Institute Cleveland Clinic, US.

For Mendelian Disorders, a single genetic variant can cause the phenotype. While the variant discovery is progressing fast, elucidating the molecular consequences of variants is still a slow process. Although each protein has a specific role in the organism, related proteins have similar functions; thus, are likely to share variant intolerant and functionally essential sites. We explore whether automated variant annotation by features derived from three-dimensional (3D) protein structures can identify protein-class-specific features associated with patient variants. This approach can advance our understanding of how a single variant changes the protein and leads to etiology of monogenic disorders. For 1,331 disease-associated genes (>80%, 1,078/1,331 annotated in OMIM), we collect missense variants from the general population (gnomAD database, N=164,915) and patients (ClinVar and HGMD databases, N=32,924), and *in silico* transfer the variants onto >15k human protein 3D structures using an in-house mapping framework. We annotate the protein positions of variants with >40 structure-based features. We then group the genes into 24 protein classes based on their molecular function and perform association analyses of the structural features to population and patient variants. Our analyses show that in general, patient variants are less exposed to solvent ($p < 2.2 \times 10^{-16}$), perturb disulfide bonds ($p < 2.2 \times 10^{-16}$), hit ligand binding sites ($p = 1.5 \times 10^{-92}$), and stay at close proximity to phosphorylation sites ($p < 2.2 \times 10^{-16}$). Besides finding the shared features, we test the association of patient variants to all the structural features for each protein class and use the association signature to study the similarity and divergence of variant pathogenicity across protein classes. We show that patient variants of nucleic acid binding and transcription factor proteins obtain similar properties; however, these properties are different than that of receptor proteins. Only in the class of signaling molecules, patient variants are enriched in amino acids of flexible structure ($p = 8.0 \times 10^{-6}$), whereas the intra-protein hydrogen bond is specifically associated with variants in kinases ($p < 2.2 \times 10^{-16}$). In summary, we study the shared and unique protein features associated with patient variants across 24 protein classes which provide novel biological and mechanistic insights. Our protein-class-specific features can be annotated at scale, and can guide variant classification and targeted drug development.

1321T

Stratifying loss-of-function variants in genome sequencing cohort. A.E.

Katz¹, X. Liu², H. Shiferaw¹, J. Dudley¹, J.J. Johnston¹, W.S.W. Wong², J.E. Niederhuber², L.G. Biesecker¹, The Genomic Ascertainment Cohort (TGAC). 1) National Human Genome Research Institute, NIH, Bethesda, MD; 2) Inova Translational Medicine Institute, Inova Health Systems, Falls Church, VA.

Predicted loss of function (pLOF) variants are common in human genomes and determining their effect, if any, on phenotype can be challenging. Individuals who have undergone genome sequencing and can participate in subsequent studies represent an opportunity to investigate the phenotypic spectrum associated with select pLOF variants. We sought to characterize the pLOF variants in a large dataset for the purpose of targeted clinical evaluation. We restricted our search to genes for which haploinsufficiency has plausible phenotypic consequence by examining the 3,230 genes for which ExAC data suggest evolutionary constraint with a loss of function intolerance (pLI) >0.9. In our genome sequenced cohort (N=3,817) of individuals unselected for any phenotype from Inova Translational Medicine Institute's Childhood Longitudinal Cohort Study, we queried these genes for nonsense, canonical splice site disrupting, and frameshift variants. A variant allele frequency cutoff of 0.005 in both gnomAD and our cohort was applied to remove common variants. The resulting 2,704 pLOF variants were stratified for known disease-gene association using NHGRI's Clinical Genomic Database. 491 variants were in genes associated with autosomal dominant, X-linked or autosomal recessive disease, 170 variants were in genes associated exclusively with autosomal recessive disease, and the remaining 2,043 variants were in genes with no known disease association. Among these 2,043 variants, we found 105 variants were located in exons present in every transcript of its gene (as defined by Gencode's comprehensive annotation) and 475 variants in at least 50% of that gene's transcripts. None of the 475 prioritized variants were found in a homozygous state. Our filtering strategy represents a computational method by which variants with plausible clinical manifestations can be identified without using phenotypic information. Because individuals in our cohort are eligible for downstream secondary studies, next steps include clinical evaluation of individuals carrying rare pLOF variants of interest. Variants in genes with no known disease association represent an opportunity for novel gene-phenotype correlation discovery. Thus, ascertainment of study participants by genotype can be used to guide clinical research aiming to broaden our understanding of the phenotypic consequence of genomic variation.

1322F

Initiative on Rare and Undiagnosed Diseases in Pediatrics (IRUD-P) in Japan: Statistics of 1st stage for three years. Y. Matsubara¹, K. Hata¹, T. Kaname², K. Kosaki², *IRUD-P consortium*. 1) Research Institute, National Center for Child Health and Development, Tokyo, Tokyo, Japan; 2) Center for Medical Genetics, Keio University School of Medicine, Tokyo, Japan.

The Initiative on Rare and Undiagnosed Diseases (IRUD) is a national consortium designed to help patients and their families suffering from rare and undiagnosed disease conditions in Japan. The project started in July 2015. The pediatric version of IRUD (IRUD-P) is coordinated by dual centers, the National Center for Child Health and Development (NCCHD) and Keio University. The IRUD-P developed a nation-wide network for patient recruitment involving 17 regional core clinical centers, and mainly performed whole exome sequencing on children with undiagnosed diseases and their parents. For three years, more than 3,500 patients, who passed the first screening in the core clinical centers, were consulted to the IRUD-P centers in total. Specimens accompanied with medical information (n=7,100) were collected from patients and their families, mainly in trios, and sent to the centers. In NCCHD, of 881 patients analyzed, genetically confirmed diagnosis in 330 patients (diagnostic yield was 37.5%). Of the confirmed patients, 18.6%, 76.5% and 3.5% were found in known gene with reported variants, known gene with novel variants, and novel genes, respectively. Patient cohort represents children with a wide range of undiagnosed disorders (malformation syndromes 37.4%, neuromuscular disorders 17.4%, skeletal and connective tissue diseases 8.7%, endocrine disorders 7.4%, cardiovascular diseases 7.4%, renal and/or liver diseases 6.4%, immunologic or hematologic disorders 3.0%, otolaryngologic/ophthalmologic diseases 1.5%, others 10.9%). For the next turn, IRUD-P has combined with adult version of IRUD and the consortium started disease model analysis for the remaining undiagnosed patients. The IRUD also built international collaboration and data sharing on rare and undiagnosed diseases.

1323W

The Rare Genomes Project: Improving our ability to diagnose rare genetic conditions through a nationwide partnership with families. C.E. Williamson, A.H. O'Donnell-Luria, A. Kern Lovgren, M. O'Leary, E. England, L. Pais, Z.M. Valivullah, K.R. Chao, L. Xu, J.K. Goodrich, M.H. Wojcik, J. Chang, M. Getaneh, I. Ndon, G. VanNoy, K. Blakeslee, H. Snow, N. Lopez Zalba, S. Maiwald, E.O. Baker, J. Hendry, A. Zimmer, S. Baxter, B. Weisburd, D.G. MacArthur, H.L. Rehm. Broad Center for Mendelian Genomics, Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA.

A majority of patients in the US with a suspected rare genetic disease do not yet have a molecular genetic diagnosis. A diagnosis can have many medical implications including management strategies, prognoses, and family planning using preimplantation genetic diagnosis and, in rare instances, can even identify effective treatments. Many families reach the end of their clinical diagnostic odyssey without answers, typically after exome sequencing or when their insurance will not cover further genetic testing options. The Rare Genomes Project (RGP; raregenomes.org), launched in June 2017, is a direct-to-families research study leveraging online participation and partnerships with patient advocacy groups to dramatically increase the reach of rare disease genomic research by engaging families as stakeholders at each step. English-speaking families who live in the United States and have an undiagnosed suspected monogenic disease are eligible for the study. We enroll families by phone or video consent, mail kits for blood collection and perform whole genome sequencing followed by standard variant and structural variant analysis. Despite advances in electronic health medical systems, it is challenging to efficiently extract the pertinent information from medical records. In our pilot, we are focusing on patient- and family-provided medical information through surveys with linked Human Phenotype Ontology terminology. We describe results from our primary analysis where we are able to make accurate diagnoses based on participant-provided information, and these diagnoses are confirmed through secondary review of medical records. In cases where we identify the genetic cause of the participant's condition, we performed clinical (CLIA) validation in coordination with a current treating physician identified by the family and provide a genetic testing report to the participant and their doctor. A central goal of this project is broad data sharing, so sequence and phenotype data will be rapidly shared through the Broad's Data Use Oversight System (DUOS, duos.broadinstitute.org) and other means. Our goal is to provide a proof of principle of the value of a patient-engaged approach to diagnosis and gene discovery in rare disease while creating a shared genomic dataset of undiagnosed families to empower the rare disease community.

1324T

Duarte galactosemia: Shedding new light on an old problem. *J.L.*

Fridovich-Keil¹, G. Carlock¹, S.T. Fischer¹, M.E. Lynch², N.L. Potter³, C.D. Coles^{2,4}, M.P. Epstein¹, J.G. Mülle¹, J.A. Kable². 1) Department of Human Genetics, Emory University School of Medicine, Atlanta, GA; 2) Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, GA; 3) Department of Speech and Hearing Sciences, Washington State University, Spokane, WA; 4) Department of Pediatrics, Emory University School of Medicine, Atlanta GA.

Duarte galactosemia (DG) is a variant form of the potentially lethal disorder classic galactosemia (CG). For decades, infants with DG have been identified by newborn screening but there has been no uniform standard of care regarding diet. Specifically, it has been unknown (1) whether infants with DG experience increased risk of developmental complications later in childhood, and (2) whether milk exposure in infancy, which causes elevated galactose metabolites, is associated with outcome severity and therefore should be contraindicated. To address this question we conducted a study of infant diet and developmental outcomes of 206 children with DG and 144 unaffected sibling controls. We recruited children with DG and their siblings, ages 6-12, in 18 states where they were identified as newborns. Diet in infancy was ascertained by retrospective parent survey and diagnosis was confirmed by DNA testing. Child developmental outcomes were assessed directly by a team of trained professionals using established instruments. Following testing, data sets were divided randomly into independent discovery (n=87) and validation (n=263) cohorts for analysis. We tested a total of 73 outcome measures for possible association with DG status in the discovery set, and then tested the top 10 of these outcomes for possible association with DG status in the validation set. We also tested the top 10 outcomes from the discovery set for possible association with milk exposure in infancy among children with DG in the validation set. We found no significant associations in any of these tests. Our results demonstrate no significant difference in developmental outcomes tested between children with DG and controls, or among children with DG as a function of exposure to milk in infancy. This study provides a robust and long-needed foundation to empower evidence-based decisions by healthcare providers and families of newborns with Duarte galactosemia choosing what to feed their baby.

1325F

Oligogenic obesity: An explanation for apparent variable penetrance of an *SH2B1* mutation. *S. Almansoori^{1,2}, Al. Blakemore^{1,2}, N. Lessan³.* 1) Imperial College, London; 2) Brunel University London; 3) Imperial College London Diabetes Centre.

Obesity is a complex, heterogeneous disorder: including a number of monogenic or syndromic subtypes. Here, we present the case of a 22 year old man of South Asian origin, with a history of progressive weight gain starting in his mid-teens, coupled with poor school performance. He had generalised obesity (BMI 38.6 kg/m²) with profound hyperinsulinaemia and prominent acanthosis nigricans, but with normal blood pressure and no Cushingoid features. There was no family history of note. The proband was found to carry a mutation in *SH2B1*, C539T/p.S180F, which was predicted to be deleterious, with a CADD score of 23.6, and, at first considered likely to be causative. On investigation, however, the non-obese father was revealed to carry the same mutation, raising questions about penetrance. Examination of exome sequencing data from the proband revealed two further potentially causative mutations A3539T/p.D1180V in *MBD5* and C3022T/p.L1008F in *POGZ*. These variants are also predicted to be deleterious with CADD scores of 22.9 and 24.2, respectively. These findings suggest that an oligogenic inheritance model might be responsible for the phenotype. We urge caution in interpretation of sequencing results from individual genes which, in this case, could have given rise to inaccurate genetic counselling.

1326W

Interallelic interactions in autosomal recessive disorders: A population-genetic and a molecular approach. A. Miko^{1,2}, A. Kaposi¹, G. Schay¹, E. Balogh^{1,2}, D. Seidl¹, K. Schnabel¹, D.K. Menyhard³, K. Tóty^{1,2}. 1) MTA-SE Lendület Nephrogenetic Laboratory, Budapest, Hungary; 2) Semmelweis University, 1st Department of Pediatrics, Budapest, Hungary; 3) MTA-ELTE Protein Modeling Research Group, Budapest, Hungary; 4) Semmelweis University, Department of Biophysics and Radiation Biology, Budapest, Hungary.

We recently identified the first variant (*NPHS2*, R229Q), that is pathogenic only when trans-associated to specific mutations in an autosomal recessive (AR) disorder. We aimed to identify other incompletely penetrant variants (IPV), and to characterize the R229Q podocin. We collected clinical and genotype data from 9,109 Caucasian patients with biallelic HGMD mutations in 17 genes (*ASL*, *ATP7B*, *CAPN3*, *CFTR*, *CTNS*, *DHCR7*, *GAA*, *GALNS*, *GALT*, *IDUA*, *MUT*, *NPHS1*, *NPHS2*, *PAH*, *PKHD1*, *PMM2*, *SLC26A4*) from the medical literature. The penetrance of each pathogenic variant ($n=2141$) was calculated as its enrichment in the patient population vs. the general population as compared to that of the loss-of-function (LOF) mutations. Out of the 78 variants that were frequent enough to assess their penetrance, we found 23 (28%) to be incompletely penetrant, including the three known IPVs (*NPHS2* R229Q, *CFTR* R117H and L997F). Based on the associated phenotype, most of the IPVs are hypomorphic (16/22, 73% vs. 19/56, 34% of completely penetrant variants, $p=0.002$). Only the *NPHS2* R229Q variant was significantly associated to specific mutations in the patient population, indicating interallelic interaction as the cause of incomplete penetrance. In order to understand how R229Q makes prone podocin to be subject of a dominant negative effect, we compared the binding capacity of R229Q and wt podocin. Podocin variants with and without dominant negative effect ($n=11$ types) were transiently expressed in HEK293 cells, extracted, eluted and stained by either Alexa 488 and 555 maleimides. Podocin variants were mixed with wt and R229Q podocin to allow heterooligomer-formation and FRET was measured by fluorescence lifetime spectrometer. We found a significantly higher FRET efficiency in the R229Q (34% [30-36%], median [quartiles]) than in the wt heterooligomers (13% [4%-22%], $p=0.01$), indicating a higher binding capacity of R229Q podocin. The highest FRET efficiencies were found in the pathogenic R229Q heterooligomers (A284V, R291W, A297V). In conclusion, among the variants with a high allele frequency, we found 23 to be incompletely penetrant, but only one (R229Q) to be subject of interallelic interactions. We found an increased binding capacity of the R229Q podocin which can partially explain its intracellular retainment in the pathogenic associations. Supported by MTA-SE Lendület Research Grant (LP2015-11/2015), NKFI/OTKA K109718, K116305, KH125566 and MedinProt Synergy grant VII.

1327T

Residual function of cystic fibrosis mutants predicts therapeutic response. S.T. Han¹, A. Rab², M.J. Pellicore¹, E.F. Davis¹, A.F. McCague¹, T.A. Evans¹, A.T. Joynt¹, Z. Lu¹, Z. Cai², K.S. Raraigh¹, J.S. Hong², D.N. Sheppard³, E.J. Sorscher², G.R. Cutting¹. 1) Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Department of Pediatrics, Emory University School of Medicine, Atlanta, GA; 3) School of Physiology, Pharmacology and Neuroscience, University of Bristol, United Kingdom.

Treatment of individuals with cystic fibrosis (CF) has been transformed by modulator therapies that target select molecular mechanisms affecting function of the CF transmembrane conductance regulator (CFTR). However, allelic heterogeneity and low frequency of alleles present significant challenges to therapeutic classification of pathogenic variants. Approximately 1,650 variants have been identified in individuals with CF; most variants have been observed in fewer than 10 individuals worldwide and only 39 genotypes are FDA approved for CFTR modulator therapy. To generate data useful for expansion of treatment eligibility, we stably expressed 43 *CFTR* variants from a single integration site in the genome of CF Bronchial Epithelial (CFBE) cells and evaluated CFTR function by measuring transepithelial ion transport. We found that most variants demonstrated a modest elevation of CFTR function in response to all treatments tested. Importantly, we found that the magnitude of therapeutic response was highly correlated with residual CFTR function for the potentiator ivacaftor ($r = 0.96$), the corrector lumacaftor ($r = 0.83$), and the ivacaftor-lumacaftor combination ($r = 0.93$). Response of a partially overlapping set of 16 variants studied in a similar manner using Fisher Rat Thyroid (FRT) cells, a robust system for studying CFTR biology, showed nearly identical correlations. As expected, only a small subset of variants demonstrated statistically significantly greater response to specific treatments, presumably due to direct interaction between modulator action and molecular mechanism of disease. Furthermore, 40/43 variants studied in CFBE cells and 13/16 variants studied in FRT cells demonstrated highest response to combination therapy. The variants studied here represent 87% of individuals in the CFTR2 database who carry at least one missense variant, indicating that most individuals with CF caused by missense variants are likely to demonstrate modest response to currently available modulator therapy while a small fraction will show significantly elevated responses. Additionally, our results indicate that most individuals with missense variants should derive greatest benefit from combination therapy and we found no variant for which monotherapy was more effective than combination therapy. These findings provide a path to expanding modulator therapy to a significant number of individuals with CF who have rare genotypes that cannot be studied by clinical trials.

1328F

Reassessing the pathogenicity of *RYR1* autosomal dominant variants with modified ACMG criteria. D. Ng¹, S.G. Gonsalves¹, J.J. Johnston¹, L.G. Biesecker^{1,2}. 1) Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 2) NIH Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD.

RYR1 is one of the 59 genes recommended by the American College of Medical Genetics and Genomics/Association for Molecular Pathology (ACMG-AMP) for opportunistic screening. *RYR1* variants (n=289) listed in the European Malignant Hyperthermia Group (EMHG), Human Gene Mutation Database (HGMD), or ClinVar and designated as a disease-causing mutation, pathogenic or likely pathogenic and associated with malignant hyperthermia susceptibility (MHS), central core disease or a dominant *RYR1*-associated phenotype(s) were assessed for pathogenicity using modified ACMG-AMP criteria and the Bayesian calculator of Tavtigian et al. Select criteria were modified to better fit MHS, i.e., BS1 criteria was invoked if the minor allele frequency (maf) >0.0004. PVS1 was not used as there is no evidence for *RYR1* loss of function as a mechanism for MHS. We used the Bayesian calculator to give a posterior probability for pathogenicity: pathogenic ≥ 0.99 , likely pathogenic ≥ 0.90 - < 0.99 , and VOUS to benign < 0.90 . Among the 289 variants analyzed, 52 were designated pathogenic, 75 were designated likely pathogenic, and 162 were in the VOUS to benign category. Of the 42 *RYR1* "diagnostic mutations" listed on the EHM website, 34 were designated pathogenic, 5 were likely pathogenic, and 3 were in the VOUS to benign category per modified ACMG-AMP criteria. Of the 3 EHM variants in the VOUS to benign category, p.Arg530His scored in the high VOUS (posterior probability 0.82), while p.Glu2348del and p.Ile403Met scored low (posterior probability 0.05) because of published functional evidence indicating no damaging effect (BS3). We tailored existing ACMG-AMP criteria to MHS for reassessing *RYR1* variants as an initial step in developing a disease-specific set of guidelines so that assessment for pathogenicity can be standardized and to facilitate secondary findings analysis. We used the Bayesian analysis as a mathematical basis of the pathogenicity scoring. This is a first attempt to make judgement of *RYR1* variant pathogenicity more explicit and scientifically rigorous.

1329W

Preliminary findings from the Colorado Center for Personalized Medicine Biobank. N. Rafaels¹, T. Brunetti¹, A. Roff¹, S. Wicks¹, K. Barnes¹, K. Crooks^{1,2}. 1) Colorado Center for Personalized Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO; 2) Department of Pathology, School of Medicine, University of Colorado, Aurora, CO.

The Colorado Center for Personalized Medicine (CCPM), established in 2014, has created a CLIA-certified Biobank that, to date, has enrolled over 65,000 adult subjects from across the University of Colorado Health (UCHealth) system and collected blood samples from over 28,000 subjects. CCPM has developed an automated pipeline that enables high-throughput processing of the Illumina Infinium CU Custom MEGA-Ex genotyping array with 45,000 custom variants added with input from 32 principal investigators from the Anschutz Medical Campus, including over 200 additional pharmacogenetic variants. We present preliminary findings from the first 4,230 samples genotyped. Genotyping was performed on 47 96-well custom plates on the Illumina iScan, with 6 controls each, selected from Coriell to maximize genotype diversity of clinically relevant variants for validation. GTC files and sample manifests were then automatically uploaded into BC|Platforms, BC|PIPE, where genotype calling and quality control (QC) analysis was performed, bypassing Illumina GenomeStudio. We employed the following QC cutoffs for this analysis: callrate < 98.5%; 10% GC score < 0.39; LOGR_DEV > 0.4; F(inbreeding on X chromosome) for males > 0.7 and females < 0.3; and $-0.2 < F(\text{inbreeding autosomes}) < 0.3$. In addition, any subjects identified to have a bone marrow transplant were also removed. In total, 4,161 samples passed our QC metrics. In summary, 65.0% of subjects identified as female, and 75.4% of subjects identified as non-Hispanic and Caucasian. From the list of 59 ACMG genes, 36 individuals have a pathogenic or likely pathogenic variant confirmed by multiple submitters or expert panel in CLINVAR for autosomal recessive diseases in OMIM, 14 of which were in BRCA1 or BRCA2 genes. Also, we discovered poor and intermediate metabolizers of CYP2C9 (3, 452); CYP2C19 (109, 1085); CYP3A5 (114, 696); DPYD (0, 4); and SLC01B1 (80, 1064), based on PharmGKB guidelines. Finally, we found 326 subjects with either Factor V Leiden or Prothrombin thrombophilia mutations and 1408 subjects with either C282Y or H63D hemochromatosis mutations. Genotyping at the CCPM Biobank is ongoing with the first 30,000 subjects to be genotyped on the current Illumina chip. Additional samples will be run on a new custom Illumina MEGA Global chip with additional custom content. Work to return clinical and pharmacogenetic results to patients, and connecting genotype data to patients' electronic health records is underway.

1330T

Discovery of URAT1 and GLUT9 novel variant in hypouricemia subjects using whole exome sequencing analysis. S. Cho, C.A. Winkler. Molecular Genetic Epidemiology Section, National Cancer Institute (NCI) / NIH, Frederick, MD.

Objective. Renal hypouricemia is a rare disorder associated with genetic mutant of renal transporters. Two types are currently reported (Type 1 (OMIM: 220150) and Type 2 (OMIM: 612076)). Because of the low prevalence and asymptomatic presentation in the initial stage, genetic studies are lacking. Here, we investigated the genetic background of renal hypouricemia using whole exome sequencing. **Methods.** Korean cases (N= 31) with extreme hypouricemia (<1.2mg/dl) were selected from an urban cohort of 179,381 subjects; selection criteria included 1) abstinence from alcohol or smoking and 2) an absence of underlying conditions (i.e., hypertension, diabetes and taking anti-hypertensive medication). Whole-exome sequencing was performed for the discovery of diagnostic markers. Two known genetic marker (*SLC22A12* c.774G>A (p.Trp258Stop) and *SLC22A12* c.269G>A (p.Arg90His)) were identified, explaining 90.32 % (28/31) renal hypouricemia in Koreans. We used SNaPshot for the 2 *SLC22A12* SNPs and screened 50 additional hypouricemia subjects from the Korean Cancer Prevention Study. 47 carried known *SLC22A12* markers; the unexplained three hypouricemic cases were whole-exome sequenced (WES). **Results.** In total, we conducted WES of 34 individuals with extreme hypouricemia. 82.35 (28/34)% carried known hypouricemia variants. 844~1238 missense variants were identified after filtering out common variants (>1%) in each subject. 5 novel variants of *SLC22A12* (c.408C>A (p.Asn136Lys), c.674C>A (p.Thr225Lys), c.851G>A (p.Arg284Gln), c.1253T>G (p.Leu418Arg) and c.1285G>A (p.Glu429Lys)) were found. 1 novel variant c.463A>G (p.Met155Val) of *SLC2A9* was identified. Only one compound heterozygote of *SLC2A9* was identified (c.1138C>T (p.Arg380Trp) and c.463A>G (p.Met155Val)). Homology modeling confirmed that all newly discovered variants are functionally related to uric acid transport. After filtering out known genes of renal hypouricemia (*SLC22A12* and *SLC2A9*), 6 unsolved patients remained. *ASB12*, *NEB* and *LRCH2-RBMXL3* were overlapping genes for 6 unsolved patients. **Conclusion.** This is the first attempt to determine the value of genetic diagnostic screening of patients for hypouricemia in the clinical setting. Screening of just 2 SNPs (p.Trp258Stop and p.Arg90His) identified 75/81 (92.6%) of patients with hypouricemia. Early identification and intervention of hypouricemia (hydration and cautious exercise) may prevent acute kidney injury.

1331F

Inherited retinal dystrophy (IRD) pedigrees with complex genetic causality. P. Biswas¹, A. Soto Hermida¹, K. Branham², H. Matusi³, A. Chekuri⁴, A. A. Guru⁵, A. Alapati⁶, B. Huang⁷, A. L. Villanueva⁸, S. Riazuddin^{5,6}, J. R. Heckelively⁹, P. A. Sieving⁷, J. L. Duncan⁸, S. Amer Riazuddin⁹, K. A. Frazer¹⁰, R. Ayyagari¹. 1) University of California San Diego, La Jolla, CA, USA; 2) Kellogg Eye Center, University of Michigan, Ann Arbor, MI, USA; 3) Institute for Genomic Medicine, University of California San Diego, San Diego, CA, USA; 4) Mejora Vision MD/Virtual Eye Care MD, Mérida, Yucatán, México; 5) National Centre of Excellence in Molecular Biology, University of Punjab, Lahore, Pakistan; 6) National Centre for Genetic Diseases, Shaheed Zulfiqar Ali Bhutto Medical University, Islamabad, Pakistan; 7) National Eye Institute, NIH, Bethesda, MD, USA; 8) Ophthalmology, University of California San Francisco, San Francisco, CA, USA; 9) Department of Ophthalmology, The Wilmer Eye Institute, Johns Hopkins University School of Medicine, Baltimore, MD, USA; 10) Department of Pediatrics and Rady Children's Hospital, Division of Genome Information Sciences, University of California, San Diego, La Jolla, CA, USA.

Purpose: To identify genetic causality of retinal degeneration by using whole genome sequencing (WGS). **Methodology:** Whole genome sequencing was performed on 213 members of 55 pedigrees, using Illumina HiSeqX10. The reads were aligned to hg19 and variant calling was performed using GATK. The genotyping quality of single nucleotide variants (SNVs) and indels was assessed using the variant quality score recalibration approach implemented in GATK. The annotation was done using SnpEff, PolyPhen2 and CADD score and copy number variations (CNVs) were called using GenomeSTRiP and SpeedSeq. Segregation analysis of the candidate variants was carried out using Sanger sequencing followed by functional evaluation of identified novel gene variants. Mutations found in a novel gene *IFT88*, were introduced in mammalian cells using CRISPR-Cas9 system to evaluate the impact of the identified variants. **Results:** Analysis of 55 pedigrees identified 30 novel and 11 reported variants in IRD associated genes in 23 pedigrees (42%). Among these, 10 structural variants were detected in 9 pedigrees and a de-novo variant was found in one pedigree. IRD due to more than one gene involvement was noted in eight large pedigrees with multiple consanguineous marriages. A novel causative variant was found in the promoter region of *TMEM216* in a single pedigree. Interestingly, affected individuals from four additional pedigrees inherited causative mutations in multiple genes. Potential candidate causative variants were detected in 9 novel genes segregating with disease (16%). One of the novel candidates, *IFT88* is known to be localized to the cilium in photoreceptors. Two heterozygous causal variants, p.R266* and p.A568T in *IFT88* segregated with the phenotype. HeLa cells with *IFT88* mutations introduced using CRISPR-Cas9 system demonstrated nonsense-mediated decay of the mutant p.R266*-*IFT88* transcript, mis-localization of p.A568T-*IFT88* and abnormal ciliogenesis. **Conclusion:** WGS of 55 pedigrees found causative mutations in 32 pedigrees (58%). The findings reveal the involvement of CNVs, de-novo mutation, mutations in multiple genes and novel gene involvement as the underlying cause of IRD in our cohort. Functional evaluation demonstrated the potential mechanism underlying IRD due to mutations in the novel ciliary protein *IFT88*. Potential involvement of novel genes, non-coding region mutations or structural variants in known genes may be responsible for IRD in the remaining 23 pedigrees.

1332W

Challenges to increasing the rate of Mendelian disease gene discovery.

J.X. Chong¹, J.C. Carey², A. Hamosh^{3,4}, D.A. Nickerson⁵, M.J. Bamshad^{1,5}, *Baylor-Hopkins Center for Mendelian Genomics, University of Washington Center for Mendelian Genomics*. 1) Department of Pediatrics, University of Washington, Seattle, WA; 2) Department of Pediatrics, University of Utah, Salt Lake City, Utah; 3) McKusick-Nathans Institute of Genetic Medicine; 4) Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, MD; 5) Department of Genome Sciences, University of Washington, Seattle, WA.

The introduction in 2010 and rapid adoption of exome sequencing (ES) to identify genes for Mendelian conditions (MCs) was a disruptive innovation that replaced incumbent processes for gene discovery and syndrome delineation. Prior to ES, gene discovery was driven by positional cloning, which requires information about the genomic location and function of a candidate gene, the phenotype, or both. Computational approaches based on ES required neither. Accordingly, application of ES increased the rate of gene discovery for MCs by ~50% and from 2013 to 2018, genes underlying ~1,500 MCs were reported, 40% of which were newly-delineated MCs. Moreover, ~40% (~1,500/3,520) of all the genes now known to underlie a MC have been reported since 2010. Yet, the overall number of MCs for which the genetic basis was reported peaked between 2012-2015 and has declined slightly each year thereafter. Prior to 2010, delineation of new MCs was virtually always "phenotype-driven," that is, a MC was characterized prior to the discovery of a "causal gene." In contrast, among 266 newly delineated MCs added to OMIM in 2013, for only 17 (6%) was the genetic basis unknown. For the vast majority of new MCs reported in 2013 (249 or 94%), syndrome delineation was driven by gene discovery (i.e. "genotype-driven"). By 2017, genotype-driven syndrome delineation accounted for 100% of new MCs added to OMIM. If, as is widely accepted, perturbation of most human genes results in a MC and given that only ~3520 (18%) genes are, to date, known to underlie a MC, delineation of MCs associated with the majority of human genes will be genotype-driven. These observations suggest that even if the current rate of gene discovery for MCs remained constant, it would take up to five decades to identify genes underlying the ~3000 remaining unexplained, known MCs and for genotype-driven delineation to discover ~10,000 new MCs. The factors limiting the rate of MC gene discovery are myriad. Major reasons are the much slower co-evolution of the infrastructure and regulatory framework to share genomic data openly and at scale world-wide (hundreds of putative gene discoveries are unreported), limited access to high-throughput functional validation of candidate variants, and the challenges (insufficient resources, lack of organized efforts) of ascertaining and deeply phenotyping families with high priority candidate genotypes. We discuss various ways to maintain, if not increase, the rate of gene discovery for MCs.

1333T

Stratification by organ systems and phenotypes in rapid genomic testing results in differences in diagnostic rates.

T. Wong, M. Wright, M. Tokita, K. Ellsworth, S. Chowdhury, S. Nahas, D. Dimmock, S. Kingsmore. *Rady Children's Institute for Genomic Medicine, San Diego, CA.*

Rapid genomic testing in neonatal and pediatric intensive care units has resulted in improved diagnostic rates compared to current standard of care, and has the potential to impact healthcare management and decision making. The diagnostic rates from genomic testing can range from 10% to 50%, dependent on the study and patient population. In this study, we sought to investigate the diagnostic rates of rapid genomic testing by organ system and phenotype. Peripheral blood DNA from 391 consenting families were characterized by whole genome sequencing (WGS) or whole exome sequencing (WES). Variant prioritization was performed using manually curated Human Phenotype Ontology (HPO) terms that were used for analysis via the Fabric Genomics Interpretation Platform. Cases were classified by affected organ systems (e.g., cardiovascular, nervous, digestive, etc.) and by phenotype (e.g., seizures, respiratory distress, hypotonia, etc.). The overall diagnostic rate across the entire cohort was 32.7%. When parsed by affected organ system, the diagnostic rate ranged from 20% for cases with a reproductive or growth phenotype to over 40% for cases reporting an immune or muscular phenotype. Diagnostic rate also varied by phenotype, including conditions such as congenital diaphragmatic hernia (9.1%, n=11), epilepsy/seizures (31.0%, n=58), and immunodeficiency (75.0%, n=8). These findings highlight the difference in diagnostic rates across disease populations in this rapid genomic testing cohort, and an exploration of how automated phenotyping methods may enhance this process will be explored.

1334F

FMR1 premutation women: High rate of reported co-morbid health conditions with an earlier onset among those with FXPOI. *E.G. Allen¹, K. Charen¹, H. Hipp², L. Shubeck¹, A. Amin¹, W. He¹, S.L. Sherman¹.* 1) Human Genetics, Emory University, Atlanta, GA; 2) Department of Gynecology and Obstetrics, Emory University, Atlanta, GA.

Fragile X-associated primary ovarian insufficiency (FXPOI) is seen in about 20% of women who carry an *FMR1* premutation (PM) allele (55-200 CGG repeats). These women develop hypergonadotropic hypogonadism with amenorrhea before the age of 40. A non-linear association with repeat size in the PM range and risk for FXPOI has been documented: those with a mid-range repeats are at highest risk (~70-100 CGG repeats). Importantly, not all female carriers with 70-100 repeats experience FXPOI. We have collected general medical history information along with reproductive histories on 337 women with a PM to test the hypothesis that co-morbid conditions are more frequent among PM carriers with FXPOI compared with those without FXPOI. Of the 337 women, 81 were classified as having FXPOI (amenorrhea < age 40) and 159 were classified as non-FXPOI (still cycling or onset of menopause after age 40). The table below lists conditions that were reported among all PM carriers at the highest frequencies. Notably, anxiety, depression, tension headaches, and migraines were reported by about one-third of all of the women. Greater than 29% of women reported at least one autoimmune disorder (specifically fibromyalgia, Raynaud's, Sjogren's, Lupus, multiple sclerosis, type 1 diabetes, rheumatoid arthritis, or psoriasis). Anxiety was the only condition seen at a higher frequency among women with FXPOI, unadjusted for age at evaluation. Using survival analysis to examine the age of diagnosis of FXPOI, adjusting for the age at evaluation, women with FXPOI compared with women without FXPOI had an earlier age at onset for anxiety, depression, tension headaches, hypothyroidism, fibromyalgia, and autoimmune disorder (see table). The age of diagnosis for anxiety was significantly associated with age at FXPOI or menopause ($p=0.001$; $r^2=0.22$). Overall, our data indicate that women with a PM report a high rate of multiple co-morbid medical conditions, and the age of onset is frequently earlier for women with FXPOI.

Condition	All women (337) %	FXPOI (81) %	No FXPOI (159) %	Difference in frequency between groups (χ^2 p-value)	Difference in age of onset between groups (survival analysis p-value)
Anxiety	38.1	45.7	31.4	0.03	0.005
Depression	35	33.3	34.6	NS	0.033
Tension headache	31.8	33.3	28.3	NS	0.023
Migraines	32.6	33.3	26.4	NS	NS
Hypothyroidism	17.8	24.7	18.2	NS	0.027
Fibromyalgia	12.2	17.3	10.1	NS	0.005
Inflammatory Bowel Syndrome	20.5	21	18.2	NS	NS
Sleep problems	28.5	33.3	30.2	NS	NS
≥ 1 Autoimmune disorder	29	37	28.7	NS	0.038
NS=not significant					

1335W

A systematic analysis of De Novo mutations in 837 trios with clinical suspicion of Mendelian phenotypes. *Z.H. Coban Akdemir¹, X. Song¹, S.N. Jhangiani², T. Gambin³, D.M. Muzny², E. Boerwinkle^{2,4}, R.A. R.A.^{1,2}, J.E. Posey¹, C.M.B. Carvalho¹, J.R. Lupski^{1,2,5,6}, Baylor-Hopkins Center for Mendelian Genomics.* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA; 3) Institute of Computer Science, Warsaw University of Technology, Warsaw 00-665, Poland; 4) Human Genetics Center, University of Texas Health Science Center at Houston, TX 770530, USA; 5) Texas Children's Hospital, Houston, TX 77030, USA; 6) Department of Pediatrics, Baylor College of Medicine, Houston, Texas 77030, USA.

Widespread availability of whole exome sequencing (WES) has enabled systematic identification of *de novo* mutations in personal genomes in both research and the clinic. Recent studies applying clinical WES show that 60-75% of the molecular diagnostic rate in sporadic cases could be attributed to *de novo* mutations. Moreover, in the research setting, *de novo* variants have been shown to contribute extensively to multiple Mendelian phenotypes. However, there remains a lack of systematic identification of *de novo* variants in child-parent trios wherein the sample of sporadic probands presents with a wide range of Mendelian phenotypes. To systematically assess the association of *de novo* variants with Mendelian phenotypes, we examined data from ~8000 exomes presenting with a wide variety of Mendelian disease traits generated at Baylor College of Medicine through Baylor-Hopkins Center for Mendelian Genomics. Our analysis in 837 trios resulted in identification of 3,043 *de novo* coding SNVs/indels (~3.63/trio) detected by an in-house developed tool, DNM-Finder. Out of those, 387 (12.41%) were annotated as truncating variants present in 319 genes, of which ~23% were defined as loss-of-function (LoF)-intolerant in ExAC (pLI score ≥ 0.9). We identified 91 genes with more than 1 potentially damaging *de novo* variant in unrelated individuals. Those include recently published novel disease genes: *DVL1* (MIM 616331), *NALCN* (MIM 616266) and *PURA* (MIM 616158). Computational detection of *de novo* CNVs (≥ 100 Kb) in 837 trios using XHMM revealed 102 (~0.12/trio) and 182 (~0.217/trio) *de novo* deletions and duplications, respectively. Those deletions and duplications include 71 and 107 distinct genes, respectively, of which ~30% (21) and ~28% (30), respectively, were defined as LoF-intolerant. Taken together, *de novo* variant analyses in the Baylor-CMG enable novel disease gene discovery and improved understanding of the genetic basis of Mendelian phenotypes and gene function allowing insight into the biology of disease.

1336T

Using phenome risk scores to systematically test the hypothesis that genetically determined expression of Mendelian genes associate with Mendelian phenotypes. J. Brown¹, E. Gamazon¹, J.C. Denny¹, L. Basterache¹, H. Im², X. Zhong³, N. Cox⁴. 1) Vanderbilt University, Nashville, TN; 2) University of Chicago, Chicago, IL.

In order to conduct a more systematized test of whether the genetically predicted expression of Mendelian genes is associated with features of the Mendelian phenotype, we have combined two relatively novel methods, PrediXcan, a method to predict gene expression based on the cis-regulatory component of variation, and phenotype risk scores (PheRS), a method to quantify clinical features of Mendelian disease from its associated phenome. PheRS is calculated by adding up the number of phenotypes within a set phenome an individual has, after weighting each phenotype by prevalence in a population. Phenome definitions of each Mendelian disease use an established HPO-phenocode map using the HPO terms linked to a Mendelian disease created from OMIM. Previously when using PrediXcan to study associations between predicted gene expression and a phenome suggested to be associated with the gene, decisions regarding enrichment for particular phenotypes depended on qualitative support. A quantitative approach is more reliable and allows for better comparisons across genes. An initial study using just over 10,000 patients from Vanderbilt's biobank, BioVU, focused on cystic fibrosis (*CFTR*), phenylketonuria (*PAH*), and acrodermatitis enteropathica (*SLC39A4*). For each of these Mendelian diseases, a regression was run to test the association of predicted gene expression for the disease-associated gene and the PheRS calculated from the HPO terms that summarize the disease features, while adjusting for sex and the first three principal components. In whole blood, the coefficient on the *CFTR* gene expression term is significant. Neither predicted gene expression for *PAH* or *SLC39A4* was significant with this sample size. This result was expected for phenylketonuria, as previous research showed no differentiation in PheRSs between cases and controls. Moving forward, all Mendelian diseases with HPO-linked terms will be studied in a larger patient sample. In addition, PheRS will be built more directly from ICD codes rather than from OMIM, and then applying those codes to subjects with genotyped data. Ultimately, this will measure for each gene, how much of a pre-specified phenome is associated. This will allow systematic comparisons among genes and flexible extensions, such as being used to identify new associations with genes that appear to be associated with a particular phenome, allowing us to better understand the relationship of Mendelian disease genes to more common complex diseases.

1337F

Identification of rare-disease genes from whole blood RNA-seq of diverse undiagnosed cases. L. Fresard¹, C. Smail², K.S. Smith¹, N.M. Ferraro³, N.A. Teran³, K. Kernohan⁴, X. Li¹, S. Marwaha⁵, Z. Zappala⁶, B. Balliu¹, J.R. Davis³, B. Liu⁶, D. Bonner⁶, J. Kohler⁶, D. Zastrow⁶, D. Fisk⁷, M.E. Grove⁸, J. Davidson⁹, T. Hartley⁴, R. Joshi⁷, B. Strober⁹, S. Utiramerur⁹, A. Battle⁹, G. Bejerano¹⁰, J. Bernstein¹¹, E.A. Ashley^{2,8}, K. Boycott⁴, J.D. Merker^{1,7}, M.T. Wheeler^{3,8}, S.B. Montgomery^{1,3}, Care4Rare Canada Consortium, Undiagnosed Diseases Network. 1) Department of Pathology, School of Medicine, Stanford, CA; 2) Biomedical Informatics Program, Stanford, CA; 3) Department of Genetics, School of Medicine, Stanford, CA; 4) Children's Hospital of Eastern Ontario, Ontario, Canada; 5) Stanford Center for Undiagnosed Diseases, Stanford, CA; 6) Department of Biology, School of Humanities and Sciences, Stanford, CA; 7) Stanford Medicine Clinical Genomics Program, Stanford, CA; 8) Division of Cardiovascular Medicine, School of Medicine, Stanford, CA; 9) Department of Computer Science, Johns Hopkins University, Baltimore, MD; 10) Departments of Computer Science, Stanford, CA; 11) Department of Pediatrics, Stanford School of Medicine, Stanford, CA.

RNA sequencing (RNA-seq) has demonstrated its potential as a complementary tool for Mendelian disease diagnosis. For both rare mitochondrial and muscle disorders, its application has improved genetic diagnoses. However, despite these successes, there still remains a lack of information on the generalizability of this approach to solve diverse undiagnosed Mendelian disorders. We sequenced whole blood RNA from 56 cases with undiagnosed rare diseases spanning 11 diverse disease categories to study the general application of RNA-seq as a diagnostic tool. We developed a robust approach to compare rare disease cases to existing large sets of RNA-seq controls (N=974 external and N=31 family-based controls). We demonstrate how comparison to a large cohort of healthy controls enables identification of gene expression outliers in loss-of-function intolerant genes. We next demonstrate the utility of whole blood- RNA-seq in diagnosing two separate neurological cases; intuitively, blood would not be viewed as a relevant tissue for these disorders. In the first case, two sisters presented with profound global developmental delay, seizures, microcephaly, hypotonia, and progressive scoliosis. We identified *RARS2* as an under-expression outlier, a gene which had previously been associated with neurodegenerative disorders. Whole exome sequencing further revealed compound heterozygous pathogenic variants in the *RARS2* gene in both sisters. In another example, a female presented with global developmental delay, loss of milestones, tremors and seizures. We identified a new splicing junction in *KCTD7*, a gene previously associated with the symptoms presented. A synonymous mutation is responsible for the aberrant splicing event. Across diverse cases in our cohort, we report that blood RNA-seq matches the 10% diagnostic rate obtained from disease matched targeted tissue approaches. This opens the door to a diagnostic tool for clinicians considering more routine application of blood RNA-seq to help unravel undiagnosed rare disease cases.

1338W

Mild Canavan disease presenting with ocular migraines and atypical brain MRI. *P. Tanpaiboon¹, R. Sharma¹, D. Salazar¹, R. Bonilla-Guerrero¹, FL. Lacbawan^{1,2}, SM. Asai³.* 1) Advanced Diagnostics, Biochemical Genetics, Quest Diagnostics, 33608 Ortega Highway San Juan Capistrano, CA. 92675 USA; 2) Advanced Diagnostics, Molecular Genetics, Quest Diagnostics, 33608 Ortega Highway San Juan Capistrano, CA. 92675 USA; 3) Child And Adolescent Neurology, Sutter health, 2825 Capital Ave Sacramento, CA 95816 USA.

Canavan disease (CD) is a rare neurodegenerative disease characterized by spongy degeneration of the white matter. CD is caused by mutations in ASPA gene, which encodes an enzyme that hydrolyzes N-acetyl-L-aspartic acid (NAA): aspartoacylase (ASPA). ASPA deficiency causes accumulation of NAA, which is the biochemical hallmark of CD. In the majority of patients, the first symptom occurs within the first year of life (neonatal/infantile form). The characteristic features are macrocephaly, head lag, and developmental delay. Classic brain MRI findings for CD include diffuse symmetrical white matter degeneration in the subcortical areas, cerebral cortex, and brain stem. Atypical findings, such as basal ganglia hyperintensity, have also been reported. A milder form or juvenile CD is characterized by a mild developmental delay or learning disability with the normal head circumference. Herein, we present findings for an adult patient with a mild CD phenotype: nonspecific neurological problems without abnormal neurological findings. This 22-year old female was initially evaluated by neurology for chronic headaches at 17-years of age. Cognitive function, neurological, and fundus examination were normal. Brain MRI showed symmetrical hyperintensity in bilateral caudate and anterior putamen in T2 and FLAIR images, and asymmetrical bilateral high signal intensity of cortical gray matter. The patient was diagnosed with a migraine with persistent visual aura. A few years later, she developed new episodes of visual field defect and syncope. The repeat brain MRI findings were unchanged from the previous study and mimicking mitochondrial diseases. Consequently, biochemical testing was ordered. Urine organic acid testing revealed marked elevation of NAA consistent with CD. In summary, the patient presented with mild and nonspecific symptoms that first appeared as a teenager and caused a delay in diagnosis. To our knowledge, this is the oldest age of onset for CD reported to date; the previous was ~7 years old. Although atypical, similar brain MRI findings have been reported previously in mild CD phenotypes. Our report is an example of the wide spectrum of phenotypes in genetic conditions and demonstrates the importance of testing for metabolic disorders in differential diagnosis. A definitive diagnosis leads to appropriate treatment, end of the diagnostic odyssey, genetic appropriate counseling, and support for clinical trial eligibility.

1339T

Clinical naming of genetic phenotypes provides unique insights into molecular mechanisms and disease etiology. *A. Hamosh, C. Bocchini, A.L. Stumpf, M.J.F. O'Neill, C. Arnold, J.S. Amberger.* McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD.

Since 1969, OMIM (omim.org) has been at the forefront of the nosology and naming of genetic disease. In the same way that grouping genes of similar function or structure by name opens a window to clinical medicine, grouping similar phenotypes using clinical naming provides a clinical view of the human genome. Clinical naming as an organizing principle of OMIM permits collection and comparison of underlying genes, which opens the possibility of revealing clinically related pathways and disease mechanisms. New disease names are given within the context of existing OMIM phenotypes and are organized and classified first by clinical features and then numbered if genetically heterogeneous (Phenotypic Series). To accommodate necessary evolution in nosology, alternative names are retained under a unique, stable MIM number. The value of clinical naming is seen in the following examples. The MSX1 gene underlies 3 distinct phenotypes - selective tooth agenesis, orofacial cleft, and ectodermal dysplasia - that are also caused by 9, 13, and 15 other genes as represented in their respective Phenotypic Series. Clinical naming highlights the relationship of MSX1 mutations to its phenotypes as well as the relationship of MSX1 to the other causative genes in each Phenotypic Series. Conversely, the Ehlers-Danlos syndrome has been classified into 13 different clinical subtypes. Some of the subtypes have been found to be genetically heterogeneous, for a total of over 20 different EDS syndromes. Arraying the genes responsible for these subtypes reveals etiologic relationships that are not apparent without the clinical naming structure. OMIM has over 415 Phenotypic Series comprised of over 3,400 disorders that can be leveraged accordingly. Naming disorders after genes is ill advised because (1) one third of known disease genes cause more than one disorder; (2) using the same name for a gene and phenotype hinders text mining and machine learning; and (3) phenotypes and genes are distinct concepts whose names should evolve appropriately and independently. Clinical naming of genetic disorders enhances analysis of the relationship between phenotypes and their molecular basis and provides new avenues of research through increased molecular understanding. .

1340F

VariantMatcher: Community-based genomic variant matching to identify new Mendelian disorders. E. Wohler¹, F. Schiettecatte², R. Martin¹, A. Hamosh¹, J. Posey³, C. Akdemir³, Z. Hande³, S. Jhangiani³, J.R. Lupski¹, D. Valle¹, N.L.M. Sobreira¹. 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD; 2) FS Consulting LLP, Seattle, WA; 3) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

The Baylor-Hopkins Center for Mendelian Genomics (BHCMG) seeks to understand the genetic causes of rare Mendelian phenotypes. We have sequenced more than 8394 samples and identified more than 256 novel disease genes, more than 288 known genes, and 190 known genes associated with phenotypic expansion. Sharing phenotypic and genomic information is a pivotal step in elucidating the genes responsible for these phenotypes as we learned from our 5 years' experience with GeneMatcher (genematcher.org). As a step further, to share more of the genomic data generated by the BHCMG, we have created a new web-based tool, VariantMatcher (variantmatcher.org), with the goal of connecting users in the community and participants of the BHCMG project that are interested in the same variant. VariantMatcher is a freely accessible website that can be used by researchers, clinicians or other healthcare providers. To comply with strict patient privacy and safety codes users of the site must register and be approved by site administrators. Users upload the genomic coordinates of interest to the site (up to 10 coordinates per day) and are notified if there is a match or not. The VariantMatcher database contains the rare (MAF <1%), coding, single nucleotide variants (SNVs) identified in the VCF files of the affected and unaffected individuals sequenced by the BHCMG. When there is a match, both parties will be notified by a simultaneous email that contains the email address of both parties so they can connect and exchange further information about the variant, mode of inheritance, phenotype, etc. If a match is not made, the user will be notified but the queried coordinates will not be stored in the database. The coordinates may be uploaded again later if the user chooses. The full dataset is not searchable by users. Data from 251 VCF files are already available for matching in the database. Matching on genomic location should be more specific than matching on features and/or gene name but incomplete penetrance, variable expressivity of the phenotype, age of onset, and zygosity are some of the factors that should be considered when the phenotypes under investigation are being compared. We expect that this tool will facilitate new collaborations among researchers and/or clinicians around the world and improve the next-generation solve rate allowing the discovery of novel disease genes and novel phenotypes.

2115W

Pilot gene expression profiles of surgical weight loss in adipose and blood. D.C. Croteau-Chonka¹, A.M. Weigl², R.P. Chase¹, E.J.M. Rudge³, T. Lo³, S. McNutt¹, L. Farnam¹, A. Tavakkoli³, E.G. Sheu², B.A. Raby¹. 1) Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 2) Center for Clinical Investigation, Brigham and Women's Hospital, Boston, MA; 3) Center for Metabolic and Bariatric Surgery, Department of Surgery, Brigham and Women's Hospital and Harvard Medical School, Boston, MA.

Laparoscopic sleeve gastrectomy (LSG) is a highly effective surgical procedure for inducing weight loss in obese individuals. There is robust epidemiological evidence that substantial weight loss dramatically improves asthma symptoms. Our interest is in the biology behind why and how surgical weight loss positively impacts asthma. We hypothesize that the effects of this intervention are reflected in longitudinal gene expression profiles that represent coherent biological pathways. From among 16 enrolled subjects undergoing LSG, we performed a pilot RNA-sequencing study of 24 samples from the first six subjects (two men and four women, of whom two had asthma). Each subject had two adipose tissue biopsies (subcutaneous (SAT) and omental (OAT)) collected at the time of surgery. Whole blood (WB) was also collected at the time of surgery and at three months afterwards. In these six subjects, body mass index (BMI) significantly decreased from the pre-operative timepoint (44.5 kg/m² (std. dev. = 6.3)) to the post-operative timepoint (35.7 (6.0)) (paired *t*-test $P = 1.8E-05$). Percent predicted forced expiratory volume in one second (FEV₁), a measure of lung function, also significantly increased from pre-op (93.2% (12.7)) to post-op (102.0% (11.4)) ($P = 0.007$). We performed univariate differential gene expression analyses using the DESeq2 R package. Our longitudinal study design allowed us to make two initial exploratory comparisons. We identified 8,823 genes differentially expressed (adjusted $P < 0.05$) cross-sectionally between pre-op SAT and OAT. We identified 208 genes differentially expressed (adj. $P < 0.05$) in pre-/post-op WB across three months of weight loss. We then performed gene set enrichment analyses to identify biological pathways underlying these expression profiles using 50 curated hallmark gene sets from the Molecular Signatures Database. The strongest gene set enrichments in the cross-sectional adipose comparison were for mTOR signaling and adipogenesis (false discovery rate (FDR) < 0.0001). The strongest gene set enrichments in the longitudinal blood comparison were for heme metabolism and downregulation in response to ultraviolet radiation (FDR < 0.001). These preliminary results from three tissues relevant to obese asthma suggest that gene expression profiles implicate pathways with plausible disease connections. Future analyses will include a larger sample size and additional expression and phenotype data at six, nine, and twelve months post-op.

2116T

Person-specific analysis of HLA-presented peptide repertoires reveals novel epitopes associated with type 1 diabetes. J. Arora¹, S. Onengut-Gumuscu², W.M. Chen², A. Lernmark³, S. S. Rich², S. Raychaudhuri¹, T. L. Lenz¹. 1) Research Group for Evolutionary Immunogenomics, Max Planck Institute for Evolutionary Biology, Plön, SH., Germany; 2) Center for Public Health Genomics, University of Virginia, Charlottesville, Virginia, USA; 3) Department of Clinical Sciences, Lund University/CRC, Skåne University Hospital, Malmö, Sweden; 4) Division of Rheumatology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA.

The increasing incidence of Type 1 Diabetes (T1D), along with a poor understanding of its immune-pathology and no available cure to date, represents a major health challenge. Extensive research on T1D consistently indicate that both environmental triggers and genetic predisposition account for disease initiation and progression. A significant proportion of the genetic risk is associated with the Human Leukocyte Antigen (HLA) locus, recently genetically mapped and analyzed, resolving the majority of the T1D-HLA risk to 3 amino acid residues in the peptide-binding groove of HLA-DQ and -DRB1 molecules that together account for 29.6% of familial T1D risk. This finding implicates a major role of DQ- and DR-bound epitopes for disease risk. Although a number of candidate proteins have been proposed to contain potential self-antigens, the exact T1D-relevant epitopes remain elusive. We used the established Peptidome-wide Association Study (PepWAS) approach on 6,651 T1D cases and 9,378 controls in order to screen for T1D-relevant HLA-bound peptides from the large pool of all possible 15-mer peptides (preferred epitope size for DQ/DR molecules) derived from 17 T1D candidate self-proteins (N = 8,018 peptides). This approach identifies disease-associated peptides by integrating HLA genotype, computationally predicted individual self-peptide repertoires, and disease status. Using PepWAS, we identified 347 and 222 potential T1D-relevant epitopes bound by HLA-DQ and DR alleles, respectively. We also simulated deamidation of the candidate proteins and predicted 136 and 113 T1D-relevant deamidated epitopes bound by DQ and trans-encoded DQ alleles, respectively. Inter-individual variation in T1D case-specific epitope repertoires bound by their HLA-DQ and DR variants additively accounted for 34.2% disease risk, significantly exceeding the T1D risk associated with genetic variants. Notably, identified epitopes varied substantially in the strength of their disease association, allowing for prioritization in experimental assays. Additionally, molecular mimicry between predicted T1D-relevant self-epitopes and specific pathogen-derived peptides provides a potential mechanism for a potential environmental trigger of T1D. Results from experimental assays to validate the predicted T1D-relevant epitopes will also be shown. Overall, our PepWAS-based approach promises a novel strategy to identify and prioritize disease-relevant epitopes for targeted therapy in HLA-associated complex diseases.

2117F

Association analysis of type 1 diabetes in >56,000 ancestrally diverse subjects identifies novel risk loci. C.C. Robertson¹, J.R.J. Inshaw², S. Onengut-Gumuscu¹, W.M. Chen¹, J.A. Todd², S.S. Rich¹, *Type 1 Diabetes Genetics Consortium*. 1) Center for Public Health Genomics, University of Virginia, Charlottesville, VA; 2) JDRF/Wellcome Diabetes and Inflammation Laboratory, Wellcome Centre for Human Genetics, University of Oxford, UK.

Background: Type 1 diabetes (T1D) is a complex disease driven by diverse genetic and environmental factors. Approximately 60 genomic regions are implicated in T1D, yet ~20% of familial clustering in Europeans remains unexplained by known T1D alleles and little is known about the contribution of risk alleles in non-European populations. To address these gaps, we analyzed genetic association with T1D in an ancestrally diverse cohort twice the size of previous studies. **Methods:** DNA samples were provided to the Type 1 Diabetes Genetics Consortium (T1DGC) and genotyped on the custom Illumina ImmunoChip, which provides dense SNP coverage in 186 regions associated with immune-mediated disease. After stringent quality control, genotypes were imputed using the 1000 Genomes Project Phase 3 reference panel applying SHAPEIT for phasing and Minimac3 for imputation. Imputed variants were filtered for imputation quality ($r^2 > 0.5$). Association analyses with T1D status were performed in seven matched case-control series and one affected-family collection. Case-control association was analyzed using logistic regression, adjusting for population stratification using the first five genetic principal components. Family-based analyses used transmission disequilibrium tests. Fixed effect meta-analysis was used to combine summary statistics from case-control and family-based analyses. **Results:** After sample and variant QC, there were 130,335 genotyped ImmunoChip variants in 56,693 participants, including 15,258 cases, 19,071 controls, and 6,475 affected parent-offspring trios, from diverse ancestry groups (3,194 African, 1,733 Asian Pacific, and 51,766 European or other admixture). After imputation and filtering for high-confidence imputed variants, 1,665,998 variants were analyzed for association with T1D. Forty loci previously associated with T1D are represented on the ImmunoChip. In meta-analysis, we replicated 36/40 of these associations at $p < 5 \times 10^{-8}$ and all 40 at $p < 5 \times 10^{-6}$. We also identified 24 novel T1D-associated regions at genome-wide significance ($p < 5 \times 10^{-8}$), with the strongest associations at 19q13.31 (OR=1.22), 8q12.1 (OR=1.19), 11q12.2 (OR=0.81), and 3q28 (OR=0.82). **Conclusion:** Leveraging our improved power for discovery, we identified novel T1D-associated chromosome regions. Characterization of these additional T1D loci will permit detection of candidate causal genes, elucidate mechanisms underlying disease biology, and improve genetic risk prediction.

2118W

Mitochondrial DNA copy number is a heritable trait that is significantly associated with multiple metabolic phenotypes. L. Ganel^{1,2}, L. Chen^{1,2}, H.J. Abel^{1,3}, D.E. Larson^{1,3}, A. Regier², I. Das¹, S.K. Service⁴, X. Yin⁵, A.U. Jackson⁵, M.I. Kurki⁶, A.S. Havulinna^{7,8}, P. Paltas⁹, S.K. Dutcher^{1,3}, V. Salomaa⁷, C.W.K. Chiang⁹, A. Palotie^{6,8,10,11}, S. Ripatti^{8,12}, N.B. Freimer⁴, M. Boehnke⁵, A.E. Locke^{1,2}, M. Laakso^{1,3}, N.O. Stitzel^{1,2,3}, I.M. Hall^{1,2,3}, *FinMetSeq Consortium*. 1) McDonnell Genome Institute, Washington University School of Medicine, St. Louis, MO, USA; 2) Department of Medicine, Washington University School of Medicine, St. Louis, MO, USA; 3) Department of Genetics, Washington University School of Medicine, St. Louis, MO, USA; 4) Department of Psychiatry, The Jane and Terry Semel Institute for Neuroscience and Human Behavior, David Geffen School of Medicine, University of California, Los Angeles, CA, USA; 5) Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA; 6) Broad Institute of MIT and Harvard, Cambridge, MA, USA; 7) National Institute for Health and Welfare, Helsinki, Finland; 8) Institute for Molecular Medicine Finland (FIMM), HiLIFE, University of Helsinki, Helsinki, Finland; 9) Department of Ecology and Evolutionary Biology, University of Southern California, Los Angeles, California, USA; 10) Massachusetts General Hospital, Boston, MA, USA; 11) Harvard Medical School, Boston, MA, USA; 12) Public Health, Faculty of Medicine, University of Helsinki, Finland; 13) Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland.

Mitochondrial DNA copy number (mtDNA-CN) varies among individuals and tissues and has tentatively been associated with a range of metabolic traits. However, prior reports of these associations have been inconsistent and often underpowered. To study the association between mtDNA-CN and metabolism, we leveraged deep ($>20\times$) whole genome sequencing (WGS) data in 4,163 Finnish individuals with extensive phenotypic information for >100 metabolic traits. To increase power, we obtained imputed GWAS array genotype data from 17,718 Finnish individuals (including 3,923 that overlap with WGS samples), for which we also had whole exome sequencing (WES) data. We estimated mtDNA-CN from the WGS and WES data by normalizing the mitochondrial genomic coverage to that of the nuclear genomic coverage and controlling for sex, age, and age². Across the 4,163 WGS samples, mtDNA-CN exhibits substantial inter-individual variation and follows an approximately normal distribution (mean=235.3; std=66.6). mtDNA-CN heritability was 31%, similar to previous estimates. In the WGS data, we observed significant associations between mtDNA-CN and several metabolic phenotypes, including fasting insulin ($p=2.0 \times 10^{-21}$), fat mass ($p=4.5 \times 10^{-16}$), and body mass index ($p=5.7 \times 10^{-12}$). In a subset of 2,453 non-diabetic individuals, mtDNA-CN was significantly associated with insulin sensitivity ($p=7.3 \times 10^{-17}$) and disposition index ($p=0.023$) after adjusting for fat mass. We identified one genome-wide significant association ($p=1.26 \times 10^{-10}$) in the *MYB-HBS1L* locus, a region previously associated with blood monocyte and platelet counts. Conditional analysis reveals that this known association is located on the same haplotype as our mtDNA-CN association. We also performed rare variant association analyses using the WES data and found a statistically significant association between mtDNA-CN and rare, predicted high-impact variants in *TMBIM1*, a gene whose expression has recently been reported to be protective against non-alcoholic fatty liver disease. Our results suggest that mtDNA-CN is significantly associated with multiple metabolic phenotypes related to obesity and glucose metabolism. Furthermore, we have confirmed mtDNA-CN as a heritable trait and identified one quantitative trait locus and one gene with a significant association of rare variants.

2119T

Association study between gene expression and body mass index for middle aged Danish twins using weighted correlation network analysis.

W. Li¹, J. Baumbach^{2,3}, L. Christiansen¹, Q. Tan^{1,4}. 1) Department of Public Health, University of Southern Denmark, Odense, Denmark; 2) Department of Mathematics and Computer Science, University of Southern Denmark, Denmark; 3) Chair of Experimental Bioinformatics, TUM School of Life Sciences, Technical University of Munich, Munich, Germany; 4) Department of Clinical Research, University of Southern Denmark, Denmark.

Body mass index (BMI) is generally recognized as an indication for underweight and obesity. As a multifactorial trait, BMI is influenced by both genetic and environment factors. Multiple genome-wide association studies have been performed to identify the genetic variants with limited success. Understanding the molecular basis that mediates environmental exposure and BMI can be more important for prevention of metabolic disorders. In this study, we focus on gene expression changes associated with BMI using monozygotic twins. Monozygotic twins that discordant for BMI are perfect subjects for transcriptomics studies where the genetic variants, age and sex are controlled. Twins are selected from The Middle Aged Danish Twins Study with BMI measured and their gene expression profiles obtained using gene expression microarray. Gene expression was summarized to gene level and weighted gene-expression correlation network analysis (WGCNA) was used to cluster correlated genes into modules and test their association with BMI. We applied pair-wise analysis in testing both module-trait and gene-trait associations. There were 4 out of 24 modules found significantly associated with BMI with $p < 0.036$, and gene ontology enrichment analysis of 919 genes in one modules suggested that these genes are highly related to several metabolic and cellular processes (cellular metabolic process, FDR $9.3e-7$; organic substance metabolic process, FDR $1.0e-5$; primary metabolic process, FDR $3.7e-5$). Our results suggested differential gene expression regulation involved in multiple metabolic and biological processes that lead to differential BMI. Moreover, we show that the discordant twin design in combination with WGCNA, is an efficient and powerful approach in transcriptomic studies.

2120F

A single-cell transcriptomics roadmap to investigate diabetes remission induced by the central action of fibroblast growth factor 1 (FGF1). T.H.

Pers², M.A. Bentsen¹, D. Rausch², J. Scarlett^{1,3}, K. Alonge¹, Z. Mirzadeh¹, A. Secher², R. Jorgensen², M.W. Schwartz¹. 1) UW Medicine Diabetes Institute, University of Washington, WA, USA; 2) Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, DK; 3) Department of Pediatric Gastroenterology and Hepatology, Seattle Children's Hospital, Seattle, WA, USA; 4) Barrow Neurological Institute, Phoenix, AZ, USA; 5) Global Research, Novo Nordisk A/S, Måløv, DK.

Recently, lasting remission of hyperglycemia was achieved in rodent models of type 2 diabetes (T2D) by a single intracerebroventricular (icv) injection of FGF1. While the mechanism underlying this effect is unknown, the lack of association with changes of body fat mass or risk of hypoglycaemia raises the possibility that icv FGF1 normalizes the defended level of glycemia, rather than simply lowering blood glucose levels. Recent work from our lab has identified gluco-regulatory neurocircuits in the mediobasal hypothalamus (MBH) as targets for this FGF1 effect. To investigate the mechanism underlying FGF1 action in this brain area, we used large-scale single cell RNA-sequencing to identify and characterize FGF1-responsive cells in mouse MBH. Based on >70,000 single cell MBH transcriptomes from diabetic ob/ob mice harvested 5d after a single icv injection of either FGF1 or vehicle, we identified >20 cell clusters and >900 cell type-specific differentially expressed genes ($p < 0.01$). Among these potential targets of the MBH response to FGF1 in diabetic mice, we found that FGF1 treatment robustly increased the proportion of newly formed oligodendrocytes, a response that can be predicted to yield new mature oligodendrocytes and associated myelination of axons. These preliminary results demonstrate that transcriptional identification and characterisation of icv FGF1-responsive cell clusters may both advance our understanding of brain control of glucose homeostasis and identify novel targets for diabetes prevention and treatment.

2121W

Screening the telomeric regions and a rare variant in the development of DN in the North Indian population. A.J.S. Bhanwer¹, G. Singh^{1,2}, P. Raina¹, H.S. Sandhu¹, V. Kilotra¹, V. Sharma², I. Sharma², I. Sethi², V. Singh³, A. Marwaha⁴, E. Rai², S. Sharma². 1) Human Genetics, Guru Nanak Dev University, AMRITSAR, Punjab, India; 2) Human Genetics Research Group, School of Biotechnology, Shri Mata Vaishno Devi University, Katra, J&K, India; 3) Kidney Clinic and Dialysis Centre, Opposite GMC Amritsar; 4) Joshi Hospital and Trauma Centre, Jalandhar.

Diabetic nephropathy (DN) is the leading cause of end stage renal disease (ESRD) worldwide. In population based studies and registry analyses, DN has emerged as the most common cause of ESRD in India. Both genetic and environmental factors likely to contribute to its development and progression, but the precise mechanisms underlying are still unknown. Different epidemiological studies suggest that genetic susceptibility plays an important role in the pathogenesis of DN, leading to extensive efforts to identify the novel gene(s) involved or replicate the known genes in different populations to uncover the role of genetics in the development and progression of this condition. Unfortunately, telomeric regions and rare variants were overlooked in most of the previous studies either on T2D or its secondary complications, so in this study two SNPs from telomeric regions and one from rare variants were included. In the present study, three SNPs from three susceptible genes of T2D viz. rs7675998 (*NAF1*), rs74019828 (*CSNK2A2*) and rs61741902 (*SGSM2*) were selected to reveal their contribution in the development of the secondary complication of T2D. A total of 1740 samples from North India were included [374 DN cases and 1366 Non-DN controls (1056 Non-DN & Non-T2D controls and 310 T2D without DN controls)] and were genotyped using Sequenom MassArray platform. All these genes were found to be significantly associated with DN, but first two of these were providing protection towards DN, when compared with three control groups i.e. NDN, NDNT and T2D without DN, except rs7675998 which conferred risk when compared with T2D without DN control group. This study concludes that neither telomeric regions nor rare variants can be disregarded, instead should be replicated in other populations with more polymorphisms from these genes to corroborate their role in DN as well as T2D development.

2122T

Whole-genome sequence analysis of body mass index in the trans-omics for precision medicine (TOPMed) program. J.A. Brody¹, J.R. O'Connell², J.A. Perry² on behalf of the TOPMed Anthropometry Working Group. 1) University of Washington, Seattle, WA; 2) University of Maryland, Baltimore MA.

While genetic analyses of common variants for body mass index (BMI) have leveraged huge sample sizes, relatively few studies have examined rare non-coding variation in relation to BMI. Through whole-genome sequencing (WGS) in TOPMed we have a unique opportunity to examine BMI genetics in a large multi-ethnic sample of deeply sequenced individuals. Analyses were performed in a sample of 45,159 individuals from 20 cohorts and 6 ancestry groups. Age-, study- and PC-adjusted BMI residuals were created within ancestry and sex strata, then rank-normal transformed and rescaled by strata standard deviation. Pooled residuals were analyzed with linear mixed models including a variance component for empirical kinship plus separate residual variance components for each sex-ancestry group. Single-variant analyses were performed using MMAP on all variants (N=87,683,863) with a minor allele count of ≥ 5 . Analyses identified single variants associations at six loci that surpassed a suggestive significance threshold of $p < 1e-8$: in addition to known loci at *FTO*, *MC4R* and *SEC16B*, three novel rare variant associations were identified (*ARHGAP24*, rs546474633, MAF 0.11%, B=-4.3, p=8.4E-9; *ISPD*, rs757499854, MAF=0.02%, B=10.8 kg/m², p=3.7E-9 and *FRMD5*, rs897162855, MAF=0.01%, B=16.6 kg/m², p=9.6E-9). Complete WGS coverage also allows a better understanding of the genetic architecture of previously identified loci. For example, we identified an independent rare variant signal (rs62106258, MAF=3.2%, B=-0.77 kg/m², P=3.5E-8) in the known intergenic region upstream of *TMEM18* (rs2867125, B= 0.33 kg/m², MAF=15, P=1.4E-5). Rare variants (minor allele frequency < 1%) were aggregated into gene-based units utilizing strategies that infer the target gene for enhancer regions along with variants that fell within the FANTOM5 promoter regions and nonsynonymous coding variants. Enhancer regions and their inferred target gene were identified using GeneHancer and EnhancerAtlas neuronal data sets. Gene-based tests were analyzed using SKAT in the GENESIS R package. No genes surpassed a Bonferroni correction for multiple testing. These findings require further investigation and we are seeking performing replication using the next set of samples (50K) soon to be released by TOPMed. .

2123F

Large trans-ethnic discovery identifies novel loci, and differing genomic and expression signatures between glycemic traits. J. Chen, *Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC) Investigators*. Wellcome Sanger Institute, Hinxton, Cambridge, United Kingdom.

Glycemic traits are used to diagnose and monitor diabetes and can inform pathophysiology in type 2 diabetes (T2D). Within MAGIC we analyzed large-scale trans-ethnic genetic data (imputed to 1000Genomes) in up to 281,416 individuals without diabetes across five ancestries (71% European, 13% East Asian, 7% Hispanic, 6% African-American and 3% South Asian) relating to fasting glucose (FG), fasting insulin (FI), 2-hour glucose (2hGlu) and glycated hemoglobin (HbA1c). Within-ancestry meta-analyses were conducted in METAL (fixed-effects), followed by trans-ethnic meta-analyses in MANTRA. We identified 242 genome-wide significant loci, of which 99 have not been previously associated with any glycemic trait or T2D. We detected ancestry-driven associations, e.g. two African-American signals for FI (rs12056334) and FG (rs61909476), and a Hispanic signal for FG (rs12315677). In addition, 22 trans-ethnic loci have a $p > 1 \times 10^{-5}$ in the European meta-analysis, and their discovery was enabled by the different ancestries and not just overall sample size. Combining genetic results with additional genomic resources has provided insights into genomic and expression signatures of each glycemic trait. FG-associated loci were enriched for genes expressed in pancreas/islets and liver, FI-associated loci with adipose and skeletal muscle, and HbA1c-associated loci with all these plus blood/immune system. These expression signatures were corroborated by corresponding enrichment of cell-type specific stretch enhancers. Furthermore, HbA1c expression and genomic signatures were consistent with the likely mode of action of HbA1c-associated loci, based on lookups of association results in glycemic, red blood cell and iron traits and using fuzzy clustering for locus classification. Overall, 36.6% of loci principally affect HbA1c via erythrocytic effects, ~22.8% via glycemic effects, and 29.1% were pleiotropic. T2D risk was 1.05 times higher in loci affecting HbA1c through glycemic mechanisms than others ($p = 2.2 \times 10^{-6}$). In DEPICT pathway analyses, FG-associated loci were enriched for insulin sensitivity, glucose tolerance and pancreas development pathways; FI-associated loci were enriched for pathways involved in glucose homeostasis, embryogenesis, growth and cancer; and HbA1c-associated loci were enriched for hematopoiesis and glucose metabolism pathways. Overall, this effort identified 99 novel loci and provides distinct expression, genomic and pathways signatures for each glycemic trait.

2124W

Uncovering the genetic architecture of data-driven dietary habits in UK Biobank. J.B. Cole^{1,2,3}, J.C. Florez^{1,2,4}, J.N. Hirschhorn^{1,3,4}. 1) Broad Institute, Cambridge, MA; 2) Massachusetts General Hospital, Boston, MA; 3) Boston Children's Hospital, Boston, MA; 4) Harvard Medical School, Boston, MA.

Unhealthy dietary habits are among the leading risk factors for several life-altering metabolic diseases such as obesity, type 2 diabetes, and coronary artery disease, which substantially increase mortality and decrease quality of life worldwide. There has been limited success in the use of traditional summary measures of diet (i.e. macronutrient composition) in genetic and epidemiological studies, emphasizing the potential for data-driven approaches to derive biologically meaningful and genetically based dietary habits for nutrigenomic research that might be easily harmonized across cohorts. We derived novel and data-driven dietary traits from two complementary and comprehensive diet questionnaires in UK Biobank, the largest cohort of complex dietary variables and genetics collected to date (N~500K). We employed statistical approaches that leverage repeated measures and correlated traits, including Empirical Bayes and principal components analysis. Preliminary heritability results of our derived dietary habits surprisingly found that 89.3% of dietary habits were significantly heritable ($P < 0.05/309$ traits analyzed to date = 1.6×10^{-4}), with milk type (soy milk vs. never drinks milk = 28.3%; full cream vs. never drinks milk = 22.8%), alcohol intake (with meals = 16.5%; overall frequency = 12.1%), and water intake (glasses per day = 11.7%) being the most heritable. Additional dietary habits with moderate heritabilities include pieces of fresh fruit per day (10.8%), coffee type used (10.6%), and amount of salt added to food (10.2%). Preliminary GWAS of these eight heritable traits have identified 394 genome-wide significant loci with minor allele frequency $> 0.05\%$ ($P < 5 \times 10^{-8}$), with 137 of them meeting study-wide significance ($5 \times 10^{-8}/309$ traits = 1.6×10^{-10}). The strongest association is the well-known *ADH1B* variant, rs1229984, associated with overall alcohol frequency ($P = 5.9 \times 10^{-179}$), but we also identified other novel and highly significant associations including rs4410790 with glasses of water per day ($P = 3.3 \times 10^{-79}$), rs10249294 with fresh fruit intake ($P = 4.4 \times 10^{-65}$), rs11336847 with overall alcohol frequency ($P = 3.0 \times 10^{-63}$), and rs2736748 with amount of salt added to food ($P = 1.4 \times 10^{-27}$). In summary, these preliminary findings not only highlight the importance of using wider dietary information but also represent novel dietary habit genetic discoveries, laying the groundwork for future investigation into the biology of diet and its impact on human health and medicine.

2125T

Socioeconomic status associates with gene expression and DNA methylation profiles within genes of the hypothalamic-pituitary-adrenal axis in Mexican Americans. V.P. Diego¹, J.M. Peralta¹, M. Almeida¹, T.E. Howard¹, E. Manusov¹, M. Johnson¹, J.E. Curran¹, S. Williams-Blangero¹, M.A. Carless², J. Blangero¹. 1) Department of Human Genetics, School of Medicine, University of Texas Rio Grande Valley, Brownsville, TX; 2) Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX.

Low socioeconomic status (SES) is a potent determinant of morbidity and mortality worldwide, particularly in Mexican Americans living close to the US-Mexico border. While this epidemiological relationship is well established, its underlying physiological mechanisms are poorly understood. It is thought that the hypothalamic-pituitary axis (HPA) axis, which is the main physiological stress response system, links SES stress in a causal manner with pathophysiology leading to increased morbidity and mortality. To investigate the effect of SES variables on genetic regulation within the HPA axis, we examined gene expression data for 206 genes in the HPA axis and DNA methylation sites for 249 genes in the HPA axis (including those for which gene expression data was available) in peripheral blood samples obtained from up to 710 individuals. We examined three dichotomized (high versus low) SES variables, namely Duncan's socioeconomic index (SEI), education years (EDU), and household income (INC). These variables are moderately correlated, with tetrachoric correlation coefficients of 0.53, 0.42, and 0.43 for SEI-EDU, SEI-INC, and EDU-INC, respectively. To test if gene expression and/or DNA methylation differ substantially between low and high SES states, we used a vector subspace comparison approach, which is a method of systematically comparing the dominant eigenvectors explaining most of the variation in gene expression and DNA methylation. Under this approach, the null hypothesis of interest is that the subspace distance is 0. We tested this hypothesis by a permutation approach to generate an empirical distribution against which the observed subspace distance was compared in order to compute an associated p-value for formal inference. For the gene expression SES comparisons, the SEI, EDU, and INC contrasts all yielded significantly different subspaces ($p < 0.0001$). Similarly, for the epigenetic SES comparisons, all contrasts yielded significantly different subspaces ($p < 0.0001$). As these inferences are based on comparisons of the dominant eigenvectors, our results show that such high versus low SES contrasts are associated with significantly different genetic/epigenetic effects. We are currently investigating the unique and shared genes associated with SES across the 249 gene expression and DNA methylation data.

2126F

Using genetics to test the birth weight effects of higher maternal body fat uncoupled from its adverse metabolic consequences. R.M. Freathy^{1,2}, W.D. Thompson¹, N.M. Warrington³, R.N. Beaumont¹, Y. Ji¹, M.C. Borges³, J. Tyrrell¹, A.R. Wood¹, D.M. Evans³, T.M. Frayling¹, H. Yaghootkar¹, D.A. Lawlor². 1) Institute of Biomedical and Clinical Science, University of Exeter, Exeter, United Kingdom; 2) MRC Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom; 3) University of Queensland Diamantina Institute, University of Queensland, Brisbane, Australia.

Maternal obesity is associated with high offspring birth weight (BW) and consequent obstetric problems and later-life disease. An improved understanding of this association would offer the potential to reduce the risk. We previously showed, using genetic instruments, that higher maternal BMI is causally associated with higher offspring BW, but the mechanisms underlying this effect are poorly understood. Recently genetic variants associated with higher adiposity but lower cardio-metabolic risk ("favourable adiposity" variants) were identified. We aimed to estimate the causal effect of maternal adiposity on offspring BW using two genetic instruments: one that captures adiposity using BMI-related variants, and one that couples adiposity with favourable metabolic effects. We performed 2-sample Mendelian randomization using summary data on genotype-BMI (76 SNPs) and genotype-favourable adiposity (14 SNPs) associations, and maternal SNP-offspring BW associations from 190,406 UK Biobank women. We calculated effects using Wald ratios and used MR-Egger to adjust for pleiotropic effects. Since maternal and offspring genotypes are correlated ($r=0.5$), we used a structural equation model (SEM) to estimate the independent maternal and offspring genetic effects using data on own BW ($n=215,444$) and offspring BW (n as above) from the UK Biobank. A 1 SD (4 kg/m²) genetically higher maternal BMI was associated with a 54g (95%CI: 40, 68) higher offspring BW, but there was evidence that pleiotropic effects contributed to this (MR Egger estimate 14g (-22, 49)). Without controlling for offspring genotype, a 1 SD (8.5%) higher maternal body fat percentage (captured by favourable adiposity variants) appeared to cause a 25g lower offspring BW (-25g (-82, 32)), with similar results using MR Egger. However, adjusting for the maternal-fetal genotype correlation dramatically changed the results to a 150g lower BW (-232, -69). In contrast, a 1 SD higher offspring body fat percentage (captured by favourable adiposity variants) was associated with a 229g (151, 307) higher BW. Using genetic variants that couple maternal adiposity with favourable metabolic consequences, we found a BW-lowering effect, suggesting that fetal growth is not influenced by the amount of maternal body fat per se, but by metabolic correlates. The contrasting effects of maternal and fetal favourable adiposity on BW point to a fetal insulin-related mechanism, which will require investigation in further studies.

2127W

Using genetics to understand the ethnic differences in adiposity and diabetes risk. J. Harrison, Y. Ji, S.E. Jones, R.N. Beaumont, A.R. Wood, J. Tyrrell, M.N. Weedon, T. Frayling, H. Yaghootar. University of Exeter Medical School, Exeter, United Kingdom.

Studies of non-Europeans are particularly important because they develop obesity related diseases at a lower body mass index (BMI). This additional risk is strongest for people of South Asian ancestry. To investigate the genetic aetiology behind this, we tested the hypothesis that variants associated with higher adiposity but lower risk of disease – “favourable adiposity” – would have different frequencies and different effects in South Asians. We used 10,203 South Asians and 451,000 Europeans from UK Biobank. We generated a genetic score based on 14 favourable adiposity variants and assessed the collective effect of these variants on measures of adiposity including body fat percentage and BMI, and risk of type 2 diabetes (1,150 cases in South Asians and 14,371 in Europeans) and heart disease (1,387 cases in South Asians and 37,741 in Europeans). Favourable adiposity alleles were less common in the South Asian population than in the European population (average number of alleles 10.5 vs 11.6; $P < 0.0001$) despite having similar effects on adiposity and risk of type 2 diabetes. In the South Asian population, carrying an additional favourable adiposity allele was associated with higher body fat percentage (0.148%, 95% confidence interval (CI) [0.0986, 0.197]) and higher BMI (0.062 Kg/m² [0.027, 0.10]) but lower risk of type 2 diabetes (0.95 Odds Ratio (OR) [0.92, 0.97]) and heart disease (0.97 OR [0.95, 0.99]). In the European population, an additional favourable adiposity allele was associated with higher body fat percentage (0.128% [0.119, 0.136]) and higher BMI (0.038 Kg/m² [0.034, 0.043]), but lower risk of type 2 diabetes ((0.97 OR [0.94, 0.97]) and heart disease (0.98 OR [0.98, 0.99]). Our study demonstrates that favourable adiposity alleles have similar effects in South Asians and Europeans, lowering the risk of type 2 diabetes and heart disease for a given BMI, but these alleles are less frequent in South Asians. The “favourable adiposity” alleles are associated with higher subcutaneous but lower liver fat in Europeans. The lower frequency of these favourable adiposity alleles in South Asian populations suggests a potential mechanism for some of the ethnic differences in adiposity and diabetes risk.

2128T

Obesity revisited: Is there a genetic basis to obesity with favorable cardiometabolic profile? L.O. Huang¹, U. Schick², N. Yang², T.O. Kilpeläinen^{1,3}, R.J.F. Loos^{2,4,5}. 1) Section for Metabolic Genetics, Novo Nordisk Foundation Center for Basic Metabolic Research, København N, København, Denmark; 2) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, New York, United States of America; 3) Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, New York, United States of America; 4) The Genetics of Obesity and Related Metabolic Traits Program, The Icahn School of Medicine at Mount Sinai, New York, New York, United States of America; 5) The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, New York United States of America.

Worldwide obesity has nearly tripled since 1975 and is described as a global pandemic. Obesity is often associated with cardiometabolic comorbidities such as insulin resistance, hypertension, dyslipidemia, and increased risk of cardiovascular disease and type 2 diabetes (T2D). However, a substantial fraction of obese individuals retain a cardiometabolically healthy profile, also called the metabolically healthy obese (MHO). While the biological mechanisms that underlie MHO are poorly understood, a number of recent single-trait GWAS have shown that there are genetic variants that predispose to increased adiposity, yet are associated with a favorable cardiometabolic profile. To identify novel adiposity-increasing variants with protective cardiometabolic effects, we performed pairwise analyses -- using the CPASSOC method -- of adiposity and cardiometabolic traits using available summary statistics from 11 GWAS, including three adiposity traits [body fat percentage (BFP), body mass index (BMI), and waist-to-hip ratio (WHR)] and eight cardiometabolic traits [high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), triglycerides, fasting glucose and insulin, blood pressure, and incident coronary artery disease (CAD) and T2D]. We identified genome-wide significant associations ($P < 5 \times 10^{-8}$) with any of the 24 trait pairs consisting of one of three adiposity traits and one of the eight cardiometabolic traits. This resulted in the discovery of 62 independent loci residing in gene regions such as *COBLL1*, *PPARG*, *VEGFB* and *VEGFA*. Results of tissue and cell enrichment analyses using DEPICT implicate strongest enrichment in adipose tissue, adrenal cortex, liver, and arteries. Gene set enrichment analyses using DEPICT highlighted several pathways ($P < 5 \times 10^{-4}$), such as “abnormal muscle cell glucose uptake”, “decreased percent body fat”, and “response to insulin stimulus”. Our results provide insights into biological pathways that may underlie MHO and could shed light on future drug development targeting improvement in cardiometabolic profile of unhealthy obese individuals.

2129F

Exome-wide association study reveals missense variants associated with diabetic retinopathy in Koreans. B. Kim¹, S. Moon¹, J. Ahn², J. Kong¹, Y.G Kim¹, M.Y Hwang¹, S. Han¹, K.H Park³. 1) Korea National Institute of Health, Cheongju-si, Chungcheongbuk-do, Korea, Republic of; 2) Department of Ophthalmology, Seoul National University, College of Medicine, SMG-SNU Boramae Medical Center; 3) Department of Ophthalmology, Seoul National Bundang Hospital.

Diabetic retinopathy (DR) is the most common diabetic complication and a leading cause of blindness in Korean. According to the Korea National Health and Nutrition Examination Surveys (KNHANES) of the Korean Center for Disease Control and Prevention, the weighted prevalence of DR was 0.9% (95% confidence interval (CI), 0.7-1.1) in the whole adult population and 11.0% (95% CI, 8.9-13.6) in Korean adults with diabetes. With the prevalence of diabetes expected to be increased, the prevalence of DR in Korea will have the highest prevalence of diabetes among OECD countries by the year 2030. DR is associated with the combination of genetic and environmental factors. In particular, progression of DR is closely related to the longer duration of Type 2 Diabetes (T2D) and poor glycemic control. With technological advances of genome-wide association study (GWAS), many risk factors associated with T2D have been identified. However, the aetiology of this complication still remains unclear. Here, we aimed to find functional relevance between variation and phenotype. To achieve this, we conducted exome-wide association analysis using Illumina Infinium HumanExome BeadChip comprising >250,000 markers. After sample quality control process, a total of 538 exome data were used (460 DR and 78 T2D patients). In addition, 97,300 markers were used for logistic regression analysis adjusting age, gender, and body mass index. From the regression analysis between case-control samples, we identified 11 genetic variants including 8 missense variants with statistical cut-off of $P < 10^{-4}$. In particular, among genes encompassing identified variants, SOAT1 at chromosome 1 ($P = 3.9 \times 10^{-5}$) is known that this gene has been implicated in the formation of beta-amyloid and atherosclerotic plaques by controlling the equilibrium between free cholesterol and cytoplasmic cholesteryl esters. Moreover, RHBDF1 at chromosome 16 ($P = 7.95 \times 10^{-5}$) were reported that this gene regulates the secretion of several ligands of the epidermal growth factor receptor. The functions of these genes are deeply connected with pathogenic mechanism of DR. In conclusion, our findings suggest new candidate genes contributing to DR.

2130W

Global, multi-ethnic genome-wide association meta-analysis of body mass index. A.E. Locke¹, S. Vedantam^{2,3}, E. Marouli⁴, S. Berndt⁵, L. Yengo⁶, A.R. Wood⁷, T. Ferreira⁸, S. Graham⁹ on behalf of the Genetic Investigation of ANthropometric Traits (GIANT) Consortium. 1) Department of Medicine, Division of Genomics and Bioinformatics, Washington University School of Medicine, Saint Louis, MO; 2) Center for Basic and Translational Obesity Research, Division of Endocrinology, Boston Children's Hospital, Boston, MA; 3) Broad Institute of MIT and Harvard, Cambridge, MA; 4) Heart Centre, William Harvey Research Institute, Barts & the London Medical School, Queen Mary University of London, Charterhouse Square, London, UK; 5) National Cancer Institute, Bethesda, MD; 6) Program in Complex Trait Genomics, Institute for Molecular Bioscience, University of Queensland, St Lucia, Brisbane, Australia; 7) Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Exeter, UK; 8) Big Data Institute, Li Ka Shing Center for Health and Health Information and Discovery, Oxford University, UK; 9) Department of Internal Medicine, Division of Cardiovascular Medicine, University of Michigan, Ann Arbor, MI.

Obesity is a health crisis affecting more than 1/3 of the US population and contributing to increased risk for other serious health conditions such as heart disease, diabetes, and numerous cancers. Using body mass index (BMI) as a proxy for overall obesity (heritability estimated at ~40-60%), GIANT and other groups have committed substantial effort to understand the genetic basis and underlying biology of obesity. The nearly 200 loci that have been associated with BMI show strong signal enrichment in brain tissues and the central nervous system. However, the BMI-associated loci discovered to date account for <3% of phenotypic variance. Here, we describe our continuing efforts to identify genetic factors contributing to overall obesity through GWAS meta-analyses of BMI. We combined published results from previous GIANT GWAS meta-analyses of BMI ($n \sim 250,000$) with a BOLT-LMM association analysis of a recent UK Biobank release ($n = 456,426$) for a combined total of 681,275 European-descent participants. We limited fixed effects meta-analysis to the ~2.4M HapMap2 variants common to both studies. Through meta-analysis and approximate conditional and joint analysis using GCTA-COJO we identified 941 loci at a stringent threshold of $P < 1 \times 10^{-8}$, including 656 primary associations and 285 secondary signals. These associations account for 6.0% of the BMI phenotypic variance. In tandem, we initiated a global GWAS meta-analysis of BMI based on 1000 Genomes and HRC reference panels covering >40M imputed variants. To date, we have received association summary statistics from >170 studies across six ethnic groups totaling >820,000 individuals, nearly half of whom are from non-European ancestry studies. The upcoming final meta-analysis of BMI including all these studies plus UK Biobank and several other large-scale GWAS efforts will comprise a sample size of >2 million individuals. Our analytical pipeline, built to deal with this sample size, will enable us to perform flexible and adaptable *ad hoc* meta-analyses, including full conditional meta-analysis and rare variant group-wise testing of both coding and non-coding regions. This large-scale multi-ethnic endeavor provides an unparalleled opportunity for fine-mapping associations to pinpoint candidate functional variants, which will speed the translation of genomic discoveries to the clinic by providing insight into functional biology of obesity.

2131T

Analysis of whole-exome sequence data to identify variants which influence Body Mass Index (BMI) via a role in adipogenesis. Y.L. Muller¹, S. Day¹, C. Koroglu¹, G. Elson¹, S. Kobes¹, R.L. Hanson¹, W.C. Knowler¹, H.I. Kim², C. Van Hout¹, N. Gosalia², A.R. Shuldiner², C. Bogardus¹, L.J. Baier¹. 1) Phoenix Epidemiology and Clinical Research Branch, Phoenix, AZ; 2) Regeneron Genetics Center, Regeneron Pharmaceuticals, Tarrytown, NY.

Body fat is regulated by changes in adipogenesis. To identify variants that potentially impact human body weight via an effect on adipogenesis, we analyzed whole-exome sequence data from a population-based sample of 6809 American Indians with longitudinal measures of BMI. ~180,673 analyzed exonic variants that met the following criteria were prioritized for functional screening 1) non-synonymous, stop-gain or frameshift and occurred in ≥ 5 individuals 2) a Combined Annotation Dependent Depletion (CADD) score ≥ 10 (probability of being the top 10% of deleterious variants) 3) evidence for association with maximum BMI recorded during adulthood ($n=6002$) or maximum BMI z-score recorded during childhood ($n=4882$) ($P<0.001$, adjusted for age, sex, birth year and the first five genetic principal components) and 4) mapped to a gene highly expressed in adipose tissue with literature support for a possible role in adipogenesis. Twenty variants met these criteria. To date 11 variants have been functionally assessed *in vitro* for roles in lipid accumulation in OP9 cells, and 2 of them showed functional evidence. A variant P891L (rs748552530) in *PTPRF* with a CADD score of 28 was identified in 7 subjects. Heterozygous carriers had a higher BMI in adulthood (54.2 vs. 36.3 kg/m², $P=0.02$) and BMI Z-score in childhood (3.0 vs. 0.33SD, $P=0.0006$) as compared to non-carriers. *PTPRF* encodes Protein Tyrosine Phosphatase, Receptor Type F, which inhibits phosphorylation of the insulin receptor. Prior studies have shown that knockdown of *PTPRF* increased adipogenesis in human mesenchymal stem cells and 3T3-L1 preadipocytes. Another variant E37K (rs144101476) in *LAMTOR1* with a CADD score of 24 was identified in 5 subjects. Heterozygous carriers had a higher BMI in adulthood and Z-score in childhood as compared with non-carriers (BMI 51.6 vs. 36.4 kg/m², $P=0.0006$; Z-score 1.5 vs. 0.33SD, $P=0.01$). *LAMTOR1* encodes Late Endosomal/Lysosomal Adaptor, which is involved in activation of the mTORC1 pathway during adipogenesis. Analyses of both variants in OP9 cells support their roles in adipogenesis, where cells expressing the BMI-risk allele 891L in *PTPRF* or 37K in *LAMTOR1* had an increase in triglyceride content compared to the wild type allele (20% with $P=0.007$; 51% with $P=0.02$, respectively). In conclusion, whole-exome sequencing followed by *in vitro* high-throughput analysis of lipid accumulation in OP9 cells has identified coding variants that may impact body weight via a role in adipogenesis.

2132F

Cardiometabolic GWAS variants colocalize with subcutaneous adipose tissue cis-eQTLs. C.K. Raulerson¹, A. Ko², T.S. Furey^{1,3}, M. Laakso⁴, P. Pajukanta^{2,5}, K.L. Mohlke¹. 1) Department of Genetics, University of North Carolina, Chapel Hill, NC; 2) Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, CA; 3) Department of Biology, University of North Carolina, Chapel Hill, NC; 4) Department of Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland; 5) Molecular Biology Institute at UCLA, Los Angeles, CA.

Genome-wide association studies (GWAS) have identified hundreds of loci for cardiometabolic traits such as type 2 diabetes (T2D), cholesterol, triglyceride levels, body fat distribution, and adiposity traits. However, the variants, genes, and mechanisms that underlie these signals are not well understood. To investigate genetic variants that influence the expression of nearby genes, we used genotypes and subcutaneous adipose tissue RNA-seq data from 434 Finns from the METabolic Syndrome in Men (METSIM) study. We tested for association between 7.8M variants and levels of 21,735 expressed genes using FastQTL. We identified 14,028 genes associated with ≥ 1 variant ± 1 Mb of the TSS at $FDR<5\%$ (local eQTLs). Among these, we focused on genes near 3,015 cardiometabolic GWAS loci from the NHGRI-EBI catalog and a recent DIAMANTE T2D GWAS preprint. We identified 200 local eQTLs targeting 241 genes for which (1) the eQTL variant is in LD ($r^2\geq 0.8$) with the lead GWAS variant; and (2) after conditioning on the GWAS variant, the conditional p-values were no longer significant. Within the colocalized signals, 79 genes corresponded to lipid loci, 68 genes to BMI loci, and 53 genes to T2D loci. GO biological process analysis of eQTL genes at GWAS loci corresponding to lipid traits showed enrichment for 8 categories, including cholesterol efflux, lipid biosynthetic process, and cholesterol homeostasis. We further conditioned on the lead eQTL variant at each significant gene and identified 2nd eQTL signals for 2,689 genes, 24 of which were colocalized with a nearby GWAS variant. For 17 loci, only a 2nd eQTL signal, and not a 1st eQTL signal, colocalized with the GWAS variant. At these 17 GWAS loci, analysis of second eQTL signals was required to identify a colocalized gene. At another 6 loci, a GWAS variant colocalized with the 1st eQTL signal for one gene and a 2nd eQTL signal for a different gene. For example, T2D GWAS variant rs3111316 colocalized with the 1st eQTL signal for *FARSA* ($p\text{-val}=2E-14$, $\beta=-0.558$) and the 2nd eQTL signal for *GCDH* ($p\text{-val}=1E-06$, $\beta=0.471$). *FARSA* encodes a tRNA synthetase and *GCDH* encodes a dehydrogenase involved in degrading lysine, hydroxylysine, and tryptophan. The GWAS signal may affect one or both of these genes to contribute to T2D. Taken together, these results suggest that careful analysis of conditionally distinct eQTL signals can aid with the identification of target genes and the interpretation of pathways and mechanisms at GWAS loci.

2133W

Understanding how natural genetic variation alters the ER stress response.

N.D. Russell¹, K.G. Owings¹, C.Y. Chow¹. Department of Human Genetics, University of Utah, Salt Lake City, UT.

The phenotypic expression of a disease can vary greatly between individuals and can manifest as differences in age-of-onset, severity, and other disease associated phenotypes. Differences in genetic background between patients can be a major contributor to these differences in disease outcomes. We are studying how natural genetic variation between individuals contributes to differences in disease outcomes by incorporating inter-individual genetic variation in my experimental design in the context of endoplasmic reticulum (ER) stress. The ER is a large eukaryotic organelle involved in protein folding. ER stress is the result of the accumulation of misfolded proteins in the lumen of the ER. If left unresolved, this can lead to cell death and disease. ER stress is an important component to many diseases and can impact disease outcomes. We hypothesize that inter-individual genetic variation in the ER stress response pathway is an important contributor to variable disease phenotypic expression. We have previously shown, in *Drosophila* and mouse, that the ER stress response is highly variable across different genetic backgrounds. Our goal is to uncover how genes and genetic variants of the ER stress response contribute to this variable phenotype. To achieve this goal, we are performing an *in vivo* study that will allow us to determine how the *cis*- and *trans*- regulatory landscape drives variation in the ER stress response in a tissue-specific manner. We are using two mouse strains, C57BL/6J (B6) and CAST/EiJ (CAST), which are very genetically divergent. B6 and CAST strains are crossed to produce F1 mice. For each response gene, we can identify *cis*- and *trans*- regulatory effects by comparing the relative allelic expression in the F1 to that of the ratio of expression between the parental strains. In each mouse, we are inducing ER stress and performing RNA-seq on multiple tissues. This will allow us to analyze the regulatory landscape and allele specific expression after the induction of ER stress in a strain and tissue dependent manner. In addition, we are using the *Drosophila* model to functionally test and validate putative genetic modifiers of the ER stress response. A modifier gene of the ER stress response could potentially be used as a therapeutic target in a number of diseases with a strong ER stress component.

2134T

Insights into the genetic determinants of diabetic kidney disease (DKD).

R.M. Salem¹, N. Sandholm^{2,3}, J.B. Cole^{4,5,6}, J.N. Todd^{4,5,6}, E.A. Valo^{4,5,6}, L.T. Hiraki^{7,8}, C.D. Liao^{7,8}, M.G. Pezzolesi⁹, C. Qui¹⁰, V. Nair¹¹, A. Smiles¹², S. Onengut-Gumuscu¹³, W.M. Chen¹³, C. Forsblom^{2,3}, M. Kretzler¹¹, K. Susztak¹⁰, C. Godson¹⁴, A.J. McKnight¹⁵, A.P. Maxwell¹⁵, H.M. Colhoun¹⁶, A.J. Krolewski¹², A.D. Paterson^{7,8}, P.H. Groop^{2,3}, S.S. Rich¹³, J.N. Hirschhorn^{4,6}, J.C. Florez^{4,5}. 1) Family Medicine and Public Health, University California San Diego, La Jolla, CA; 2) Helsinki University Hospital, Helsinki, Finland; 3) Folkhälsan Research Center, Helsinki, Finland; 4) Broad Institute of MIT and Harvard, Cambridge, MA; 5) Massachusetts General Hospital, Boston, MA; 6) Boston Children's Hospital, Boston, MA; 7) University of Toronto, Toronto, Canada; 8) SickKids, Toronto, Canada; 9) University of Utah, Salt Lake City, Utah; 10) University of Pennsylvania, Philadelphia, PA; 11) University of Michigan, Ann Arbor, MI; 12) Joslin Diabetes Center, Boston, MA; 13) University of Virginia, Charlottesville, VA; 14) University College Dublin, Dublin, Ireland; 15) Queen's University Belfast, Belfast, United Kingdom; 16) The University of Edinburgh, Edinburgh, Scotland.

DKD is a heritable, common and devastating complication affecting ~40% of those with type 1 diabetes (T1D) and is the leading cause of end-stage renal disease (ESRD). Despite strong evidence for a genetic component in DKD (heritability ~35% in persons with T1D), few DKD loci have been discovered and replicated. We performed the largest GWAS to date for T1D DKD in 19,406 European ancestry participants drawn from 17 cohorts genotyped on the Illumina HumanCoreExome BeadChip. Our GWAS identified 16 novel loci attaining genome-wide significance ($p < 5 \times 10^{-8}$) with at least one DKD-related phenotype. Four of the 16 lead variants were common (minor allele frequency [MAF] > 5%), eight were low frequency ($5\% \geq \text{MAF} \geq 1\%$) and four were very low frequency ($\text{MAF} < 1\%$). The variant with strongest association is rs55703767 ($p = 5.3 \times 10^{-12}$, MAF = 21%, missense) in the collagen type IV alpha 3 chain (COL4A3) gene, which encodes a major structural component of basement membranes implicated in Alport Syndrome. The COL4A3 association has a protective effect of the minor allele for macroalbuminuria and/or ESRD ($\text{OR} = 0.78$; $p = 5.3 \times 10^{-12}$); genome-wide significance was also reached for three other DKD subphenotype definitions. Stratified analyses in the FinnDiane cohort showed that variants near COL4A3, SNCAIP, and ALLC-COLEC11 loci were more strongly associated with DKD in individuals with HbA1c $\geq 7.5\%$ (chronically high glucose levels) than in those with HbA1c $< 7.5\%$, suggesting that these loci may be related to DKD rather than general renal disease. Similarly, the COL4A3 effect was stronger among those with longer T1D duration and those randomized to conventional treatment (with resulting higher HbA1c) in the DCCT-EDIC cohort. In sex-stratified meta-analysis of all studies, the COL4A3 rs55703767 association was stronger in men than women ($\text{OR}_{\text{men}} = 0.73$, $p = 1.3 \times 10^{-11}$, $\text{OR}_{\text{women}} = 0.85$, $P = 1.4 \times 10^{-3}$). COL4A3 expression in renal biopsy samples was also positively correlated with the degree of fibrosis in tubules ($p = 3.2 \times 10^{-4}$). Four other loci are in or near genes with known or suggestive connections to renal biology (BMP7, COLEC11, DDR1, and TAMM41). Our results implicate determinants of glomerular basement membrane function as playing a key role in DKD, suggesting that this aspect of kidney biology should be a focus of future research on the pathology, prevention, and treatment of DKD.

2135F

Deconstructing the genetic architectures Of BMI and eating behavior phenotypes. J.F. Shelton, J.R. Ashenhurst, R.I. Tennen, P. Fontanillas, 23andMe Research Team. 23andMe, Mountain View, CA.

Seventy-one percent of Americans over the age of 20 are overweight or obese. A wide range of behavioral and genetic factors contribute to high body mass index (BMI), and a significant number genes related to higher BMI act on neurological rather than metabolic pathways, highlighting the importance of addressing behavioral and emotional factors with regard to the encouragement of achieving or maintaining a healthy weight. To better understand the genetic basis of eating behaviors and their relationship with BMI, we fielded a web-based survey to over 300,000 consented 23andMe research participants covering phenotypes related to restrictive eating, emotional eating, situational eating, disinhibition, preoccupation with food, satiety, snacking behaviors, and mealtime atmosphere. Principal components analysis (PCA) of the 21 survey items yielded two broad constructs (i.e. two first principal components) that we describe as situational eating and food avoidance behavior. Regressing BMI on situational eating resulted in a model that explained 8.6% of the variance in BMI. Food avoidance behavior, on the other hand, explained very little of the variance in BMI ($r^2 = 0.2\%$). Genome-wide association analyses (GWAS) on situational eating scores showed, as expected, a strong signal in *FTO* ($\beta = -0.135$, $P\text{-value} = 6.6\text{e-}53$), and others genes associated with BMI including *TMEM18* ($\beta = 0.083$, $P\text{-value} = 9.7\text{e-}13$) and *MCR4* ($\beta = -0.069$, $P\text{-value} = 1.3\text{e-}13$). On the other hand GWAS on sub-constructs showed that anger and irritability during hunger were associated with *ERI1* ($\beta = 0.031$, $P\text{-value} = 7.2\text{e-}10$) and *VRK2* ($\beta = 0.03$, $P\text{-value} = 7.4\text{e-}10$), genes involved in neuropsychiatric pathways.

2136W

T2D disease status alters the effects of genetic variation on molecular phenotypes: A direct study. A. Viñuela¹, J. Fernandez², A. Kurbasic³, M.G. Hong⁴, S. Sharma⁵, C. Brorsson⁶, J. Adamski⁶, J.M. Schwenk⁴, E.R. Pearson⁷, S. Brunak⁶, P.W. Franks³, M.I. McCarthy², E.T. Dermizakis¹, DIRECT Consortium.

1) Department of Genetic Medicine and Development, University of Geneva, Geneva, Switzerland; 2) WTCHG, University of Oxford, Oxford, UK; 3) University of Lund, Malmö, Sweden; 4) SciLifeLab, Karolinska Institutet, Solna, Sweden; 5) Genome Analysis, Helmholtz Zentrum, Munich, Germany; 6) Systems Biology, DTU, Lyngby, Denmark; 7) Molecular & Clinical Medicine, University of Dundee, Dundee, UK.

Despite the success of GWAS and other studies in identifying genomic regions implicated in T2D, we still lack an understanding of the cellular processes which underline this increased disease risk. To better understand how genetic variants perturb molecular networks, and how these perturbations influence disease progression and status, we present data from the DIRECT consortium cohort of 3,029 subjects. This comprised of 2,234 prediabetic and 795 newly diagnosed T2D subjects, with genotype information deep clinical data and molecular phenotypes such as RNAseq, targeted and untargeted metabolite quantifications and protein (immunoassays) measurements. To explore the effects of genetic variation in cellular phenotypes, we looked for evidence of eQTLs in blood identifying 59,704 independent eQTLs associated with 94.2% genes (FDR 1%), with 78.7% genes having more than one eQTL. Compared to the eQTLs in 44 tissues produced by GTEx, we observed that 81% (testis) to 98% (brain) of the eQTLs were also active in blood, demonstrating the value of large studies in accessible tissues to uncover genetic effects in other tissues. Co-localization analysis between eQTLs and 153 GWAS variants for T2D identified 47 candidate genes with shared genetic effects. Of those, the CaVEMaN fine-mapping method proposed 13 specific causal variants (Probability > 0.9). For example, the T2D associated GWAS variant at the *RREB1* locus (rs35742417) co-localized with the eSNPs for the same gene (RTC score 0.95) and the variant showed a high probability of being causal ($P = 0.99$). Adding information on 116 targeted metabolites and 263 proteins, we found that 67.2% of the expressed genes (10,905) were significantly correlated (FDR 1%) with at least one metabolite, while 71.4% of the genes (11,577) were associated with at least one protein. Finally, to investigate how the molecular network responds to T2D progression, we looked for evidence of T2D specific genetic regulation. We identified 182 significant genotype-by-T2D interactions (FDR 1%), including rs11235825 for *ARHGEF17*. We are currently further exploring the relationship between variants, gene expression, metabolites, and disease status to understand how they are mediated. Our results will increase the understanding on the complex ways genetic regulation affects multiple cellular phenotypes and leads to disease development.

2137T

Multi-ethnic genotyping in the GENNID study: Elucidating metabolic syndrome etiology underlying linkage regions. J. Y. Wan¹, E. L. Willems², T. M. Norden-Krichmar³, S. A. Santorico^{2,3,4}, K. L. Edwards¹, American Diabetes Association GENNID Study Group. 1) Epidemiology Dept, University of California, Irvine, NC; 2) Mathematical and Statistical Sciences Dept, University of Colorado, Denver, CO; 3) Human Medical Genetics and Genomics Program, University of Colorado, Denver, CO; 4) Biostatistics & Informatics Dept, University of Colorado, Denver, CO.

Metabolic Syndrome (MetS) is characterized by a collection of impaired quantitative metabolic traits, including weight, waist circumference, triglyceride (TG) levels, high-density lipoprotein (HDL) cholesterol, fasting glucose and insulin levels, and systolic and diastolic blood pressure. The GENetics of NonInsulin-dependent Diabetes mellitus (GENNID) is a resource of multiplex families across diverse ethnic groups which includes African-American (73 families, 288 individuals), European-American (79 families, 519 individuals), and Mexican-American (113 families, 610 individuals). Previous linkage analyses using microsatellite markers identified several regions on multiple chromosomes harboring variants influencing variation among MetS traits. All samples were genotyped using the Illumina Infinium Multi-Ethnic Global array for follow-up by fine mapping of these regions. These dense markers were analyzed using both variance components linkage and association testing under the previously identified linkage peaks to identify quantitative trait nucleotides (QTNs) for individual MetS traits and combinations of traits. Our genome-wide analysis focuses on both genotyped and imputed common and rare variants and their genetic association with MetS in the presence of linkage. The results show evidence of heterogeneity among underlying QTNs and different clustering of MetS traits across ethnic groups. Further, association analysis detected QTNs in only a subset of the previously identified linkage regions and also identified putative QTNs outside these linkage regions. These results suggest that although both linkage and association testing are useful tools in family-based analyses, further work is needed to resolve the basis of these different results.

2138F

Trans-ancestral fine-mapping and causal inference of variants within urate- and gout-associated loci *SLC2A9* and *ABCG2*. W. Wei¹, R. Takei^{1,2}, C. Li^{3,4}, Z. Li^{5,6}, H. Hu⁷, M. He⁷, H. Guo⁸, Y. Shj^{3,4,5,6}, T. Merriman². 1) Department of Women's and Children's Health, University of Otago, Dunedin, Otago, New Zealand; 2) Department of Biochemistry, University of Otago, Dunedin, New Zealand; 3) Shandong Gout Clinical Medical Center, Qingdao 266003, China; 4) Shandong Provincial Key Laboratory of Metabolic Disease, The Affiliated Hospital of Qingdao University, Qingdao 266003, China; 5) Bio-X Institutes, Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders (Ministry of Education), Shanghai Jiao Tong University, Shanghai 200030, China; 6) Institute of Social Cognitive and Behavioral Sciences, Shanghai Jiao Tong University, Shanghai 200240, China; 7) Department of Occupational and Environmental Health and the Ministry of Education Key Lab of Environment and Health, School of Public Health, Tongji Medical College, Huazhong University of Science & Technology, Wuhan 430030, Hubei, China; 8) Center for Biostatistics, School of Health Sciences, The University of Manchester, Manchester, UK.

Gout is a type of complex inflammatory arthritis affecting a quarter of people with elevated serum urate levels. The prevalence of gout varies widely across ancestral groups, from 6-8% in NZ Māori and Pacific (Polynesian) people, to 3% in Europeans, 1 ~ 2% in East Asians, and 0.4% in South Americans. Genome-wide association studies (GWASs) of gout have identified two major urate-regulating loci *ABCG2* and *SLC2A9* in European and Asian samples. Identification of causal variant(s) in these loci, however, is challenging because of extensive linkage disequilibrium and / or different haplotype structures. We hypothesised that leveraging multiple ancestral genetic backgrounds will enhance fine-mapping of *ABCG2* and *SLC2A9* and inference of causal variants. We used the MANTRA software package to perform trans-ancestral meta-analysis to fine map the two loci in gout and urate separately, and then the *coloc* package to identify likely causal variants colocalised in associations with urate and gout. Using European (gout: 5,712 cases, 105,569 controls; urate: 110,347) and East Asian (gout: 1,255 cases, 1,848 controls; urate: 36,547) samples, we fine mapped the associations to a single variant resolution for both loci: rs4697701 in both urate and gout for *SLC2A9*, and rs2622621 in gout and rs4148155 in urate for *ABCG2*. The MANTRA fine-mapping results were confirmed in the colocalisation analysis of European samples where only rs4697701 in *SLC2A9* was identified as a common causal variant of urate and gout. These results are consistent with a previous inference of causal variants for urate in Europeans (Farh et al. *Nature* 518:337-43), where rs12498742 was claimed to be a causal variant in *SLC2A9* that is adjacent (2043 bases away) and in linkage disequilibrium ($r^2 = 0.77$) with rs4697701, whereas no causal variants in *ABCG2* were reported. The fine-mapping and colocalisation analyses of East Asian samples might be underpowered but suggested different causality.

2139W

Association of adiponectin gene polymorphism (-3971A/G and +276G/T) with type 2 diabetes in North Indian Punjabi population. V. Bains, B. Badaruddoza, H. Kaur. Human Genetics, Guru Nanak Dev University, Amritsar, India.

Adiponectin is the most abundant circulating adipocytokine in human that plays important role in regulation of glucose and lipid metabolism and positively regulates insulin actions. Human adiponectin is encoded by the ADIPOQ gene is located on the chromosomal locus 3q27 and spans approximately 16 kilobases (kb) of DNA in a region identified as a susceptibility locus for type 2 diabetes. Previous studies have shown that the Adiponectin gene polymorphisms are associated with insulin resistance and an increased risk of type 2 diabetes mellitus in several ethnic populations. The present study was aimed to investigate the association of adiponectin gene polymorphisms (ADIPOQ-3971A/G and +276G/T) with type 2 diabetes in North Indian Punjabi population. The study comprised of 510 unrelated (300 type 2 diabetics T2D and 310 normal controls) subjects from North Indian Punjabi Population. Anthropometric and physiometric measurements were taken from all the study participants under standardized protocols. Biochemical analysis including glucose and lipid profile was done for all the participants. The ADIPOQ variants were screened by polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) and for the validation of genotyping 10% samples of each polymorphism were confirmed by Sanger sequencing method. The differences of overall genotype distribution for ADIPOQ (-3971A/G and +276G/T) gene polymorphisms are statistically significant ($p=0.021$ and 0.046 , respectively) in cases and controls. Risk genotype (3971GG) had significantly higher TC, TG, LDL-C and VLDL-C. However, +276TT genotype had significantly higher TG and HDL-C levels only. Under genetic models, -3971A/G (rs822396) were significantly associated with T2DM under the co-dominant (OR:1.32, CI:1.04-1.68, $p=0.021$) and recessive (OR:2.05, CI:1.21-3.46, $p=0.006$) genetic models. whereas, +276G/T (rs1501299) was found to be associated with T2DM under recessive model (OR:2.00, CI:1.14-3.49, p -value=0.013) only. Results of present study supported that the ADIPOQ gene polymorphism (ADIPOQ-3971A/G and +276G/T) plays a important role in type 2 diabetes susceptibility in North Indian Punjabi Population.

2140T

Fine mapping of T2D linked 12q24 region in Finnish families. O. Dwivedi¹, T. Tuomi², L. Groop^{1,3}. 1) Institute for Molecular Medicine Finland FIMM, University of Helsinki, Finland, Helsinki, Finland; 2) Folkhälsan Institute of Genetics, Folkhälsan Research Center, Helsinki, Finland; 3) Department of Clinical Sciences, Diabetes and Endocrinology, Lund University, Malmö, Sweden.

Aim: Previously, a linkage for type 2 diabetes (T2D) was reported at the 12q24 (104-133Mb, b37) region in Finnish families participating in the Botnia Study (Mahtani M et al., Nat Genet. 14: 90-94, 1996). The linkage was observed after stratification for a low 30-minute insulin (INS 30) level, thus suggesting that genetic variation in this region could result in impaired insulin secretion and T2D. We here aimed to fine-map the linkage peak in order to identify the putative causal variant/haplotype in the region. **Material and Methods:** The linkage study included 34 individuals (20 affected and 14 non-diabetics) from six families. All family members underwent a dense GWAS (2.5 millions markers) and whole genome sequencing (WGS) for at least two affected member per family. We first identified identity by descent (IBD) shared haplotypes among families (present in ≥ 2 families) using high dense phased data. We used 253 Finnish individuals with genotype data as population controls to allow identification of family specific IBD haplotypes. We further filtered the variants present in shared IBD haplotype and looked their effect on T2D risk, insulin secretion and expression of nearby genes in 191 human islets. **Results:** Analysis of haplotype sharing revealed only one long (>3Mb) IBD haplotype (126 to 130 Mb) under the linkage peak spanning 104 to 133 Mb; two families with the lowest INS 30 levels contributed most to the linkage. We further refined this shared haplotype to a 500kb region (128 to 128.5 Mb, b37) using population controls and high density marker data (2.5 M and WGS data as available). IBD sharing in this region was rarely seen in the population controls whereas the most severely affected patients shared both haplotypes in the region. The shared IBD haplotype contains variants (frequency 0.2% to 5%) that are associated with T2D risk (OR=1.5, $P=0.044$), reduced trend of INS30 ($P=0.09$) levels and increased islet expression of nearest gene *TMEM132C* ($P=0.01$), which is negatively correlated with insulin secretion in human islets (Ottosson-Laakso E. et al. Diabetes 66:3013-3028, 2017). **Conclusion:** The *TMEM132C* gene is a plausible candidate to explain linkage of impaired insulin secretion and T2D to chromosome 12q in these families.

2141F

Phenome-wide prediction of additional effects of antidiabetic therapies using human genetics: The UK Biobank. S.M. Figarska^{1,2,6}, S. Gustafsson³, D. Lindholm^{4,5}, T. Quertermous^{1,2}, T. Fall³, J.W. Knowles^{1,2,6}, E. Ingelsson^{1,2,6}. 1) Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA; 2) Stanford Cardiovascular Institute, Stanford University, Stanford, CA 94305; 3) Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 4) Department of Medical Sciences, Cardiology, Uppsala University, Uppsala Sweden; 5) Uppsala Clinical Research Center, Uppsala University, Uppsala, Sweden; 6) Stanford Diabetes Research Center, Stanford University School of Medicine, Stanford, CA 94305, USA.

Background: For several type 2 diabetes (T2D) drugs, it is uncertain whether they have other potentially adverse or beneficial outcomes, including influencing risk for cardiovascular disease (CVD). **Methods:** We investigated downstream effects of functional variants in genes encoding T2D drug targets: dipeptidyl peptidase 4 (*DPP4*), glucagon-like 1 receptor (*GLP1R*), sodium-glucose cotransporter 2 (*SGLT2*), sulfonylurea receptor 1 (*ABCC8*), lysosomal alpha-glucosidase (*GAA*), and peroxisome proliferator-activated receptors- γ (*PPARG*) to predict possible additional effects of T2D therapies. For each gene, we tested associations of two variant sets: a) functional variants according to VEP and b) all variants in the gene (MAF<0.05) in a hypothesis-driven set of 16 phenotypes and a hypothesis-free set of 259 phenotypes comprising blocks of ICD-10 codes, groups of procedures (OPCS4) and self-reported variables in 337,536 unrelated individuals of white British ancestry from the UK Biobank. Analyses were performed using Sequence Kernel Association Test (SKAT), as well as single variant tests adjusting for age, sex, array and first ten principal components reflecting ancestry. A Bonferroni correction for multiple testing was applied. **Results:** Variations in all examined genes, except for *SGLT2*, were nominally associated with T2D in the SKAT test. In the hypothesis-driven approach, a loss of function (LoF) variant in *DPP4* was associated with lower T2D risk, BMI, total body fat percentage and heart rate. A LoF variant in *GLP1R* was associated with lower T2D risk and higher risk of acute pancreatitis. An intronic variant in *PPARG* was associated with higher T2D risk and lower body fat percentage, whereas another intronic variant in *PPARG* was associated with higher T2D risk and treatment with lipid-lowering medications. In addition to the phenotypes selected *a priori*, the LoF variant in *DPP4* was associated with lower leukocyte count, a 3' UTR variant in *GLP1R* was associated with higher risk of unspecified mental disorder, and the intronic variants in *PPARG* associated with higher T2D risk were associated with higher use of blood pressure medication and higher platelet count. **Conclusions:** In addition to confirming expected associations of variation in genes encoding T2D drug targets with T2D, we also observe associations with other phenotypes, predicting several previously suggested and some previously unknown potential additional beneficial or adverse effects of T2D drugs.

2142W

Rare variant discovery in type 2 diabetes using extremely low-coverage sequencing in large multigenerational pedigrees. S.G. Frodsham¹, J. Laza-ro-Guevara^{1,2}, M.H. Pezzolesi¹, C.A. Zaffino¹, M.G. Pezzolesi^{1,2,3}. 1) Division of Nephrology and Hypertension, Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City, UT; 2) Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, UT; 3) Diabetes and Metabolism Research Center, University of Utah School of Medicine, Salt Lake City, UT.

Type 2 diabetes (T2D) is a common complex metabolic disease with genetic underpinnings. Over the past decade, T2D genome-wide association (GWAS) and sequencing studies have advanced our understanding of the genetic basis of T2D by identifying more 100 loci that impact its risk. Variants at these loci, however, explain only a minority (~10%) of the heritability of T2D; these data suggest that alternative approaches are needed to fully elucidate the residual genetic variance of T2D. While deep coverage (>30X) whole-genome sequencing (WGS) offers the most comprehensive assessment of rare, low-frequency, and common variants, at scale, this approach remains cost prohibitive. Recently, extremely low-coverage (~0.5X) WGS (XLC-WGS) coupled with imputation has been shown to be a cost-effective alternative to deep WGS that increases the power of association studies by allowing for new/rare variant discovery. Here, we apply this approach in the context of large multigenerational pedigrees identified using the Utah Population Database (UPDB), a unique population-based genealogy resource of Utah pioneers and their descendants that is linked to electronic health record data from the University of Utah Health Sciences Center. Using the UPDB, we identified 20 large multigenerational pedigrees enriched for T2D. These pedigrees span as many as 10 generations and range in size from 1,450 to >15,000 members. To date, we have performed XLC-WGS in more than 150 T2D members of these pedigrees. Following XLC-WGS, these data were imputed using 1000 Genomes (Phase 3) to obtain variant calls across the genome. Among these samples, we identified an average of 3.3 million high-quality variants per genome, including 5,500 rare variants (minor allele frequency <0.01). These data show high concordance with GWAS chip data from these same individuals, validating the utility of this approach. Linkage and family-based association analyses are currently underway. Preliminary analyses suggest that XLC-WGS coupled with family-based analyses is a promising approach to identify novel rare variants associated with T2D.

2143T

Influence of vitamin D receptor gene polymorphisms and three autoantibodies on susceptibility to type 1 diabetes mellitus in Kuwaiti Arabs. *M.Z. Haider¹, M.A. Rasoul^{1,2}, M. Al-Mahdi², H. Al-Kandari³, G.S. Dhaunsi^{1,4}*. 1) Pediatrics Department, Kuwait University, Jabriya, Kuwait; 2) Pediatrics Department, Al-Adan Hospital, Kuwait; 3) Pediatrics Department, Farwania Hospital, Kuwait; 4) Medical Laboratory, Mubarak Al-Kabeer Hospital, Kuwait.

Type 1 diabetes mellitus (T1DM) is highly prevalent in Kuwait and its incidence has increased considerably during the past four decades. T1DM is thought to result from a complex interplay between genetic and environmental factors and the immune system of susceptible individuals. Vitamin D deficiency is widely spread around the world and is considered to be involved in immune modulation and contribute to autoimmune destruction of insulin producing beta cells in pancreas of T1DM patients. Vitamin D has been shown to exert its effects via a nuclear vitamin D receptor (VDR) and therefore, VDR gene has become a candidate for T1DM susceptibility. We have determined the genotypes of four VDR gene polymorphisms by PCR-RFLP analysis in 261 Kuwaiti Arab T1DM patients and 214 healthy controls in order to investigate their role in susceptibility to T1DM. Statistically significant differences were detected between the genotypes of two VDR gene polymorphisms (*FokI*, C/T, rs10735810 and *TaqI*, C/T, rs731236) between T1DM patients and controls ($P < 0.0001$). In the case of both these polymorphisms, the variant alleles were significantly associated with T1DM. In contrast, the VDR gene *Apal* (G/T, rs7975232) and *BsmI* (A/G, rs1544410) polymorphisms did not show an association with T1DM in Kuwaiti Arabs. The frequency of variant genotype ff, of the *FokI* VDR gene polymorphism was significantly higher in T1DM patients with late age-of-onset ($>6y$) compared to the early-age-of-onset cases ($<4y$). The same pattern was observed in VDR gene *Apal* polymorphisms with the higher frequency of variant genotype detected in late age-of-onset ($>6y$) patients. For VDR gene polymorphisms, *TaqI* and *BsmI*, no age-of-onset related variation was detected. Three autoantibodies, ICA (Islet cell), GADA (glutamate decarboxylase) and INS (insulin) autoantibodies were positively associated to, varying degrees, with T1DM in Kuwaiti Arabs harboring different VDR gene polymorphism genotypes. Our results support the role of two VDR gene polymorphisms (*FokI* and *TaqI*) in conferring susceptibility to T1DM in Kuwaiti Arabs.

2144F

Genome-wide screen for parent-of-origin effects in the association with type 2 diabetes in American Indians. *R.L. Hanson, W.C. Hsueh, S. Kobes, Y.L. Muller, S. Day, P. Piaggi, A. Nair, W.C. Knowler, L.J. Baier*. PECCB/NIDDK, Phoenix, AZ., USA.

Genome-wide association studies (GWAS) have identified multiple variants associated with type 2 diabetes; although some of these variants are subject to parent-of-origin effects, few GWAS have accounted for the parental origin of alleles. We analyzed GWAS data from 7659 Pima Indians (2571 with diabetes), consisting of 4719 nuclear families, with allowance for parent-of-origin effects. Genotypic data were derived from a custom Axiom array (Affymetrix, Santa Clara, CA) designed to capture common variation in Pima Indians, and constituted 474,196 autosomal variants with minor allele frequency ≥ 0.05 . Genotypic data were available for ≥ 1 parent for 65% of individuals. Parental origin of alleles was assigned probabilistically using a family-based likelihood method. A mixed model, which accounted for empirical genetic-relatedness estimates among all individuals, was used to analyze association with diabetes, with estimation of the odds ratio conferred by a risk allele inherited from the mother (OR_m), that conferred by a risk allele inherited from the father (OR_f), and the P-value for the null hypothesis $OR_m = OR_f$ (P_{diff}). Since the full likelihood method is computationally intensive, we first used a logistic regression model to analyze the difference in diabetes prevalence according to parental origin of alleles among perfectly informative heterozygous individuals (≥ 1 parent homozygous) across all variants, and applied the full method to the 1000 SNPs with the strongest evidence for parent-of-origin effects; it was also applied to 5529 SNPs in 6 regions containing clusters of genes known to be imprinted in humans. We identified SNPs from 8 different regions with evidence for parent-of-origin effects ($P_{diff} < 3.0 \times 10^{-5}$, corresponding to a false discovery rate $P < 0.05$ across the known imprinted regions). These included established diabetes-susceptibility variants in *KCNQ1* (rs2237895, risk allele=C, $OR_m = 1.72$, $P = 2.7 \times 10^{-11}$, $OR_f = 0.99$, $P = 0.90$, $P_{diff} = 2.3 \times 10^{-5}$), and additional variants near *LINC00670* (rs1460319, G, $OR_m = 0.76$, $P = 6.7 \times 10^{-4}$, $OR_f = 1.39$, $P = 3.3 \times 10^{-5}$, $P_{diff} = 4.2 \times 10^{-6}$), *LOC105378522* (rs7896335, T, $OR_m = 1.56$, $P = 2.7 \times 10^{-6}$, $OR_f = 0.79$, $P = 0.01$, $P_{diff} = 5.0 \times 10^{-6}$) and *LINC00250* (rs12443418, T, $OR_m = 1.58$, $P = 2.0 \times 10^{-5}$, $OR_f = 0.77$, $P = 0.02$, $P_{diff} = 1.7 \times 10^{-5}$). In addition to the established imprinted variants in *KCNQ1*, these analyses identify variants with potential parent-of-origin effects in the association with type 2 diabetes.

2145W

Genome-wide association study of diabetes progression in Chinese patients with type 2 diabetes. G. Jiang^{1,2}, C.H. Tam^{1,2}, A.O. Luk^{1,2,3}, H.M. Lee^{1,3}, C.K. Lim^{1,3}, R. Ozaki¹, E. Chow^{1,2}, A.P. Kong^{1,2,3}, C.C. Chow¹, W.Y. So^{1,2}, J.C. Chan^{1,2,3,4}, R.C. Ma^{1,2,3,4}.

1) Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong, China; 2) Hong Kong Institute of Diabetes and Obesity, The Chinese University of Hong Kong; 3) Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong; 4) Integrated Bioinformatics Laboratory for Cancer and Metabolic Diseases, The Chinese University of Hong Kong.

Disease progression and glycaemic deterioration in type 2 diabetes (T2D) are highly heterogeneous. Identification of genetic risk factors that affect disease progression may facilitate precision medicine in diabetes. We performed a nested case-control study from the Hong Kong Diabetes Register (HKDR), which includes more than 8,000 patients with T2D and prospective follow-up. We excluded patients with type 1 diabetes, those on insulin treatment, those requiring insulin within 1 year, or patients considered oral drug failure at baseline (defined as 2 consecutive HbA1c $\geq 8.5\%$, more than 3 months apart, when on two or more oral anti-hyperglycemia drugs). Disease progression was defined as progression to continuous insulin treatment during follow-up (up to 30th June 2014), or oral drug failure, whichever came first. Samples were genotyped using the Illumina Omni 2.5+ exome array and genotype data was imputed using minimac 3 (1000G phase 3v5). After standard quality control procedures pre- and post-imputation, ~8.0 million genetic variants (MAF $\geq 1\%$) were included in the study. Association analysis was performed using the Cox proportional hazards regression, adjusted for age, gender, BMI and principal components. Among 4,208 patients included in the genome-wide association analysis for diabetes progression, 1,991 (47.3%) progressed to oral drug failure or insulin treatment during median follow-up of 9.1 years (IQR 4.6-13.9). We identified 61 variants with suggestive association ($p < 10^{-5}$) with disease progression, with the top association signal located at 22q12.3 (MAF=0.19, HR [95% CI]=1.27 [1.17-1.38], $P=6.5 \times 10^{-5}$). Our study has identified a number of novel regions associated with disease progression in T2D. Replications and extension of these findings are currently in progress. **Acknowledgement:** This work was supported by the Research Grants Council Theme-based Research Scheme (T12-402/13N), the Focused Innovation Scheme, Vice-Chancellor One-off Discretionary Fund, and the Postdoctoral Fellowship Scheme of the Chinese University of Hong Kong.

2146T

Association of adiponectin levels and *AdipoQ* gene variants with T2D in females from Northwest Indian population. K. Kaur, A.J.S Bhanwer. Human Genetics, Guru Nanak Dev University, Amritsar, India.

Adiponectin is an important cytokine with anti-inflammatory and insulin sensitizing effects. Lower adiponectin levels have been associated with parameters of metabolic syndrome and Type 2 diabetes (T2D) and are further known to exhibit sexual dimorphism. In the present study, Adiponectin protein levels and five polymorphisms in the adiponectin gene (*AdipoQ*) were studied to determine their association with T2D in females (370 cases and 323 controls) from the North-West Indian population of Punjab. Protein levels were determined through ELISA; and SNPs *rs17300539*, *rs266729*, *rs2241766* and *rs1501299* were genotyped using PCR-RFLP method while *rs17846866* was typed through ARMS-PCR. Adiponectin levels were observed to be significantly different ($p < 0.001$) among T2D cases (13.84 ± 4.39) and controls (18.16 ± 6.62). Nearly 2-3 folds increased risk for T2D was observed with the T allele of *rs1501299* [$p = 1.55 \times 10^{-5}$, OR=2.04 (1.61-2.57)], and A allele of the promoter polymorphism *rs17300539* [$p = 4.96 \times 10^{-5}$, OR=3.38 (1.82-6.31)]. In haplotype analysis, three haplotypes: GCTTG [$p = 0.002$, OR=1.41 (1.13-1.75)], GGTGG [$p = 0.022$, OR=1.45 (1.06-1.99)] and GCGTG [$p = 0.024$, OR=1.70 (1.07-2.7)] significantly associated with increased risk of T2D. Two haplotypes GCTTT [$p = 1.20 \times 10^{-5}$, OR=0.56 (0.43-0.73)] and ACTTG [$p = 0.006$, OR=0.23 (0.07-0.78)] appeared to have protection against the development of T2D. Association of protein levels in the background of genotypes using one way ANOVA revealed a significant difference in the mean protein levels in cases ($p = 0.01$) controls ($p = 0.01$) for *rs17300539* and *rs17846866*, respectively. Overall, the observations are suggestive of association of reduced adiponectin levels with T2D in females from the North-West Indian population of Punjab. The association of promoter polymorphism *rs17300539* with increased risk of T2D and adiponectin levels, indicate the importance of regulatory region variants in modulating the protein levels and disease etiology. Association of intronic variant *rs1501299* reflects the role of non-coding polymorphisms in regulation of gene expression; however, the findings need to be exemplified through functional validations. In conclusion, the regulatory and structural region *AdipoQ* variants seems to dictate the adiponectin protein levels and hence T2D susceptibility in females from the North-West Indian population. The financial assistance was provided by DST-WOS(A) through Grant No.SR/WOS-A/LS-150/2009. There is no conflict of interest.

2147F

Exome-wide analysis in American Indians identifies an association between rare variants in *TGM2* and type 2 diabetes. *H. Kim¹, R.L. Hanson², N. Gosalia¹, B. Ye¹, C. Koroglu¹, W.C. Hsueh², W.C. Knowler², L.J. Baier², A.R. Shuldiner¹, C. Bogardus², C.V. Van Hout¹.* 1) Regeneron Genetics Center, Regeneron Pharmaceuticals, Inc., Tarrytown, NY; 2) Phoenix Epidemiology and Clinical Research Branch, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Phoenix, AZ.

Whole exome sequencing in founder populations with unique allelic architecture can lead to the discovery of novel disease-associated genes. To identify rare coding variants that are associated with type 2 diabetes (T2D), we sequenced exomes in American Indians including 2248 T2D cases and 4585 controls. The top gene-based association was with *TGM2*, in which 29 variants are either predicted loss-of-function or missense variants with minor allele frequency lower than 1% (SKAT: $p = 3.0 \times 10^{-7}$, burden test: OR = 2.21, $p = 2.1 \times 10^{-5}$). Further evaluation of the locus showed that two coding variants of *TGM2* that are enriched in American Indians were associated with T2D risk in opposite directions: variant p.Arg76His, c.227G>A, MAF = 0.9%, was associated with increased risk (OR = 3.05, $p = 2.3 \times 10^{-7}$), while variant p.Arg476Gln, c.1427G>A, MAF = 1.3%, was associated with reduced risk (OR = 0.61, $p = 0.010$). Furthermore, variant p.Arg76His was associated with earlier T2D age of onset (beta = -4.9yr, $p = 1.07 \times 10^{-3}$), while variant p.Arg476Gln was associated with later onset (beta = +3.8yr, $p = 0.017$). *TGM2* [MIM 190196] encodes transglutaminase 2 protein (TGM2) with enzymatic activity mediating protein cross-linkages. Published literature has implicated TGM2 in the regulation of glucose stimulated insulin secretion from pancreatic islets and glycemic control in mice. We hypothesized that genetic variants at the locus might modulate the expression or activity of TGM2, leading to the observed T2D association. We first tested whether p.Arg76His and p.Arg476Gln alter the enzymatic activity of TGM2 by in vitro enzymatic assays using wildtype and mutant recombinant TGM2 proteins. Enzymatic activity of TGM2 is activated by calcium and inhibited by GTP/GDP. TGM2-p.Arg476Gln protein exhibited elevated IC50 for GTPgS compared to wildtype protein (2.1μM versus 0.1μM), indicating that its enzymatic activity is more resistant to GTPgS inhibition; there was no difference in EC50 for calcium. This result suggests that variant p.Arg476Gln may increase the enzymatic activity of TGM2, which in turn, may reduce T2D risk. On the other hand, TGM2-p.Arg76His showed no difference in EC50 for calcium, or IC50 for GTPgS, which suggests that different mechanisms or potentially linkage disequilibrium with another variant in or near *TGM2* might be responsible for the association with T2D. Further studies are ongoing to characterize the potential impact of *TGM2* and its genetic variation on glycemic regulation.

2148W

Pancreatic islet specific long non-coding RNAs variants nominally associate with insulin secretory function and type 2 diabetes in American Indians. *P. Kumar, P. Piaggi, S. Kobes, C. Bogardus, R.L. Hanson, L.J. Baier.* Phoenix Epidemiology and Clinical Research Branch, National Institute of Diabetes and Digestive and Kidney Diseases, National Institute of Health.

Long non-coding RNAs (lncRNAs) are increasingly recognized as contribute to metabolic processes and disease pathogenesis. Therefore, we sought to determine common variation in lncRNAs that may increase risk for type 2 diabetes (T2D). Recent genome-wide transcriptome and regulome studies have identified nearly 2,600 intergenic and antisense lncRNAs in human islets. Several of these affect beta-cell differentiation and maturation. We merged the positions of known islet lncRNAs with whole-genome sequence data from 335 Pima Indians, who are at high risk for T2D, to identify common SNPs that mapped to lncRNAs. Tags for these SNPs constituting 11,447 variants, were genotyped in 7,700 Pima Indians informative for T2D, where a subset of this sample ($N = 404$) was also informative for insulin secretory function as assessed by an intravenous glucose tolerance test when the subject had normal oral glucose tolerance. We identified 16 variants in lncRNAs that nominally associated (P -value < 0.05) with both a reduced acute insulin response and increased risk for type 2 diabetes. The two strongest associations were rs2210585 which mapped to lncRNA HI-LNC4502 (P -value, T2D = 1.11×10^{-4} , OR [95%CI] = 1.20 [1.10-1.32]; P -value, insulin secretion = 4.75×10^{-3} , beta = -0.065] and a novel SNP chr6:147389391 which mapped to lncRNA HI-LNC292 (P -value, T2D = 5.01×10^{-4} , OR = 1.30 [1.12-1.50]; P -value, insulin secretion = 2.05×10^{-3} , beta = -0.103). The risk allele for rs2210585 (A) and a novel chr6:147389391 (A) had a frequency of 64% and 89% respectively in Pima Indians. These lncRNAs mapped to genes SNAP25-AS1 and STXBP5-AS1, respectively and anti-sense to their respective coding genes that belong to the soluble N-ethylmaleimide-sensitive factor attachment protein (SNAP) receptor (SNARE) dependent exocytosis pathway. SNARE proteins have a physiological role in insulin exocytosis in pancreatic beta-cells. Another variant, rs7732130 of lncRNA HI-LNC2744 nominally associated with T2D (P -value = 0.027, OR = 1.11 [1.01-1.22], risk-allele frequency = 35%) and insulin secretion (P -value = 0.047, beta = -0.047) in Pima Indians. This lncRNA variant mapped to the ZBED3-AS1 gene, associates with type 2 diabetes previously (T2D Portal, P -value = 3.49×10^{-7} , OR = 1.16, GoT2D+replication dataset) with a similar direction to our study. In summary, studying islet lncRNAs variants that associate with insulin secretion and type 2 diabetes may uncover novel pathways that modify risk for type 2 diabetes.

2149T

Dysregulation of microRNAs in DR patients in Chinese Han population.

Z. Li¹, Q. Guo², Y. Dong³, Q. Wang^{1*}, P. Chen^{1*}. 1) Department of Endocrinology, China-Japan Union Hospital of Jilin University, Changchun, Jilin, China; 2) College of Basic Medical Sciences of Jilin University, Changchun, Jilin, China; 3) Department of Radiotherapy, The Tumor Hospital of Jilin Province, Changchun, Jilin, China.

Background: Type 2 diabetes (T2D) can induce many severe complications. Diabetic retinopathy (DR) is one of the most common and severe complications and considered as the major cause of blindness among young adults. MicroRNAs (miRNAs) are a group of small noncoding RNAs regulating the expression of target genes. Several miRNAs have been reported to play a critical role in the development of DR via a variety of molecular mechanisms such as cell proliferation, angiogenesis, and inflammatory response modulation. In this study, we sought to identify miRNAs that are differentially expressed in DR patients, in the hope of unveiling the differentially regulated genes and pathways. **Methods:** We recruited T2D inpatients in China-Japan Union Hospital of Jilin University. All the patients were subjected to a detailed medical examination and retinopathy diagnosis. DR patients with duration of diabetes less than 5 years were selected as DR group (N = 10), while non-DR group was consist of retinopathy-free participants with duration of diabetes more than 10 years (N=11). Fasting peripheral serum samples were obtained at the end point. MiRNA transcription level was evaluated by next-generation sequencing. Differential expression of miRNA in DR group as compared with non-DR group was analyzed by miRDeep2 and DESeq2, adjusted for age. Target genes of miRNAs were predicted by miRDB. **Results:** There was no significant difference in BMI and sex between DR and non-DR groups. Twenty-seven miRNAs were differentially expressed in DR group with p value less than 0.01. Eight genes were found to be jointly targeted by these miRNAs, including KCNMA1, NEAT, MIAT, which has been reported closely correlated with the dysfunction of vascular endothelium. **Conclusion:** Serum circulating miRNAs may resemble the pathogenic changes in retina tissue. Using an extreme case-control study, we identified miRNAs that are differentially expressed in the serum of DR patients and the genes that are apparently regulated by these miRNAs. Our study provided evidences for future mechanical studies. However, studies with larger cohorts are needed to validate our preliminary findings.

2150F

Advanced identification of novel genetic variants for type 2 diabetes, childhood obesity and their pleiotropic loci.

X. Lin^{1,3}, C. Zeng², J. Greenbaum¹, Y. Chen³, C. Peng³, R. Liu³, Z. Feng³, L. Peng³, J. Shen³, H. Deng^{1,4}. 1) Department of Biostatistics and Data Science, Tulane University, New Orleans, LA; 2) Department of Endocrinology and Metabolism, The Fifth Affiliated Hospital of Guangzhou Medical University, Guangzhou, China; 3) Department of Endocrinology and Metabolism, The Third Affiliated Hospital of Southern Medical University, Guangzhou, China; 4) School of Basic Medical Sciences, Central South University, Changsha, China.

Type 2 diabetes (T2D) has become a serious global public health issue and caused severe health and economic burden. Numerous studies have demonstrated that childhood obesity (CO) is associated with increased risk of developing T2D in later life. In this study, we aim to identify some novel single-nucleotide polymorphisms (SNPs) associated with T2D and CO. In particular, we are especially interested to identify their pleiotropic loci. We integrated the summary statistics for two independent genome-wide association studies (GWASs) of T2D (n = 149,821) and childhood body mass index (CBMI) (n = 35,668) using the pleiotropy-informed conditional false discovery rate (cFDR) method. By leveraging the information of different levels of association for CBMI, we observed a strong enrichment of genetic variants associated with T2D based on the conditional QQ plot. We identified 139 T2D-associated SNPs including 125 novel ones (cFDR < 0.05). Conditioned on T2D, we identified 37 significant SNPs for CBMI (cFDR < 0.05), including 25 novel ones. It is noteworthy that, the conjunctural cFDR (ccFDR) analysis showed ten novel pleiotropic loci for T2D and CBMI (ccFDR < 0.05). Interestingly, the novel SNP rs1996023 is located at protein coding gene *GNPDA2* (ccFDR = 1.28E-02). Variations of this gene have been reported to influence the risk of T2D and CO through processes in the central nervous system. By integrating GWAS data from related phenotypes, we reported several pleiotropic loci shared by both T2D and CO. Our findings may help to explain a greater proportion of the heritability for human traits and advance the understanding of the common pathophysiology between T2D and CO. This work was partially supported by grants from the National Institutes of Health [U19AG05537301, R01AR057049, R01AR059781, D43TW009107, P20GM109036, R01MH107354, R01MH104680, R01GM109068], the Edward G. Schlieder Endowment fund to Tulane University.

2151W

Process-specific genetic risk scores for type 2 diabetes reveal differential effects on mechanism, phenotype, and disease course. A. Mahajan^{1,2}, N. Robertson^{1,2}, M.I. McCarthy^{1,2,3}. 1) Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, UK; 2) Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of Medicine, University of Oxford, Oxford, UK; 3) Oxford NIHR Biomedical Research Centre, Oxford University Hospitals Trust, Oxford, UK.

Over 400 type 2 diabetes (T2D) risk loci have been identified. However, most map to non-coding sequence and mechanisms of action are often unclear. One strategy for defining the physiological processes through which these variants act is to examine patterns of association to metabolic and anthropometric traits. Here, we extend such analyses to an expanded set of T2D signals. We considered lead variants at 177 loci with the strongest T2D-risk effects and widest availability of relevant GWAS data. From multi-trait association patterns generated from a set of 10 core traits, we used fuzzy clustering to define six genetic clusters (GCs), made up of variants with similar phenotypic consequences. We characterize these clusters as capturing variants with primary effects on: adiposity (GC1-BMI); insulin action (GC2-IA: low BMI and HDL, high TG); beta-cell function (GC3-BCF, GC4-BCF); and mixed features (GC5-MIX, GC6-MIX). Based on GC membership, we created a series of process-specific genetic risk scores (psGRS), and used these as instruments to explore their relationships to the heterogeneity in T2D phenotypes. Across GWAS data for 355 other traits/outcomes not used in the initial clustering, we observed GC-specific patterns of association ($p < 2 \times 10^{-6}$), consistent with the proposed mechanisms. We could also characterize the phenotypes driving membership of related GCs: for example, GC3-BCF and GC4-BCF showed differential association with insulin secretion/processing traits (reduced insulin secretion response $p = 7 \times 10^{-14}$ vs 0.005; high vs low proinsulin, $p = 7 \times 10^{-16}$ and 9×10^{-5} , respectively). We observed GC-specific associations with diverse phenotypes. Increased T2D-risk mediated by GC1-BMI was exclusively associated with renal failure ($p = 6 \times 10^{-9}$) and arterial fibrillation ($p = 5 \times 10^{-9}$). Hypertension and coronary artery disease risk were more strongly associated with T2D-risk mediated through GC1-BMI and GC2-IA ($p < 1 \times 10^{-32}$) vs other GCs ($p > 3 \times 10^{-14}$). Branched-chain amino acid levels were related exclusively to T2D-risk mediated through GC1-BMI and GC2-IA ($p < 1 \times 10^{-7}$ vs > 0.05 for other GCs). These findings demonstrate that process-specific GRSs capture the aetiological heterogeneity that contributes to overall T2D-risk, and that psGRS-differential effects can be detected across diverse clinical phenotypes. As well as enhancing understanding of the pathophysiological basis of T2D, these psGRSs provide a basis for a more personalised approach to the management of T2D.

2152T

Leveraging T2D specific omics data in rare variant association analysis in TOPMed. T.D. Majarian¹, P.S. de Vries², D. Jain³, B.E. Cade^{4,5,6}, A.K. Manning^{1,7,8}, TOPMed Diabetes Working Group. 1) Diabetes Research Group, Broad Institute of MIT and Harvard, Cambridge, MA; 2) Human Genetics Center, Department of Epidemiology, Human Genetics, and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, USA; 3) Department of Biostatistics, University of Washington, Seattle, Washington; 4) Division of Sleep and Circadian Disorders, Brigham and Women's Hospital, Boston, MA; 5) Division of Sleep Medicine, Harvard Medical School, Boston, MA; 6) Program in Medical and Population Genetics, The Broad Institute of MIT and Harvard, Cambridge, MA; 7) Clinical and Translational Epidemiology Unit, Massachusetts General Hospital, Boston, MA; 8) Department of Medicine, Harvard Medical School, Boston, MA.

In whole genome sequencing (WGS) studies of rare variants, multiple variants must be aggregated into units to have sufficient statistical power for association tests with type 2 diabetes (T2D). We incorporated T2D-relevant functional annotations in generating three separate gene-centric aggregation strategies as an alternative to agnostic, sliding window-based approaches. Using TOPMed WGS (44,732 individuals in 23 studies), T2D was defined by fasting glucose ≥ 7 mmol/L or HbA1c $\geq 6.5\%$ or 2-hr OGTT ≥ 11.1 mmol/L or non-fasting glucose ≥ 11.1 mmol/L or treatment, physician diagnosis, self-report, resulting in 9,651 T2D cases and 35,081 controls from 4 ancestries: African American, Asian, European, and Hispanic. We tested association of each aggregation unit using sequence kernel association tests (SKAT) as implemented in the GENESIS package. P-value based meta-analyses of summary statistics were also performed. We used Bonferroni correction to determine statistical significance within each of 3 gene-centric grouping strategies. We defined aggregation units with islet-specific expressed genes from previously published RNA-seq data of 89 individuals with and without T2D (aggregation 1). Each unit included regulatory or high-impact, protein truncating variants. Promoter and enhancer regions were derived from chromatin state predictions, filtered by predicted transcription factor binding sites, and linked to genes by distance and publicly available databases. Two gene-centric coding variant aggregation strategies were also used: high-impact, protein truncating variants (2) and high to moderate-impact variants, protein truncating and missense variants (3). Our results show 8,922 islet expressed genes with cumulative minor allele count greater than 10, comprised of 1 million variants. The *TMEM35B* aggregation unit was associated with T2D in the Asian ancestry analysis ($P = 1.3 \times 10^{-6}$, (1)). High impact variant aggregation (2) yielded 18,420 tests. *DNASE1* and *EXOC6B* units were associated with T2D in the Asian ancestry ($P = 1.4 \times 10^{-6}$, (2)). The *ZNF454* aggregation unit was significant in meta-analysis ($P = 1.6 \times 10^{-7}$, (2)). 19,048 units were identified in high- and moderate-impact aggregation. *MC4R* was associated with T2D ($P = 2.6 \times 10^{-6}$, (3)) in the Asian ancestry analysis. In conclusion, with islet-specific rare variant aggregation, we identified sets of variants that were likely to be both functional and act on common gene targets in the setting of T2D.

2153F

Genome-wide meta-analysis identifies a novel low frequency *STK39* variant of large effect on risk of type 1 diabetes. D. Manousaki^{1,2}, V. Forgetta¹, S. Ross^{1,3}, M.-C. Tessier¹, L. Marchand⁴, H.-Q. Qu⁷, J.P. Bradfield⁵, S.F.A. Grant^{6,9,10}, H. Hakonarson^{8,9,10}, A. Paterson¹¹, C. Piccirillo^{4,5,12,13}, C. Polychronakos^{2,6}, J.B. Richards^{1,2,3,4,14}. 1) Lady Davis Institute for Medical Research, Montreal, Quebec, Canada; 2) Department of Human Genetics, McGill University, Montreal, Quebec, Canada; 3) Department of Epidemiology and Biostatistics, McGill University, Montreal, Quebec, Canada; 4) Department of Medicine, McGill University, Montreal, Quebec, Canada; 5) Department of Microbiology and Immunology, McGill University, Montreal, Quebec, Canada; 6) Department of Pediatrics, McGill University, Montreal, Quebec, Canada; 7) Human Genetics Center, University of Texas School of Public Health, Houston, Texas, USA; 8) The Center for Applied Genomics, The Children's Hospital Philadelphia, Philadelphia, Pennsylvania, USA; 9) Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA; 10) Division of Human Genetics, Abramson Research Center, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA; 11) The Hospital for Sick Children Research Institute, The Hospital for Sick Children, Toronto, Ontario, Canada; 12) Program in Infectious Diseases and Immunology in Global Health, Centre for Translational Biology, Research Institute of McGill University Health Centre, Montreal, Quebec, Canada; 13) FOCIS Centre of Excellence in Translational Immunology (CETI), Montreal, Quebec, Canada; 14) Department of Twin Research and Genetic Epidemiology, King's College London, London, UK.

The genetic etiology of type 1 diabetes (T1D) has been extensively studied, with 57 loci identified to date, mainly through genome-wide association studies (GWAS). Most of these known genetic associations involve common variants, while a sizable portion of the missing heritability of T1D may be attributed to unidentified rare single nucleotide polymorphisms (SNPs) (minor allele frequency (MAF) < 5%). The recent availability of large human whole genome sequencing datasets enables the interrogation of rare genetic variation by imputation from directly genotyped data, in search of large genetic effects by rare alleles. Here, through deep imputation, GWAS, and meta-analysis of 12 European cohorts totaling of 9,684 T1D cases and 15,743 controls, we identified 47 independent genome-wide significant variants outside the major histocompatibility complex, among which 23 were common and 24 had MAF < 5%. All 24 rare SNPs had large effects ($OR_{discovery} > 1.5$), and only 5 mapped in previously described autoimmune loci. Six low-frequency variants successfully replicated at $p < 0.05$ and with $p_{combined} < p_{discovery}$ in a separate cohort (4,329 T1D cases from the T1DGC sibling pairs/9,543 controls from UK Biobank), 5 of which were novel. After combining discovery and replication results, all of them had an $OR \geq 1.5$, and 3 achieved an $OR \geq 2$. *In silico* analysis using topological domains and proximity to functional elements prioritized a low-frequency variant at 2q24.3 (rs60587303 (C), MAF 1.8%) within the first intron of *STK39*, with an $OR_{combined}$ of 1.97 (95% CI 1.58-2.47, $P_{combined} = 2.9 \times 10^{-9}$), an effect comparable to those of the insulin (*INS*) and *PTPN22* loci. Additional analysis showed that admixture did not unduly bias results for this SNP. This variant overlaps a DNase I hypersensitivity cluster in 68 cell types and a cluster of transcription factor bindings sites including *FOS*, *JUN* and *STAT3*. Pharmacological blocking of Stk39 in murine T-cells lines increased activation and proliferation of effector T cells and secretion of interferon gamma from CD4+ T cells. Thus, we identified 5 novel rare T1D variants, which have effect sizes larger than most previously described common SNPs. Our findings highlight *STK39* as a protein, with no previously known role in T1D, which appears to influence T-cell activation and effector functions. These findings add to the knowledge of the genetic architecture of T1D and demonstrate *STK39* as a new T1D gene with potential therapeutic implications.

2154W

Diabetes risk loci regulate the expression of nuclear-encoded mitochondrial genes. H. Maude^{1,2}, F. Tamanini¹, D. Huntley², M.J. Jarvelin³, W. Laur⁴, N. Maniatis⁴, T. Andrew¹. 1) Department of Genomics of Common Disease, Imperial College London, United Kingdom; 2) Department of Life Sciences, Imperial College London, United Kingdom; 3) Department of Epidemiology and Biostatistics, Imperial College London, United Kingdom; 4) Department of Genetics, Evolution & Environment, UCL, London, United Kingdom.

Mitochondrial dysfunction is commonly associated with Type 2 Diabetes (T2D) and has been observed in ostensibly healthy, but insulin resistant offspring of T2D patients. However, there remains a lack of genetic evidence supporting a heritable mechanism, thus supporting competing hypotheses that the observed changes are a consequence of disease, rather than a contributory cause. Since the majority of known T2D genetic risk loci (>200) are non-coding, there remains the possibility that common variants modify mitochondrial function through currently uncharacterised regulatory mechanisms. Given that non-coding variants can regulate the expression levels of local *cis*- (and distal *trans*-) genes in the genome, we hypothesize that individuals at risk of T2D possess an inherent mitochondrial dysfunction, driven by heritable changes in the expression of nuclear-encoded mitochondrial genes (NEMGs). In order to investigate this, subcutaneous adipose eQTL (expression quantitative trait loci) and T2D disease loci were co-located using high-resolution genetic maps, to estimate the location of variants that increase risk of T2D by altering the expression of *cis*-genes (here named Diabetes-eQTL). The current study identifies a number of NEMGs (>80) with evidence of being regulated by variants at Diabetes-eQTL. These *cis*-NEMGs include many which maintain general mitochondrial function, as well as genes which function in common metabolic pathways related to fuel partitioning of lipids, carbohydrates and amino acids, in addition to energy production. Further work has been carried out to validate the altered expression of these *cis*-NEMGs and implicated mitochondrial pathways using independent T2D case-control expression data, and to characterize the functional role of the *cis*-NEMGs in peripheral insulin resistance. We observe enrichment of *cis*-genes involved in branched chain amino acid (BCAA) catabolism and propose a genetic mechanism driving the increased levels of BCAAs, which have been demonstrated to be effective biomarkers for predicting T2D.

2155T

Integration of genome-wide polygenic risk scores from type 2 diabetes and related glycemic traits improves the prediction of the disease in Partners Healthcare Biobank participants. J.M. Mercader¹, A. Leong^{3,4}, V. Kaur², B. Porneala³, M.S. Udler^{2,1,4}, J.B. Meigs^{3,4}, J.C. Florez^{2,1,4}. 1) Programs in Metabolism and Medical & Population Genetics, Broad Institute of Harvard and MIT, 02142 Cambridge, Massachusetts, USA; 2) Diabetes Unit and Center for Genomic Medicine, Massachusetts General Hospital, 02114 Boston, Massachusetts, USA; 3) Division of General Internal Medicine, Massachusetts General Hospital, Boston, 02114 Massachusetts, USA; 4) Department of Medicine, Harvard Medical School, 02115 Boston, Massachusetts, USA.

Despite the identification of hundreds of loci associated with type 2 diabetes (T2D), the ability of these genetic markers to predict disease incidence is poor. Recently developed methods allow the generation of genome-wide polygenic risk scores (PRS), which improve prediction over using only genome-wide significant loci. Current PRSs developed for T2D are based on summary statistics from T2D genome-wide association studies (GWAS) and cannot distinguish between individuals who develop T2D due to insulin resistance, impairment of beta-cell function, or both. We hypothesized that combining the PRSs based on GWAS for T2D and related glycemic traits may be an alternative strategy to improve the prediction of T2D. To test this hypothesis we generated a PRS composed of 5.9M variants based on a T2D GWAS (12,931 T2D cases and 57,196 controls), 1.9M variants from a GWAS of homeostasis model assessment of beta-cell function, (HOMA-B, N=46,168) and 1.8M variants from GWAS data from homeostasis model assessment of insulin resistance (HOMA-IR, N=46,168). We computed the PRS for these three traits in 12,938 subjects participants from Partners Healthcare Biobank, a biobank from patients seen at Partners HealthCare-affiliated hospitals in Boston. Genotyping was performed with the Illumina Multi-Ethnic Genotyping Array, and imputation with the Haplotype Reference Consortium panel after applying standard quality control procedures. By combining the PRSs for T2D and both glycemic traits we observed that PRSs for T2D (OR per SD=1.4, $p=5 \times 10^{-22}$), HOMA-IR (OR per SD=1.26, $p=5 \times 10^{-4}$), and HOMA-B (OR per SD=0.8, $p=6 \times 10^{-4}$) were associated with T2D, with the three of them providing orthogonal information. The variance explained by the combination of the three genetic PRSs after accounting for sex, age and body-mass-index improved by 32% compared to the model that used only the T2D PRS. Additionally, we observed that individuals in the intersection of the top 5% percentile for T2D PRS and the top 5% percentile of HOMA-IR had 5.5 fold higher risk for T2D ($p=1.5 \times 10^{-5}$) compared to individuals in the middle of the T2D PRSs distribution, suggesting that intersection of tail-ends of PRS for several related traits can identify individuals with extreme genetic risk for T2D. Overall, we show the advantage of using multiple-phenotype PRSs to improve the prediction of T2D. This approach may be enhanced as data from additional and larger T2D and glycemic traits GWAS become available.

2156F

Distinct polymorphisms in *RNLS* are associated with Type 1 Diabetes (T1D) in European and African ancestry populations and implicate novel functional pathways. S. Onengut, R. Pickin, M. Faidas, N. Vogler, E. Gersz, S.S. Rich. University of Virginia, Charlottesville, VA.

Type 1 diabetes (T1D) is an autoimmune disease that arises from the action of both genetic and environmental factors. Genetic studies of T1D have focused almost exclusively on European ancestry populations and identified more than 50 risk loci. We recently conducted the largest T1D association study in African-ancestry participants and identified variants in the 10q23.31 T1D risk locus in European and in African ancestry populations, suggesting a common genetic contribution to T1D. The T1D risk variants in the 10q23.31 locus are located in the non-coding region of the FAD-dependent amine oxidase gene (*RNLS*). While 10q23.31 appears to be a "shared" T1D risk locus, the T1D-associated variants in European populations are distinct from the African-ancestry T1D risk variants; thus, there may be allelic heterogeneity. Our current understanding of the function of *RNLS* is limited. In previous studies, a positive *RNLS*-STAT3 feedback loop has been suggested. Thus, we evaluated the role of *RNLS* in the immune system, given the T1D association and the importance of JAK-STAT signalling in immune-mediated diseases. RNA-seq was used to compare transcriptomes of Jurkat cells expressing endogenous *RNLS* versus those overexpressing *RNLS*. RNA-seq analysis identified 390 genes that were more than two-fold differentially expressed, of which 74.1% (289/390) of genes were up-regulated and 25.9% (101/390) were down-regulated in cells overexpressing *RNLS*. Tissue-specific gene network analysis revealed that up-regulated genes mainly functioned in pathways related to chromosome segregation and cell division; in contrast, genes that were down-regulated by *RNLS* overexpression were involved in the interferon (IFN) and nuclear factor- κ B (NF- κ B) signaling pathways. Together, these results suggest a role for *RNLS* in cell-cycle related pathways and immune signaling pathways, providing new mechanistic insights of the role of *RNLS* in risk of T1D.

2157W

Chromatin accessibility patterns of a hiPSC model of islet development and type 2 diabetes risk. M. Perez-Alcantara¹, C. Honoré², M. Thurner^{1,3}, M. Hansson⁴, A.L. Gloyne^{1,3,5}, M. van de Bunt^{1,3}, N.L. Beer², M. McCarthy^{1,3,5}. 1) Wellcome Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) Department of Stem Cell Biology, Novo Nordisk A/S, Maaloev, Denmark; 3) Oxford Centre for Diabetes, Endocrinology & Metabolism, University of Oxford, Oxford, United Kingdom; 4) Department of Stem Cell Research, Novo Nordisk A/S, Maaloev, Denmark; 5) Oxford NIHR Biomedical Research Centre, Churchill Hospital, Oxford, United Kingdom.

Most variants associated with type 2 diabetes (T2D) predisposition in genome-wide association studies (GWAS) act through defects in insulin secretion, which could result from deficiencies in islet development and/or mature islet function. While functional studies have focused on the latter, we have explored the contribution of disturbed islet development to T2D pathogenesis. We used ATAC-seq to characterise patterns of open chromatin in three human fibroblast-derived iPSC lines (from different donors) differentiated along the pancreatic endocrine lineage. We obtained ATAC-seq profiles of 174,379 peaks at eight developmental stages (iPSC to beta-like cells). Chromatin accessibility profiles at each peak were clustered by weighted gene co-expression network analysis (WGCNA). Peaks within the resulting 26 modules (median=558, range=5841) that were highly correlated (Pearson's r p -value < mode [r p -value]) with their respective module profile were tested for enrichment in T2D-associated SNPs from the European ancestry DIAMANTE analysis (~900,000 participants) using fGWAS. Six clusters of chromatin accessibility profiles were enriched in T2D-associated signals, reaching maximum ATAC-seq peak counts at varying degrees of developmental maturity, including middle (posterior foregut) and late stages (endocrine and beta-like cells), with log₂ fold enrichment values from 2.49 (95% CI 1.06–3.32) to 3.92 (95% CI 0.15–5.24). Peaks within these modules were tested for enrichment in transcription factor (TF) binding sites with HOMER. Enriched (p -value $\leq 10^{-10}$) motifs included those bound by TFs implicated in islet development (e.g. *PDX1*, *MAFA*, *FOXA2*; which were expressed at those stages as measured by RNA-seq), confirming the biological relevance of this model. 1007 ATAC-seq peaks contained at least one variant from the 99% credible set, 47 with an individual posterior probability of association (PPA) above 10%. Several were near known important genes in islet development (*HNF1A* [98%], *HNF4A* [52%]) and less well-known candidates such as *WDR72* (99%) or *SLC2A10* (40%), with ATAC-seq peak counts across stages significantly (p -value<0.01) correlating with these genes' expression profiles. Integrating ATAC-seq data with epigenetic and gene expression data which have also been obtained from this iPSC-derived differentiation model will allow further exploration of the contribution made by genome regulation to disturbed human islet development and T2D pathogenesis.

2158T

Trans-ethnic meta-analysis of gestational diabetes reveals shared genetic background with type 2 diabetes. N. Pervjakova^{1,2}, J.P. Cook³, A.P. Morris^{1,3}, T. Ferreira^{4,5}, R. Mägi¹, – on behalf of the GenDip Consortium. 1) Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu, Tartumaa, Estonia; 2) Genomics of Common Disease, Division of Diabetes, Endocrinology and Metabolism, Department of Medicine, Imperial College London, UK; 3) Department of Biostatistics, University of Liverpool, Liverpool, UK; 4) 5 Big Data Institute, Li Ka Shing Center for Health for Health Information and Discovery, Oxford University, Oxford, UK; 5) Wellcome Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK.

Gestational diabetes mellitus (GDM), defined as glucose intolerance first recognized in pregnancy, has important implications for both mother and child. Offspring of mothers with GDM have an increased risk of birth complications associated with increased birth weight and adiposity, as well as for developing metabolic syndrome, type 2 diabetes (T2D) and cardiovascular disease in later life. Furthermore, in a 5-10 years perspective, it is estimated that around 50% of the women diagnosed with GDM will develop T2D. To date, the only genome-wide association study (GWAS) for GDM, in 1,399 affected Korean women, revealed associations at two loci: *MNTR1B* and *CDKAL1*. We conducted trans-ethnic meta-analysis of 21 GWAS in 5,374 cases and 346,506 controls of diverse ancestry (62.0% European, 16.5% East Asian, 3.2% Mexican-American and 18.3% Afro-Caribbean), each imputed up to reference panels from the 1000 Genomes Project or Haplotype Reference Consortium. The trans-ethnic meta-analysis, performed with MR-MEGA to allow for heterogeneity in allelic effects between ancestries, included 13,980,490 variants, after excluding those with minor allele count <5 and imputation quality <0.4 in each GWAS. We replicated both known GDM associations at genome-wide significance ($P < 5 \times 10^{-8}$): *MTNR1B* (rs10830963, $P = 2.3 \times 10^{-49}$) and *CDKAL1* (rs9348441, $P = 9.7 \times 10^{-15}$). We also identified three additional novel loci: *TCF7L2* (rs7903146, $P = 1 \times 10^{-14}$), *CDKN2A/B* (rs10811660, $P = 1.9 \times 10^{-9}$) and *LOC105369513* (rs143421658, $P = 4.1 \times 10^{-8}$). Allelic effects of GDM association signals were mostly homogeneous across ethnic groups. The exception was at the *CDKAL1* locus ($P_{het} = 3.68 \times 10^{-9}$), where the association signal was driven by East Asian ancestry GWAS ($OR_{East-Asian} = 1.67(1.46-1.89)$; $OR_{Europeans} = 1.06(1.01-1.12)$). Four of the loci (*MTNR1B*, *CDKAL1*, *TCF7L2*, and *CDKN2A/B*) have also been robustly associated with T2D, and lead variants are identical to those we have identified for GDM, supporting a shared underlying genetic contribution to both diseases.

2159F

Altered expression of *WFS1* & *NOTCH2* genes associated with diabetic nephropathy in T2DM patients. S.A. Sharaf¹, N.A. Kantoush², D.F. Ayoub², A.A. Ibrahim², A.A. Abdelaal¹, R. Abdel Aziz², M.M. ElHefnawi¹, A.N. Ahmed². 1) Clinical and Chemical Pathology Department-Faculty of Medicine Cairo University, Cairo, Egypt; 2) Clinical and Chemical Pathology Department, National Research Centre, Cairo, Egypt; 3) Internal Medicine, Faculty of Medicine, Cairo University, Cairo, Egypt; 4) Biomedical Informatics and Chemoinformatics Group, Informatics and Systems Department, Division of Engineering Research, National Research Centre, Cairo, Egypt.

Introduction: The increased incidence of type 2 diabetes mellitus (T2DM) and the importance of early identification and management of its complications, especially diabetic nephropathy (DN), have spotted the light on genetic factors that increase risk of T2DM and its related nephropathy. The present study aimed at investigating expression of (*KCNJ11*, *ABCC8*, *JAZF1*, *WFS1*, *PPARG*, *NOTCH2* and *EXOSC4*) genes in peripheral blood of T2DM patients. **Materials and Methods:** The study included 30 non-complicated T2DM patients, 30 patients with DN and 40 healthy controls. Quantitative Real Time PCR Array was used to study gene expression. **Results:** *NOTCH2* showed higher expression while *KCNJ11*, *JAZF1*, *WFS1* and *PPARG* genes showed lower expression in DN patients compared to non-complicated patients. *KCNJ11*, *JAZF1*, *WFS1*, *PPARG*, and *EXOSC4* expression showed significant negative correlation with microalbumin, while *NOTCH2* expression was significantly positively correlated with microalbumin. As regard HbA1c and studied genes expression, there was significant negative correlation between *WFS1* expression and HbA1c, while *NOTCH2*, *KCNJ11*, *JAZF1*, *PPARG*, *EXOSC4* expression didn't show significant correlation with HbA1c. Risk ratio of studied genes expression showed that *WFS1* and *NOTCH2* had highest risk ratio (30) and highest sensitivity and specificity, in relation to DN and they were the best predictors in the group of studied genes at cut off value of =0.861 for *WFS1* and =0.678 for *NOTCH2*. **Conclusions:** Altered expression of *WFS1* and *NOTCH2* genes may play a role in pathogenesis and development of DN in patients with T2DM. These results may contribute in early identification and management of DN.

2160W

Association of risk factors and *CAPN10*, *PGC-1α* genes in T2D pathogenesis along with potential microRNA targets: A study from Northwest India. R. Sharma¹, R. Kapoor², A.J.S. Bhanwer¹. 1) Human Genetics, Guru Nanak Dev University, Amritsar, Punjab, India; 2) Heart station and Diabetes Clinic, Amritsar, India.

Diabetes is a cluster of metabolic diseases characterized by hyperglycaemia that results from alterations in either insulin secretion, insulin action, or both. The prevalence of T2D is escalating especially in developing countries like India. Study enrolled 6195 individuals (3762 T2D cases and 2433 healthy controls) from North-West India. Waist circumference, waist-hip ratio, hypertension, Triglycerides, RBS, HbA1c, Percent body fat, SGOT, SGPT levels posed risk towards T2D. The association of *CAPN10* (rs3792267, rs3842570, rs5030952) and *PGC-1α* (rs3736265, rs3755863, rs8192678, rs17574213, rs2970847, rs2946385) polymorphisms were investigated in 1125 samples including 554 T2D cases and 571 healthy controls. Minor allele of *CAPN10* rs3792267 [$p=2.97 \times 10^{-4}$, 1.69(1.35-2.13)], rs5030952 [$p=3.26 \times 10^{-5}$, 1.59 (1.28-2)], *PGC-1α* rs3736265 [$p=0.006$, 1.49 (1.12-1.96)], rs2970847 [$p=0.02$, OR=1.22 (3.0-1.43)] conferred risk towards T2D progression. Significant association of *CAPN10* 121 haplotype combination played protective role against T2D [$p=0.0008$, 0.75 (0.63-0.89)]. Linkage disequilibrium (LD) analysis revealed strong LD for *PGC-1α* rs3736265 and rs2970847 polymorphisms and the haplotypes G-A-G-G-C-A, G-G-G-G-C-C of *PGC-1α* gene played protective role against T2D. Epistatic effect of *PGC-1α* rs3755863, rs8192678, rs2970847 variants was observed to confer 2.48-fold increased T2D risk. MicroRNAs (miRNAs) regulate the expression of protein-coding genes and represent potential biomarkers. In addition, we identified molecular networks and biological pathways, enriched for or targeted by miRNAs. Ingenuity Pathway analysis (IPA) predicted the miRNA targets mir-223, mir-132, mir-103, mir-142, mir-34a and mir-199a for T2D. The discovered miRNAs and their target genes and pathways represent potential clinically actionable biomarkers and therapeutic targets. The financial assistance to Rubina Sharma by UGC, UGC-CPEPA and DBT Scheme is acknowledged. The authors declare no conflict of interest.

2161T

Development and standardization of a type 1 diabetes genetic risk score for use in newborn screening and incident diagnosis. S.A. Sharp¹, S.S. Rich², A.R. Wood¹, S.E. Jones¹, R.N. Beaumont¹, J.H. Harrison¹, D.A. Schneider², J.M. Locke¹, M.N. Weedon¹, W.A. Hagopian², R.A. Oram¹. 1) University of Exeter Medical School, University of Exeter, Exeter, Devon, United Kingdom; 2) PNRI, 720 Broadway, Seattle, WA 98122, USA; 3) University of Virginia, Charlottesville, VA, USA.

Type 1 diabetes (T1D) exhibits strong familial aggregation with ~50% of the genetic risk associated with HLA region genes. A genetic risk score (GRS) is a simple, inexpensive approach to improve prediction of T1D and to distinguish T1D from other diabetes subtypes in research and clinical settings. We aimed to use large cohort studies to improve the capture of genetic variants associated with T1D to maximise the discriminative power of a T1D GRS. We analysed genetic variants from 6481 cases with T1D and 9247 controls of the Type 1 Diabetes Genetics Consortium (T1DGC) in both the HLA region and across the genome. We included interactions of variants with strongly associated HLA haplotypes to create an improved T1D GRS. We validated our findings in the UK Biobank (542 with T1D, 16652 with type 2 diabetes (T2D), and ~374,000 controls). The improved T1D GRS used information from 68 single nucleotide polymorphisms (SNPs) and accounted for interactions between 18 HLA DR-DQ haplotype combinations. T1D discrimination was validated in UK Biobank (ROC AUC=0.922, $p<0.0001$ vs previous T1D GRS). The new T1D GRS was most discriminative of early onset T1D (<5 years old, AUC=0.942) but performed well for all T1D (AUC=0.922), including discrimination of T1D vs T2D in adults (AUC=0.918, $p<0.0001$). We simulated use of the improved GRS in a newborn screening program, demonstrating performance nearly twice as efficient as HLA genotyping alone (reflecting increased familial aggregation captured by the T1D GRS). Those with a T1D GRS in the top 0.01% had over 22.8% increased risk of T1D in childhood, suggesting a future threshold for intervention trials. Simulation experiments showed >78% of childhood T1D occurs in the top 10% of T1D GRS in the population, suggesting targeting parents for education about the presenting symptoms of T1D. Our results demonstrate how an improved T1D GRS is highly useful for both classifying form of diabetes in incident cases, and improving newborn screening efforts. Given the cost-effectiveness of SNP genotyping, this approach has great clinical and research potential in T1D and could be extended to other autoimmune disorders.

2162F

Association study of three single nucleotide polymorphisms (SNPs) of *TERC*- a telomeric gene, in diabetic nephropathy patients. G. Singh^{1,2}, P. Raina¹, H.S. Sandhu¹, V. Kalotra¹, V. Sharma², I. Sharma², I. Sethi², V. Singh³, A. Marwaha⁴, E. Rai², S. Sharma², A.J.S. Bhanwer¹. 1) Department of Human Genetics, Guru Nanak Dev University, Amritsar, India; 2) Human Genetics Research Group, School of Biotechnology, Shri Mata Vaishno Devi University, Katra, J&K, India; 3) Kidney Clinic and Dialysis Centre, Opposite Govt. Medical College, Amritsar, Punjab, India; 4) Joshi Hospital and Trauma Centre, Jalandhar, Punjab, India.

Telomeres are the border region of the chromosomes which continuously loses their bases per cell cycle. Telomere length is regulated by a set of genes and any variations in these genes have been extensively linked with attrition of telomeres in complex and age related disorders, such as Idiopathic pulmonary fibrosis, Cancers, Cardiovascular diseases, Atherosclerosis and Diabetes. Since Diabetic nephropathy (DN) is a major late onset complication of diabetes and affects ~30–40% of all patients with either type 1 or type 2 diabetes (T1D or T2D) and continues to be the leading contributor to end-stage renal disease (ESRD) worldwide. Thus, this effort has been made to identify the role of telomeric gene *TERC* in DN development. In this case-control study, 1740 subjects from North India were included, 374 were DN cases, 1056 were Non-DN & Non T2D controls and 310 were T2D controls (Without DN) and were genotyped for the three SNPs, rs10936599, rs3772190 and rs16847897 from *TERC* telomeric gene using the MassARRAY system. All three SNPs were observed to be associated with DN, but rs10936599 was providing protection, whereas rs3772190 and rs16847897 were conferring risk towards DN development in all the three compared groups i.e. Non-DN, Non-DN & Non T2D and T2D without DN controls. This observation demonstrated that *TERC*, a telomeric gene does play a vital role in the development of the DN and more studies should be carried out to elucidate its contribution in the development of this complication. The financial assistance to GS by UGC-UPE Scheme is acknowledged. Financial Support from DST/SERB, Govt through project grant to SS and project grants from UGC, Govt to SS and ER are acknowledged. Financial support from UGC-CPEPA to AJSB is also acknowledged. The authors declare no conflicts of interest.

2163W

Identification of 28 novel susceptibility loci for type 2 diabetes in the Japanese population. K. Suzuki^{1,2,3,4}, M. Akiyama^{2,24}, M. Horikoshi², Y. Kamatani^{2,25}, A. Hozawa^{5,27}, A. Kadota^{6,28}, K. Ishigaki², K. Wakai², K. Kuriki², K. Matsuda², K. Tanno^{10,29}, M. Hirata¹¹, M. Sasaki^{12,30}, M. Naito^{7,13}, M. Kana^{2,14}, M. Ikeda¹⁵, M. Yamamoto^{5,27}, M. Kubo¹⁶, M. Iwasaki¹⁷, Y. Murakami¹¹, N. Iwata¹⁵, N. Sawada¹⁷, S. Ikegawa¹⁹, S. Maeda^{18,20}, S. Tsugane²¹, T. Yamaji¹⁷, Y. Ishigaki^{22,23}, Y. Okada^{4,2,26}, T. Yamauchi¹, T. Kadowaki¹. 1) Department of Diabetes and Metabolic Diseases, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; 2) Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 3) Laboratory for Endocrinology, Metabolism and Kidney Diseases, RIKEN Centre for Integrative Medical Sciences, Yokohama, Japan; 4) Department of Statistical Genetics, Osaka University Graduate School of Medicine, Osaka, Japan; 5) Tohoku Medical Megabank Organization, Tohoku University, Sendai, Japan; 6) Center for Epidemiologic Research in Asia, Shiga University of Medical Science, Otsu, Japan; 7) Department of Preventive Medicine, Nagoya University Graduate School of Medicine, Nagoya, Japan; 8) Laboratory of Public Health, School of Food and Nutritional Sciences, University of Shizuoka, Shizuoka, Japan; 9) Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan; 10) Division of Clinical Research and Epidemiology, Iwate Tohoku Medical Megabank Organization, Disaster Reconstruction Center, Iwate Medical University, Iwate, Japan; 11) Institute of Medical Science, The University of Tokyo, Tokyo, Japan; 12) Iwate Tohoku Medical Megabank Organization, Disaster Reconstruction Center, Iwate Medical University, Iwate, Japan; 13) Department of Oral Epidemiology, Graduate School of Biomedical and Health Sciences, Hiroshima University, Hiroshima, Japan; 14) Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA; 15) Department of Psychiatry, Fujita Health University School of Medicine, Aichi, Japan; 16) RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 17) Division of Epidemiology, Center for Public Health Sciences, National Cancer Center, Tokyo, Japan; 18) Division of Clinical Laboratory and Blood Transfusion, University of the Ryukyus Hospital, Nishihara, Japan; 19) Laboratory for Bone and Joint Diseases, RIKEN Center for Integrative Medical Sciences, Tokyo, Japan; 20) Department of Advanced Genomic and Laboratory Medicine, Graduate School of Medicine, University of the Ryukyus, Nishihara, Japan; 21) Center for Public Health Sciences, National Cancer Center, Tokyo, Japan; 22) Division of Diabetes, Metabolism and Endocrinology, Department of Internal Medicine, Iwate Medical University, Iwate, Japan; 23) Division of Innovation & Education, Iwate Tohoku Medical Megabank Organization, Disaster Reconstruction Center, Iwate Medical University, Iwate, Japan; 24) "Department of Ophthalmology, Graduate School of Medical Sciences, Kyushu University. Maedashi, Higashi-ku, Fukuoka, Fukuoka Japan."; 25) Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan; 26) Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Osaka, Japan; 27) Graduate School of Medicine, Tohoku University, Sendai, Japan; 28) Department of Public Health, Shiga University of Medical Science, Otsu, Japan; 29) Department of Hygiene and Preventive Medicine, School of Medicine, Iwate Medical University, Iwate, Japan; 30) Division of Ultrahigh Field MRI, Institute for Biomedical Sciences, Iwate Medical University, Iwate, Japan.

To understand genetics of type 2 diabetes (T2D) in Japanese, we conducted meta-analysis of the four genome-wide association studies (GWAS; 36,614 cases and 155,150 controls of Japanese ancestry). We identified 88 T2D-associated loci ($P < 5.0 \times 10^{-8}$) with 115 independent signals ($P < 5.0 \times 10^{-4}$), of which 28 loci with 30 signals were novel. Twenty-eight missense variants were in linkage disequilibrium ($r^2 > 0.6$) with the lead variants. Among the 28 missense variants, previously unreported three variants had distinct minor allele frequency (MAF) spectra between Japanese and Europeans (MAF_{JPN} > 0.05 vs. MAF_{EUR} < 0.01), such as the missense variants in genes related to pancreatic acinar cells (*GP2*) and insulin secretion (*GLP1R*). Trans-ethnic comparisons of the molecular pathways from the GWAS results highlighted both ethnically shared and distinct impacts of a series of pathways on T2D (e.g., monogenic diabetes). These findings shed light on the etiology of T2D in Japanese and Europeans.

2164T

Discovery of 176 novel loci for type 2 diabetes in 790,275 individuals in a multi-ethnic meta-analysis. M. Vujkovic^{1,2}, D.R. Miller², J.A. Lynch^{4,5}, S. Damrauer⁶, T. Assimes^{7,8}, J.S. Lee^{7,8}, Y.V. Sun^{9,10,11}, K. Cho^{12,13}, S.L. DuVall⁴, P. Wilson^{10,11,14,15}, D.J. Rader^{1,16}, C.J. O'Donnell^{12,13}, J. Meigs¹⁷, J.M. Gaziano^{12,13}, P.S. Tsao^{7,8}, P. Reaven¹⁸, L. Phillips^{11,19}, B.F. Voight^{1,16,20}, K.M. Chang^{1,21}, D. Saleheen^{1,2} on behalf of the VA Million Veteran Program. 1) Corporal Michael J. Crescenz VA Medical Center, Philadelphia, PA; 2) Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 3) ENR Memorial Veterans Hospital, Bedford, MA; 4) Department of Veterans Affairs Salt Lake City Health Care System, Salt Lake City, UT; 5) University of Massachusetts College of Nursing and Health Sciences, Boston, MA; 6) Department of Surgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 7) VA Palo Alto Health Care System, Palo Alto, CA; 8) Department of Medicine, Stanford University School of Medicine, Stanford, CA; 9) Department of Epidemiology, Emory University Rollins School of Public Health, Atlanta, GA; 10) Department of Biomedical Informatics, Emory University School of Medicine, Atlanta, GA; 11) VA Atlanta Health Care System, Atlanta, GA; 12) VA Boston Health System, Boston, MA; 13) Brigham Women's Hospital, Harvard Medical School, Boston, MA; 14) Schools of Medicine, Emory University, Atlanta, GA; 15) Department of Public Health, Emory University, Atlanta, GA; 16) Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 17) Massachusetts General Hospital, Harvard Medical School, Boston, MA; 18) Phoenix VA Health Care System, University of Arizona, Arizona State University, Phoenix, AZ; 19) Emory University Hospital, Atlanta, GA; 20) Department of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia PA; 21) Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

Background: The global epidemic of type 2 diabetes (T2D) is projected to reach ~592 million by 2035. To date, genetic studies have uncovered >240 loci associated with T2D, but they only partially explain the genetic heritability of T2D. While the incidence of T2D varies with sex, age, and anthropometry, the genetic relationships have not been fully elucidated. In addition, the genetic mechanisms leading to T2D complications are largely unknown. To characterize the genetic etiology of T2D and its major complications, we performed genome-wide association studies in the Million Veteran Program (MVP), a mega-biobank established by the Veterans Health Administration. **Methods:** The current analyses comprise 790,275 individuals (156,233 with T2D), including 63,051 European, 21,234 African and 7,342 Hispanic T2D cases enrolled in MVP. Stratified by self-reported ethnicity, we conducted primary discovery analyses for T2D in MVP controlling for age, gender and 10 PCA's followed by an overall meta-analysis involving participants from the UK Biobank, DIAGRAM, PROMIS, MDC, RACE, MedStar, PennCath and PHF. Given the availability of individual participant data in MVP, the largest analysis of the X-chromosome to date was performed. SNP-T2D effect decomposition by BMI was done with causal mediation analysis. Stratified analyses for three macrovascular complications of T2D (coronary heart disease, stroke and peripheral vascular disease) and three microvascular complications (retinopathy, nephropathy and neuropathy) were also performed. **Results:** Our preliminary analyses identified 840 autosomal variants of genome-wide significance ($P < 5 \times 10^{-8}$) at 393 distinct genomic regions; of these, 166 were previously not reported. Five of these novel loci were observed in African Americans only, which can be attributed to differences in allele frequencies, LD patterns, and potential effect heterogeneity. Chromosome X analysis identified an additional 10 novel loci, some of which displayed effect heterogeneity by sex. Finally, mediation analysis elucidated several loci that conferred T2D susceptibility through BMI. **Conclusion:** This large-scale multi-ethnic meta-analysis identified 176 novel T2D risk loci. Some T2D loci had effects that differed by African American ancestry, sex, diabetic complications, or obesity.

Study	T2D	Controls
MVP	91,627	148,082
UK Biobank	27,932	331,092
DIAGRAM	26,676	132,532
PROMIS	7,083	10,310
MDC	1,417	5,721
RACE	612	2,039
MedStar	358	964
PennCath	278	1,113
PHF	250	2,189
Total	156,233	634,042

2165F

Single cell heterogeneity analysis and CRISPR screen identify key β cell-specific disease genes. C. Weng¹, Z. Fang¹, H. Li¹, R. Tao², W. Ma³, X. Liu⁴, L. Lu¹, S. Lai¹, Q. Duan⁴, C. Alvarez⁵, P. A'rv'an⁶, A. Wynshaw-Boris¹, Y. Li¹, Y. Pei², F. Jin^{1,7}, Y. Li¹. 1) Department of Genetics and Genome Sciences, School of Medicine, Case Western Reserve University, Cleveland, OH; 2) Children's National Medical Center, Washington, D.C; 3) Department of Neurology, The Fifth Affiliated Hospital of Sun Yat-sen University, Zhuhai, Guangdong Province, China; 4) Department of Biostatistics, Department of Genetics, Department of Computer Science, University of North Carolina, Chapel Hill, NC, USA; 5) The Biomedical Sciences Training Program (BSTP), School of Medicine, Case Western Reserve University, Cleveland, OH, USA; 6) Division of Metabolism, Endocrinology, and Diabetes, University of Michigan Medical Center, Ann Arbor, MI, USA; 7) Case Comprehensive Cancer Center, Case Western Reserve University, Cleveland, OH, USA.

Identification of human disease signature genes typically requires samples from many donors to achieve statistical significance. Here we show that single cell heterogeneity analysis may overcome this hurdle by significantly improving the test sensitivity. We analyzed the transcriptome of 39,905 single islets cells from 9 donors and observed β cell heterogeneity trajectories associated with obesity or type II diabetes (T2D). We therefore developed RePACT, a general single-cell analysis algorithm to identify both common and specific signature genes for obesity and T2D with high sensitivity. We further mapped both β cell specific genes and disease signature genes to the insulin regulatory network identified from a genome-wide CRISPR screen. Our integrative analysis discovered the previously unrecognized roles of the cohesin loading complex and the NuA4/Tip60 histone acetyltransferase complex in regulating insulin transcription and release. Our study demonstrated the power of combining single-cell heterogeneity analysis and functional genomics to dissect the etiology of complex diseases.

2166W

Subjects at the phenotypic extremes of the baseline T2D phenotype spectrum differ in genetic risk and disease progression over 36 months: a DIRECT study. A. Wesolowska-Andersen¹, C. Brorsson², A. Mari³, R. Koivula⁴, A. Kurbasic⁴, A. Mahajan¹, J. Fernandez¹, A. Viñuela⁵ on behalf of the DIRECT Consortium. 1) University of Oxford, Oxford, Oxfordshire, United Kingdom; 2) Technical University of Denmark, Lyngby, Denmark; 3) National Research Council, Padova, Italy; 4) Lund University, Malmö, Sweden; 5) University of Geneva, Geneva, Switzerland.

Type 2 diabetes (T2D) is a complex common disease with heterogeneous presentation and progression. The stratification of patients into more homogeneous subgroups is the foundation of precision medicine: however current T2D diagnosis based purely on a single biomarker - glucose - is not reflective of the disease heterogeneity. Here, we investigated detailed clinical phenotypes from 725 subjects with newly-diagnosed type 2 diabetes from the DIRECT study, comprising anthropometric, biochemical and glycemic modelling variables. We applied reversed graph embedding to define a principal tree spanning the phenotype spectrum for 30 of these clinical variables. At the extremes of this baseline phenotype spectrum, we find the subjects with the highest involvement of processes underlying T2D etiology: beta-cell dysfunction, insulin resistance, obesity, and dyslipidemia. Using the tree branch points, we identified 5 subgroups of subjects who differ significantly in their clinical presentation at baseline, disease progression at 18 and 36 months, and patterns of genetic predisposition to T2D. We observed significant differences in disease progression between subgroups, evaluated in two ways: the slope of change in glycated haemoglobin (HbA1c) at 18m ($p=1.14e-03$), and the proportion of patients starting or increasing dose of diabetes medication within each subgroup (18m, $p=1.03e-06$; 36m, $p=7.13e-10$). While the overall T2D genetic risk score (GRS) constructed using all 403 reported T2D GWAS loci did not differ between subgroups, we observed significant differences with process-specific GRSs, in particular the insulin secretion GRS ($p=8.46e-08$). Additionally, we sought circulating multiomic signatures of the subgroups, detecting, at baseline, significant differences for 287 genes (RNA expression) and levels of 87 proteins and 192 metabolites at 5% FDR. These signatures can be used to better stratify patients, and to highlight additional underlying mechanisms, such as enrichment of neutrophil gene expression signatures in the fastest progressing subgroup ($p=1.18e-24$). In conclusion, the heterogeneity observed in patients with newly-diagnosed diabetes can be described by different degrees of impairment of underlying biological processes. Identification and characterization of patients at the phenotypic extremes is an important step towards implementation of precision medicine in T2D.

2167T

Type 2 diabetes and osteoporosis: A Mendelian randomization Study. Q. Zhang¹, W.D. Zhang², J. Greenbaum¹, C.Q. Sun², H.W. Deng¹. 1) Tulane University, New Orleans, LA; 2) Zhengzhou University, Zhengzhou, China.

Traditional epidemiological studies suggest that there is an association between type 2 diabetes (T2D) and bone mineral density (BMD) at the sites of the femoral neck and lumbar spine (FNK and LS BMD), indicating a potentially important relationship between T2D and the development of osteoporosis (OP). However, these findings may be influenced by unmeasured confounding factors that can obscure the true relationship between the phenotypic traits. Therefore, we performed Mendelian randomization (MR) analyses to determine whether there is a causal relationship between T2D and BMD (FNK and LS BMD), where T2D may increase the risk of developing OP. Adopting a two-sample MR approach we incorporated genome-wide association (GWAS) summary statistics from the DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) (a total of 26,488 cases and 83,964 control) consortium and the GENetic Factors for Osteoporosis (GEFOS) Consortium (n=53,236). Using this MR approach we found that T2D is not causally associated with FNK BMD ($b = 0.022$, $se = 0.018$, 95% CI (-0.013, 0.057), $p = 0.229$) and LS BMD ($b = 0.028$, $se = 0.025$, 95% CI (-0.021, 0.077), $p = 0.269$), which deviates from the traditional study results. This study demonstrates no causal relationship between T2D and OP etiology and may provide novel insights into the pathophysiology of bone related diseases. This work was partially supported by grants from the National Institutes of Health [U19AG05537301, R01AR057049, R01AR059781, D43TW009107, P20GM109036, R01MH107354, R01MH104680, R01GM109068], the Edward G. Schlieder Endowment fund to Tulane University.

2168F

Exome sequencing identifies a nonsense variant in DAO associated with reduced energy expenditure and respiratory quotient in American Indians. P. Piaggi¹, C. Koroglu¹, A.K. Nair¹, Y.L. Muller¹, W.C. Hsueh¹, S. Kobes¹, A.R. Shuldiner², H.I. Kim², N. Gosalia², C. Van Hout², M. Jones², R.L. Hanson¹, C. Bogardus¹, L.J. Baier¹. 1) NIDDK, Phoenix, AZ; 2) Regeneron Genetics Center, Regeneron Pharmaceuticals, Inc., Tarrytown, NY.

Obesity and energy expenditure (EE) are heritable and genetic variants altering EE and substrate utilization may contribute to the development of obesity. The aim of this study was to identify genetic variants that affect EE in American Indians, an ethnic group with a high prevalence of obesity where lower EE predicts long-term weight gain. Whole-exome sequencing was performed in 373 healthy Pima Indians (224 men, age: 27 ± 6 yr, BMI: 34 ± 7 kg/m², body fat: $33 \pm 8\%$) who underwent a 24-h session in a whole-room indirect calorimeter to measure EE and respiratory quotient (RQ, an index of substrate utilization) during energy balance. Genotypic analyses of 24-h EE were performed for 60,231 exonic variants (≥ 5 carriers) after adjusting for age, sex, body composition, physical activity, the first 5 genetic principal components, and accounting for genetic relatedness among individuals by pairwise identity-by-descent estimates. Due to the limited statistical power for rare variants, only coding variants annotated to be damaging by prediction tools (CADD, SIFT, or PROVEAN) were considered for follow-up studies. A total of 17 exonic variants predicted to be pathogenic by at least one algorithm showed a p -value $< 10^{-3}$ for association with 24-h EE (effect sizes ranged from 43 to 268 kcal/day). Rs752074397 (C/A, CADD=35) introduces a premature stop codon (Cys264Ter) in DAO, and demonstrated the strongest association for 24-h EE ($p = 10^{-5}$), where the Ter allele (7 carriers) associated with substantially lower 24-h EE ($b = -268$ kcal/day) and sleeping EE ($b = -135$ kcal/day, $p = 0.01$). This variant also associated with lower RQ ($b = -0.03$, $p = 2 \times 10^{-3}$), reflecting markedly reduced carbohydrate oxidation rate ($b = -197$ kcal/day, $p = 7 \times 10^{-3}$). Analysis of rs752074397 in a population-based sample of 7686 Pima Indians showed that the Ter allele has a frequency of 0.5% and 0.2% in full vs. mixed heritage Pima Indians, while public databases show that this variant is extremely rare in Latinos (ExAC frequency=0.009%) and monomorphic in other ethnic groups. DAO encodes D-amino acid oxidase, which is involved in dopamine synthesis which might explain its role on modulating EE. Functional studies are ongoing to determine the effect of rs752074397 on DAO protein expression and activity. Our results suggest that a nonsense mutation in DAO may influence EE and substrate oxidation in American Indians. Identification of variants that influence energy metabolism may lead to new pathways to treat human obesity.

2169W

Heterozygotes for a founder mutation in SLC12A3 that causes autosomal recessive Gitelman syndrome is associated with lower serum potassium in the Amish. X. Wan^{1,2}, J.A. Perry¹, H. Zhang¹, E. Streeten¹, B.D. Mitchell¹, A.R. Shuldiner^{1,3}, M. Fu¹. 1) Program in Personalized and Genomic Medicine, and Division of Endocrinology, Diabetes & Nutrition, Department of Medicine, University of Maryland School of Medicine, Baltimore, MD; 2) Department of Endocrinology and Diabetes Center, The First Affiliated Hospital of Sun Yat-sen University, Guangzhou, China; 3) Regeneron Genetics Center, Tarrytown, NY.

Potassium levels regulate multiple physiological processes, including nerve and muscle function, hormone secretion and action, glucose and insulin metabolism, and fluid and blood pressure homeostasis. To identify functional variants that contribute to normal physiologic variation in serum concentrations of potassium, we performed whole exome sequencing in 5045 Old Order Amish subjects. We identified a cluster of variants spanning ~537 Kb on chromosome 16q13 showing exome-wide significant association with age-, age²- and sex-adjusted potassium levels ($P = 5 \times 10^{-6}$ to 6.57×10^{-9}). These variants tag a large haplotype block that includes 22 genes. Among the associated variants is a missense variant in *SLC12A3* (rs200697179, R642G), the gene encoding solute carrier family 12 member 3, a renal thiazide-sensitive sodium-chloride cotransporter that is important for electrolyte homeostasis. Mutations in this gene cause autosomal recessive Gitelman syndrome. The mutation, R642G, in its recessive form is a known pathogenic variant. Consistent with a founder effect and genetic drift, the minor allele frequency (MAF) of R642G is 650 times higher in the Amish (MAF: 0.013; 4912 wild type homozygotes and 133 heterozygotes) than in other populations (MAF: 0.0002 in 1000 Genomes). The heterozygous carriers had markedly lower potassium levels than wild type carriers (4.053 mmol/L vs 4.201 mmol/L; $P = 1.89 \times 10^{-6}$). The heterozygous carriers also had slightly lower serum chloride levels, lower blood urea nitrogen (BUN) levels and higher serum CO₂ levels than wild type carriers ($P = 1.88 \times 10^{-3}$ to 7.16×10^{-3}). In a subset of 469 wild type homozygotes and 10 R642G heterozygotes subjected to 7 days of a low-salt diet (40 mmol Na/day), carriers had nominally higher resting heart rate (79.7 vs. 74.2 bpm, respectively; $p = 0.025$) but no significant difference in blood pressure or in heart rate or blood pressure in the same subjects subjected to a high salt diet (300 mmol Na/d), suggesting increased heart rate as a compensatory change for mild hypovolemia on a low-salt diet. In summary, we have identified a locus on chromosome 16q13 that is significantly associated with potassium levels and confirmed a heterozygous mutation (R642G), in *SLC12A3* that may influence potassium levels at the population level. These findings support the growing realization that pathogenic variants causing many canonically recessive disorders may manifest an intermediate often sub-clinical phenotype in heterozygotes.

2170T

Phenome-wide patterns of genotype-environment (GxE) interactions.

R. Moore^{1,2}, I. Barroso¹, O. Stegle². 1) Wellcome Sanger Institute, Cambridge, United Kingdom; 2) European Bioinformatics Institute, Cambridge, United Kingdom.

Phenome-wide association studies (PHEWAS) are a useful tool to study the effect of loci on multiple phenotypes, and have established new links between biological processes or diseases not previously known to be related. However, such pleiotropy for genotype-environment (GxE) interactions has been poorly explored. Phenome-wide GxE analyses may reveal greater aetiological overlap between traits than conventional PHEWAS, or identify loci that are enriched for interaction effects. To explore phenome-wide GxE effects, we used seven cardiometabolic traits from the UK Biobank data ($N_{\text{max}} = 251,443$), focussing on a set of 57,328 variants (1,104 loci, 1MB window, LD $r^2 < 0.1$) significantly associated ($P < 5 \times 10^{-8}$) with basal metabolic rate (BMR), and a set of 64 lifestyle factors, including dietary and physical activity variables. Interaction effects were evaluated using StructLMM (bioRxiv doi: 270611), a computationally efficient method for detection and characterisation of loci interacting with one or more environmental variables. There are two main results from our GxE analysis. First, this analysis detects many interaction loci. Whilst only eight significant interaction loci are identified for BMR (conditional FDR 5%; estimated heritability 0.29), in contrast, for other traits with lower heritability (< 0.27), including BMI and hip circumference, we detect ≥ 30 distinct interaction loci. We also note that 32% of the detected interaction loci are shared across at least three phenotypes, suggesting that pleiotropic GxE effects are common. Secondly, estimating the proportion of phenotypic variance explained by interaction effects, reveals substantial differences across traits and for some traits, interaction effects explain more of the phenotypic variation than that explained by direct genetic effects. Namely, interaction effects explain 0.20 times the amount of phenotypic variation explained by direct genetic effects for BMR, and 1.35 times for diastolic blood pressure. Together, our results provide first insights of the GxE landscape across multiple phenotypes.

2171F

Genome wide assessment for resting heart rate and shared genetics with type 2 diabetes and metabolic traits. Y. Guo¹, W. Chung¹, Z. Shan², J. Li³, S. Liu³, L. Liang^{1,4}. 1) Program in Genetic Epidemiology and Statistical Genetics, Department of Epidemiology, Harvard School of Public Health, Boston, MA; 2) Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA, USA; 3) Departments of Epidemiology, Medicine, and Center for Global Cardiometabolic Health, Brown University, Providence, Rhode Island, USA; 4) Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA.

Abstract Background: Resting heart rate (RHR) has recently been related with type 2 diabetes (T2D) and cardio-metabolic traits. However, genetic correlations and shared genetics of RHR with T2D and cardio-metabolic traits were largely unknown. **Methods:** In the current study, we performed large scale cross trait and transcriptome wide association scan (TWAS) analysis to investigate genetic correlations and shared genetics of RHR with T2D and cardio-metabolic traits using UK Biobank data (N~500K), public available GWAS data for T2D and eight cardio-metabolic traits, and GTEx tissue expression data. **Results:** We found significant genetic correlation of RHR with T2D ($r_g=0.19$, 95% confidence interval [CI]: 0.13 to 0.24, $P=2.55 \times 10^{-11}$) and five metabolic traits: FI ($r_g=0.24$ [0.15 to 0.33], $P=3.08 \times 10^{-7}$), WHR (0.15 [0.09, 0.21], $P=5.81 \times 10^{-7}$), FG ($r_g=0.13$ [0.05 to 0.21], $P=0.001$), TG ($r_g=0.13$ [0.07, 0.19], $P=1.60 \times 10^{-5}$), and BMI ($r_g=0.10$ [0.06, 0.14], $P=2.38 \times 10^{-5}$), as well as a marginal significant negative association ($r_g=-0.09$ [-0.16, -0.02], $P=0.009$) with HDL. No significant genetic correlations were observed with LDL and TC. Cross-trait LD Score regression by eleven functional categories showed similar results. TWAS analysis identified nine unique genes were shared between RHR and T2D in multiple tissues. Three out of these nine unique shared genes: *BMP8A*, *PABPC4*, and *RP11-69E11.8*, were located in the same region at chromosome 1p34.3, indicating important roles of this region in linking RHR to T2D, potentially through vascular calcification. In addition, we further identified 57 unique genes shared between RHR and at least one metabolic trait. **Conclusion:** The current study confirmed significant genetic correlation of RHR with T2D and cardio-metabolic traits, and identified shared genetics between these traits, which would help clarifying mechanisms linking RHR with T2D and cardio-metabolic disorders.

2172W

GWAS of insulin resistance measures in Hispanic/Latino adolescents from the Santiago Longitudinal Study (SLS). V.L. Buchanan¹, Y. Wang¹, E. Blanco², M. Graff³, C. Albala³, J.L. Santos⁴, B. Ange⁵, B. Lozoff⁶, V.S. Voruganti⁶, S. Gahagan⁷, H.M. Highland¹, A.E. Justice⁷, K.E. North¹. 1) Department of Epidemiology, University of North Carolina at Chapel Hill, NC, USA; 2) Division of Academic General Pediatrics, Child Development and Community Health at the Center for Community Health, University of California at San Diego, CA, USA; 3) Department of Public Health Nutrition, Institute of Nutrition and Food Technology, University of Chile, Santiago, Chile; 4) Department of Nutrition, Diabetes and Metabolism, School of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile; 5) Pediatrics and Communicable Diseases, University of Michigan, Ann Arbor, MI, USA; 6) Department of Nutrition and UNC Nutrition Research Institute, University of North Carolina at Chapel Hill, Kannapolis, NC, USA; 7) Biomedical and Translational Informatics, Geisinger Health, Danville, PA, USA.

Introduction: Genetic underpinnings of insulin resistance, which can lead to prediabetes and type 2 diabetes (T2D), have been understudied in Hispanic/Latino (HL) populations and adolescents, despite shouldering an increasing burden of obesity and T2D. We therefore conducted a genome-wide association study (GWAS) of 3 insulin resistance traits—fasting glucose, insulin, and HOMA-IR (homeostatic model assessment of insulin resistance)—in HL adolescents from the Santiago Longitudinal Study (SLS). **Methods:** Participants were genotyped on the Illumina Multiethnic Genotyping Array (MEGA) and imputed using the 1000 Genomes Phase III AMR reference panel. We conducted linear regression, assuming an additive genetic model, adjusting for sex and population substructure. In order to provide a relatively normal distribution for the 3 traits, both fasting insulin and HOMA-IR were natural log-transformed. All participants (n=543) were aged 16-17 years, and because age did not appear to have a meaningful effect on these variables, it was not included in the model. We only included study participants without diabetes, and sensitivity analyses considered adjustment for BMI. Quality control included individual call rate >90%, SNP call rate >95%, imputation quality >0.5, minor allele count >10, gender mismatch, and ancestry outliers. **Results:** For HOMA-IR, we identified a novel GWAS-significant association between rs77465890-C and decreased HOMA-IR on chromosome 8 ($p < 5 \times 10^{-8}$, MAF: 0.10, β : -0.32, SE: 0.057) within the *CSMD1* gene, and 14 additional independent signals with suggestive evidence for statistical significance ($p < 5 \times 10^{-6}$, +/-500kb). We also generalized several SNP-trait associations previously observed in other study populations. Notably, the novel SNP association we identified is in a locus previously associated with obesity-related traits in Mexican-American participants from the Viva La Familia Study. **Conclusion:** We demonstrated associations between common variants with insulin resistance phenotypes in Chilean adolescents, generalized some loci and reported concordant direction of effects with published findings. Importantly, we observed novel associations indicating possible age- or population-specific genetic effects on insulin resistance. Our findings support the importance of conducting GWAS in ancestrally diverse populations and across life stages to better understand the inherent genetic architecture of complex traits.

2173T

Genome wide association study of body weight, body mass index, adiposity, and fasting glucose in 3,184 outbred HS rats. A.S. Chitre¹, O. Polesskaya¹, K. Holl², J. Gao¹, R. Cheng¹, A. Horvath³, A. Hughson³, T. Wang⁴, A.G. Martinez¹, W. Han⁴, J.A. Trip⁵, C.P. King⁵, A. Lamparelli⁵, T. George⁶, C. Martin⁶, K. Ishiwari⁶, H. Chen¹, S.B. Flage¹, P. Meyer⁶, J. Richards⁴, T.E. Robinson⁷, A.A. Palmer¹, L.C. Solberg Woods⁸. 1) Department of Psychiatry, University of California San Diego, La Jolla, California; 2) Medical College of Wisconsin, Milwaukee, Wisconsin; 3) Molecular and Behavioral Neuroscience Institute, University of Michigan, Ann Arbor, Michigan; 4) Department of Pharmacology, The University of Tennessee Health Science Center, Memphis, Tennessee; 5) Department of Psychology, University at Buffalo, Buffalo, New York; 6) Research Institute on Addictions, University at Buffalo, Buffalo, New York; 7) Department of Psychology, University of Michigan, Ann Arbor, Michigan; 8) Department of Internal Medicine, Wake Forest Baptist Medical Center, Winston-Salem, North Carolina.

Obesity is a global health crisis that is influenced by both genetic and environmental factors. Although many genes have been identified for body mass index (BMI) and obesity in human genome-wide association studies, the majority of genetic drivers remain unknown. Studies in rodents often employ standardized housing conditions that curtail the impact of alleles that alter food consumption, exercise and other factors that are likely to be identified in humans. Our study used 3,184 N/NIH heterogeneous stock (HS) rats from the NIDA Center for GWAS in outbred rats. Although the rats were produced as part of a behavioral genetics project, we also measured numerous physiological traits, including body weight, BMI, fat pad weight (retroperitoneal, epididymal, and parametrial), and fasting glucose levels. We obtained >6 million SNPs per individual by using genotype by sequencing (GBS). We used a linear mixed model to account for complex familial relationship among the HS rats. We estimated the proportion of variance attributable to SNPs (SNP heritability h^2) using GCTA-GREML. We used the GCTA Bivariate GREML analysis to estimate genetic correlations (r_G). The SNP heritability estimates for these traits ranged from 0.1 to 0.4. We observed strong evidence of pleiotropic effects across multiple phenotypes. For example, body weight and body length showed a genetic correlation of 0.787 and shared overlapping QTLs on chromosomes 3, 7, and 12. A QTL on chromosome 1, which has been previously identified for retroperitoneal fat in rats, showed pleiotropic effects on body weight, BMI, epididymal, retroperitoneal and parametrial fat pad weight. Finally, the genetic correlation between epididymal and retroperitoneal fat weights was high ($r_G = 0.788$) and both had a QTL on chromosome 6. This work underscores the potential of rodent studies to supplement our understanding of molecular genetic factors that contribute to obesity-relevant traits and to provide novel mechanistic insights into the underlying genetic architecture of these traits.

2174F

Targeted metabolomics and genome-wide association study identifies plasma amino acids and genetic variants associated with non-alcoholic fatty liver disease. A. Huertas-Vazquez¹, E. Ocampo-Medina², P. Leon-Mimila³, L. Macias-Kauffer⁴, H. Villamil-Ramirez³, S. Hui¹, A. Lusi¹, F. Campos-Perez², N. Mendez-Sanchez⁴, F. Campos-Perez², S. Canizales-Quinteros⁴. 1) University of California, Los Angeles, Los Angeles, CA; 2) National University of Mexico. Mexico City; 3) Clinica Integral de cirugía para la obesidad y enfermedades metabólicas. Hospital Ruben Lenero. Mexico City; 4) Liver Research Unit. Medica Sur Clinic and Foundation. Mexico City.

Background: Non-alcoholic fatty liver disease (NAFLD) manifests as a spectrum of abnormalities ranging from steatosis to non-alcoholic steatohepatitis (NASH). The incidence of steatosis and its progression to NASH varies among different populations, where Mexicans have the highest prevalence. Due to the growing prevalence and lack of FDA approved therapies, increased scientific knowledge of the genetic and molecular factors underlying NAFLD is urgently needed. In this study, we aimed to identify plasma metabolites and investigate the effects of genetic variation modulating metabolite levels and risk of NAFLD in mice and humans. Methods: A total of 43 plasma metabolites were measured in 425 morbidly obese Mexican subjects who underwent bariatric surgery using a targeted mass-spectrometry based platform. We investigated the correlation between metabolites and histological features of NAFLD using the Kleiner scoring system that includes steatosis grade, hepatocellular ballooning, lobular inflammation and NAFLD activity score (NAS). Genome-wide association analysis was performed to identify genetic variants associated with plasma metabolites using a Multi-ethnic genotyping array. To investigate the potential causal role of identified variants, we performed correlations with existing liver and adipose tissue gene expression and clinical data from the Hybrid Mouse Diversity Panel (HMDP) including liver fibrosis, insulin resistance, HOMA-IR and liver triglycerides. Correlation p-values were generated from the bcor coefficient implemented in R. Results: We observed that plasma levels of tyrosine, alanine, proline and methionine were associated with NAFLD and its components including hepatocellular ballooning and lobular inflammation. In addition, we identified a population-specific missense variant in the proline dehydrogenase 1 (*PRODH*) gene, to be significantly associated with proline levels ($P < 5 \times 10^{-6}$) and NAS score ($r = 0.83$, $p = 0.0002$) after Bonferroni correction. In the HMDP, *Prodh* liver expression was strongly correlated with liver triglycerides and fibrosis. Conclusions: We identified several plasma metabolites associated with NAFLD and its components and identified a population specific genetic variant in the *PRODH* gene to be associated with proline levels and risk of NAFLD.

2175W

Identification of 12q24 locus strongly associated with sweet taste preference in Japanese populations by genome-wide meta-analysis. K. Kawafune¹, T. Hachiya^{1,2}, S. Nogawa¹, K. Saito¹, S. Takahashi¹. 1) Genequest Inc., Tokyo, Japan; 2) Genome Analytics Japan Kabushikikaisha, Tokyo, Japan.

Sweet taste preference is important for human to get carbohydrates as energy source, while it can be a risk factor for diseases such as obesity and diabetes. It is known to be strongly influenced by ethnicity, but its genetic background remains unclear. Here we performed a genome-wide meta-analysis of 12,312 Japanese participants recruited via the direct-to-consumer genetic testing service of Japan, who completed internet-based questionnaires on sweet taste preference. Their genomic DNA was extracted from saliva and was genotyped for ca. 300,000 SNPs. The participants were divided into eight groups by the residential area, and for each group the association between SNPs and sweet taste preference scale was tested by a linear regression model. Genome-wide meta-analysis of the results from all regional groups was performed using a fixed-effect model and the inverse-variance weighting method. P-value $< 5 \times 10^{-8}$ was regarded as genome-wide significance. The genome-wide meta-analysis revealed a strong association between 12q24 locus and sweet taste preference scale ($P = 2.8 \times 10^{-70}$). The lead variant rs671 for this locus which is located at a coding exon of the aldehyde dehydrogenase gene (*ALDH2*), is known for its association with alcohol drinking and commonly found in East Asian populations. The association between minor allele of rs671 and sweet taste preference scale was attenuated when additionally adjusted for alcohol consumption ($P = 7.8 \times 10^{-27}$) or frequency of alcohol drinking ($P = 6.5 \times 10^{-10}$), suggesting the mediation by drinking habits at least partially. The subgroup analysis stratified by sex showed that the effect of rs671 on sweet taste preference scale of males was greater than females (P for interaction < 0.001). Our data suggested that alcohol drinking was inversely associated with sweet taste preference, which was inconsistent with a previous study that reported that alcohol drinking was positively associated with sweet taste preference in non-Asians. In conclusion, we found the association between the 12q24 locus and sweet taste preference. Our findings may provide new clues to understand the difference among ethnicities as well as to shed light on mechanisms influencing personal food preference in Japanese.

2176T

HRC-based genome-wide association study of uric acid in the Korean population. B. Kim¹, SK. Cho^{1,3}, W. Myung², S. Kim¹, HH. Won^{1,3}. 1) Samsung Advanced Institute of Health Sciences and Technology (SAIHST), Sungkyunkwan University, Seoul, Republic of Korea.; 2) Department of Neuropsychiatry, Seoul National University Bundang Hospital, Seongnam, Republic of Korea; 3) Co-corresponding authors.

Background Abnormal level of serum uric acid (SUA) is an important marker and risk factor for complex diseases such as gout and chronic kidney disease. Although SUA level is heritable, most genetic determinants of SUA are not known in East Asians, particularly for hypouricemia. Since genome-wide association studies (GWAS) were designed to examine common variants, the differential effect between common and lower-frequency variants has yet to be fully explored. In the present study, we conducted GWAS of SUA using the Haplotype Reference Consortium (HRC)-based imputation to identify novel common and lower-frequency variants in the Korean population. **Method** We measured SUA levels of 6,881 individuals and performed a meta-analysis of GWAS of SUA in two different cohorts from the Korean Genome and Epidemiology Study; a nation-wide cohort from urban areas (3,585 individuals) and a nation-wide cohort from rural areas (3,296 individuals). Genotyping was conducted on the Affymetrix 6.0 array and the Illumina Omni1 array for the urban cohort and the rural cohort, respectively. After quality control, genotype data was phased by using Eagle 2.3 and imputed on the reference panel of the HRC (HRC r1.1 2016) by using Minimac3. We performed GWAS using linear regression adjusted for sex, age, BMI, eGFR, diabetes, hypertension, systolic blood pressure, total cholesterol, HDL cholesterol, LDL cholesterol, and triglycerides. The inverse variance-weighted fixed-effects meta-analysis was conducted using METAL. **Results** We identified 16 independent genetic variants (2 lower-frequency and 14 common) that reached a level of genome-wide significance ($P < 5 \times 10^{-8}$). Of note, a rare, nonsense variant in the *URAT1* gene (*SLC22A12*) showed the strongest association with SUA ($P = 7 \times 10^{-9}$; $\beta = -1.15$ mg/dl; $SE = 0.07$ mg/dl). Although it was reported as a pathogenic variant of familial renal hypouricemia (RHUC1, OMIM #220150), this variant has not been identified in previous GWAS of the general population. Other significant genetic variants were found to be located near genes, including *ABCG2*, *SLC2A9*, *NRXN2*, *PYGM*, *SF1*, *NAA25*, *PCNX3*, *EHBP1L1*, *SLC17A3*, *SLC17A2*, and *BCAS3*. **Conclusions** In this study, we identified novel variants, including one rare variant, that were significantly associated with SUA in the Korean population. We also demonstrated the utility of the large-scale, high-density imputation panel to identify novel lower-frequency variants from GWAS array data.

2177F

Meta-analysis of genome-wide association studies for body fat distribution in 694,649 individuals of European ancestry.

S.L. Pulit^{1,2,3}, C. Stoneman⁴, A.P. Morris^{5,6}, A.R. Wood⁴, C.A. Glastonbury⁴, J. Tyrrell⁴, L. Yengo⁷, T. Ferreira¹, E. Maroulis⁸, Y. Jie¹, J. Yang^{7,9}, S. Jones⁴, R. Beaumont⁴, D.C. Croteau-Chonka¹⁰, T. Winkler¹¹, A.T. Hattersley⁴, R.J.F. Loos¹², J.N. Hirschhorn^{3,13,14,15}, P.M. Visscher^{7,9}, T.M. Frayling⁴, H. Yaghootkar⁴, C.M. Lindgren^{1,3,6}, *The GIANT Consortium*. 1) Big Data Institute, Li Ka Shing Center for Health Information and Discovery, Oxford University, Oxford, UK; 2) Department of Genetics, Center for Molecular Medicine, University Medical Center Utrecht, Utrecht, The Netherlands; 3) Program in Medical and Population Genetics, Broad Institute, Boston, MA, USA; 4) University of Exeter Medical School, University of Exeter, Royal Devon and Exeter NHS Trust, Exeter, UK; 5) Biostatistics Department, University of Liverpool, Liverpool, UK; 6) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 7) Institute for Molecular Bioscience, University of Queensland, Brisbane, Australia; 8) William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK; 9) Queensland Brain Institute, The University of Queensland, Brisbane, Queensland, Australia; 10) Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; 11) Department of Genetic Epidemiology, University of Regensburg, Regensburg, Germany; 12) The Charles Bronfman Institute for Personalized Medicine, The Mindich Child Health and Development Institute, the Icahn School of Medicine at Mount Sinai, New York, USA; 13) Department of Genetics, Harvard Medical School, Boston, MA, USA; 14) Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, MA, USA; 15) Department of Pediatrics, Harvard Medical School, Boston, MA, USA.

Obesity has reached pandemic proportions: 25% of adults worldwide are either overweight (body mass index, BMI ≥ 25 kg/m²) or obese (BMI ≥ 30) and >10% of children meet obesity criteria. While obese individuals are at higher risk of some cancers, infertility, and cardiometabolic disease, epidemiological studies show that the location of fat storage, rather than general adiposity, is more informative for predicting disease risk. Independent of BMI, individuals with higher central adiposity have higher cardiometabolic disease risk and Mendelian randomisation studies indicate a putative causal effect of higher central adiposity on cardiometabolic disease. To pinpoint genetic variants associated to fat distribution, we performed a meta-analysis of waist-hip ratio adjusted for BMI (WHRadjBMI). WHR is an easily-measured fat distribution phenotype well-correlated with imaging-based fat distribution measures. Implementing a linear mixed model to account for relatedness and ancestral heterogeneity, we performed genome-wide association studies (GWAS) of WHRadjBMI in UK Biobank, followed by meta-analysis with publicly-available data. The meta-analysis comprised 694,649 samples and ~27.4M SNPs and identified 346 associated loci ($p < 5 \times 10^{-8}$) containing 463 independent signals. Given the known sex-dimorphism of fat distribution in humans, we also performed sex-specific meta-analyses. We identified ~3x more loci associated in women (266 loci) than men (91) and found SNP-based heritability to be strikingly different between the sexes ($h^2_{\text{women}} = 25.6\%$, $h^2_{\text{men}} = 16.7\%$, $p = 9 \times 10^{-8}$). Of all index SNPs ($p < 5 \times 10^{-8}$ in the combined or sex-specific analyses), 105 were sex-dimorphic (Bonferroni $p < 3.3 \times 10^{-4}$) and 97 of these 105 showed stronger effects in women than men. In the combined analysis, index SNPs in 12 loci correlated with nearby loss-of-function variants, including stop gains in an insulin like growth factor binding protein (*IGFBP3*) and a gene expressed in testis (*AKR1E2*). Pathway analyses mapping associated SNPs from the combined analysis to eQTLs in 53 tissues revealed enrichment of association signal in adipose tissue, while analysis of associated SNPs in women revealed an enrichment not only for adipose but also in female reproductive tissues including uterus, cervix and vagina. Additional work, including fine-mapping and further tissue and cell-enrichment analyses, will help elucidate the causal mechanisms and biological interplay underlying fat distribution and its comorbidities.

2178W

Genetic study of birth weight resolves maternal and fetal genetic effects and provides insight into the role of the intrauterine environment on future risk of offspring cardiometabolic health.

N.M. Warrington¹, R.N. Beaumont², M. Horikoshi^{3,4,5}, F.R. Day⁶, K.K. Ong^{6,7}, M.I. McCarthy^{3,4,8}, J.R.B. Perry⁶, D.M. Evans^{1,9}, R.M. Freathy^{2,9}, *EGG Consortium Birth Weight Collaborators*. 1) Diamantina Institute, University of Queensland, Brisbane, Australia; 2) Institute of Biomedical and Clinical Science, University of Exeter, Exeter, UK; 3) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 4) Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, UK; 5) RIKEN, Centre for Integrative Medical Sciences, Laboratory for Endocrinology, Metabolism and Kidney diseases, Yokohama, Kanagawa, 230-0045, Japan; 6) MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Cambridge, UK; 7) Department of Pediatrics, University of Cambridge, Cambridge, UK; 8) Oxford National Institute for Health Research (NIHR) Biomedical Research Centre, Churchill Hospital, Oxford, UK; 9) MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK.

Birth weight (BW) is influenced by both fetal and maternal factors, and has been reproducibly associated with adult cardiometabolic health. These associations are hypothesised to reflect lifelong consequences of an adverse intrauterine environment. We previously demonstrated that genetic effects contribute to the negative correlation between BW and adult cardiometabolic traits. However, no previous study has systematically distinguished the direct fetal genetic effect on BW and cardiometabolic health from the indirect effects of their mother's correlated genotype, mediated by the intrauterine environment. We aimed to (1) discover novel BW loci, (2) distinguish between fetal and maternal genetic contributions to BW, using a statistical method based on structural equation modelling (SEM) and (3) use the fetal- and maternal-specific genetic effects to elucidate BW-later health relationships. We report a large genome-wide association meta-analysis (GWAS) of BW using fetal (N=321,223) and maternal (N=230,069) genotypes from 50 studies including the UK Biobank and studies in the EGG Consortium. We identified 305 lead SNPs at 274 loci (210 novel) associated with BW ($P < 5 \times 10^{-8}$). We decomposed the direct fetal and indirect maternal genetic influences on BW using SEM, and were able to categorize lead SNPs at the majority of loci into those with fetal genetic effects only (N=83), maternal effects only (N=45), directionally-concordant (N=36), or directionally-opposing maternal and fetal effects (N=24). Our findings implicate several fetal-specific or maternal-specific mechanisms/tissues involved in the regulation of BW. We used Mendelian randomization to explore the causal relationships between BW and cardiometabolic traits using the fetal- and maternal-specific genetic effects. These analyses indicated that the inverse genetic correlation between BW and systolic blood pressure (SBP) can be explained by a combination of (1) the indirect effects of maternal SBP-raising alleles reducing BW and (2) the direct effects of SBP-raising alleles inherited by the offspring increasing their own SBP. Conversely, direct fetal genotype effects dominate the inverse relationship between BW and type 2 diabetes risk. This work represents a significant advance in GWAS of perinatal outcomes by using novel methods to separate fetal from maternal genetic effects, and has set a precedent for future analyses interrogating the causal role of the maternal environment on offspring outcomes.

2179T

Whole-exome sequencing identifies rare coding variants associated with erythrocyte membrane fatty acids. X. Yin^{1,2}, A. Locke³, A. Jackson^{1,2}, H. Stringham^{1,2}, M. Laakso⁴, M. Boehnke^{1,2}. 1) Biostatistics, University of Michigan, Ann Arbor, MI; 2) Center for Statistical Genetics, University of Michigan, Ann Arbor, MI; 3) McDonnell Genome Institute, Washington University School of Medicine, Saint Louis, MO, 63108, USA; 4) Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland.

Introduction: Erythrocyte membrane fatty acids (EMFAs) are circulating biochemical compounds that are associated with the risk for a wide spectrum of human diseases including type 2 diabetes and neurological disorders. Recent EMFAs genome-wide association studies have detected > 60 associations within > 20 genetic loci; most of them are common non-coding variants. The role of rare coding variants has not been studied. **Methods:** We measured 20 EMFAs (six saturated, five mono-unsaturated, and nine poly-unsaturated), three summed measures of saturated, mono-unsaturated, and poly-unsaturated fatty acids, and six ratios of EMFA concentrations using gas chromatography technology in 2,223 Finnish men from the METabolic Syndrome In Men (METSIM) study. These study participants have been exome sequenced on the HiSeq2500. We performed single variant and gene-based association analyses for each of the 29 EMFA traits under the linear mixed model framework to account for population structure and sample relatedness. **Results:** We validated known single-variant associations at *ELOVL5*, *AGPAT4*, *AGPAT5*, *FADS1*, *CPT1A*, *LPCAT3*, *PDXDC1*, and *CERS4*. We also identified significant novel associations ($P < 5 \times 10^{-6}$) at two rare missense variants (MAF=0.045%) within *CSMD2* and *ATP11A*. In gene-based association analysis, we constructed four gene masks of putative coding variants. We found the *CERS4* gene attained significance even after conditioning on known *CERS4* variants. We also discovered novel significant gene-level association evidence in *ABHD3* ($P_{\text{gene}} = 8.37 \times 10^{-10}$), *CPSF1* ($P_{\text{gene}} = 1.70 \times 10^{-6}$), *FAM83E* ($P_{\text{gene}} = 1.69 \times 10^{-6}$), *TRPV1* ($P_{\text{gene}} = 2.42 \times 10^{-10}$), and *AKR1E2* ($P_{\text{gene}} = 2.20 \times 10^{-6}$), due to the aggregated effect of rare coding variants. These genes newly-implicated for EMFAs have been previously suggested for multiple human diseases/traits, for example *ABHD3* in the hemostasis of plasma phospholipid levels.

2180F

Systematic identification of genes underlying obesity risk in adipocytes. G.T. Hansen, Y.I. Li, D.R. Sobreira, A.C. Joslin, L.E. Montefiori, M.A. Nobrega. University of Chicago 920 E 58th St, Chicago IL, 60637.

Obesity is a highly heritable condition affecting a third of individuals in the United States. Over 700 loci have been previously associated to obesity risk from genome-wide association studies (GWAS), yet like many other diseases, the total identified heritability is distributed over many variants, each of which contribute a miniscule amount to overall disease risk. Because the vast majority of obesity risk variants identified in GWAS are located in non-coding regions of the genome, it has been challenging to pinpoint exactly which genes drive obesity. Here, we performed a transcriptome-wide association study (TWAS) using GTEx gene expression data in subcutaneous adipose tissue to identify genes whose expression underlies obesity risk. We used summary statistics from a GWAS meta-analysis on obesity in ~750,000 individuals (Yengo et al., 2018) and identified 113 genes whose expression is associated with increased obesity risk. Interestingly, these genes show functional enrichment for responses to oxidative stress and protein localization. To systematically validate the link between obesity-associated variants and genes nominated by our TWAS, we collected ATAC-seq and promoter capture HiC data in a primary human adipocyte cell culture. We investigated the chromatin state and physical interactions between GWAS SNPs and identified genes. ATAC-seq and promoter-capture HiC show colocalization between chromatin accessibility and interactions at the promoters of TWAS-identified genes. Thus, our TWAS identified genes represent promising targets for further research, and potential new targets of therapies for the prevention and treatment of obesity.

2181W

Looking beyond the nearest gene: A high throughput functional approach using tissue-specific RNAi knockdowns to move from “statistical loci” to “functional genes”. M.A. Province¹, J.L. Fink², P.A. Lenzini¹, I.B. Borecki³, C.T. Liu⁴, L.A. Cupples⁵, K.E. North⁵, T.J. Baranski⁵. 1) Department of Genetics, Division of Statistical Genomics, Washington University School of Medicine, SAINT LOUIS, MO; 2) Department of Internal Medicine, Division of Endocrinology, Metabolism and Lipid Research, Washington University School of Medicine, SAINT LOUIS, MO; 3) Department of Biostatistics, University of Washington, Seattle, WA; 4) Department of Biostatistics, Boston University School of Public Health, Boston, MA; 5) Department of Epidemiology, University of North Carolina, Chapel Hill, NC.

GWAS have identified thousands of novel complex trait loci but knowledge of function is lagging. The loci are undeniably tagging real biology with overwhelming statistical significance, and independent replication, but effect sizes are clinically trivial. Few loci seem to tag coding variation (perhaps only 12%), making identification of the gene of action difficult. Regulome databases are helpful but much of the annotation is either lacking, insufficient, or not specific enough to allow definitive gene mapping. Functional experiments in high order organisms is daunting, expensive, and time consuming. We chose a high throughput *Drosophila* system, to efficiently functionally screen the set of nearby human genes that have fly orthologs for many loci simultaneously. Starting with 50 known Waist-Hip Ratio GWAS loci, we identified all human genes within 250kb of the most significant tag SNP (442 genes). 140 of these had *Drosophila* orthologs. We crossed RNAi knock down (KD) lines for each gene with cg-GAL4 tissue-specific driver line to obtain brain and fat body KD crosses. We measured adult percent body fat compared to wild type (WT), using Dunnett's multiple comparison test in ANOVA to assess significance. For 17 of the 50 WHR loci (34%), we found at least one -fat functional gene in the fly whose human ortholog was near the GWAS locus. All of these significant KDs were fatter than WT. For 12 of the 17, GTEx validated that the index SNP was an eQTL for the same nearby genes in human tissue corresponding to the fly brain or fat body. For 5 loci, more than one nearby gene was fat-functional in the fly, suggesting that the initial GWAS locus was tagging the influence of multiple nearby genes. The significant, GTEx validated SNP-human gene pairs are: *RS10804591*→*IFT122*, *RS11231693*→*ESRRA*, *RS11231693*→*FERMT3*, *RS17819328*→*MKRN2*, *RS17819328*→*SYN2*, *RS303084*→*NUDT6*, *RS3805389*→*TMEM165*, *RS4765219*→*ATP6V0A2*, *RS6090583*→*EYA2*, *RS6090583*→*TP53RK*, *RS6784615*→*NT5DC2*, *RS6784615*→*TWF2*, *RS714515*→*SUCO*, *RS7917772*→*SUFU*, *RS905938*→*DCST1*, *RS905938*→*SCAMP3*, *RS9991328*→*HERC3*. For only 2 of the 17 loci was the nearest human gene the fly-functional one (*EYA2* and *FGFR4*). For all others, other nearby genes were the functional in the fly and validated in GTEx. This “low hanging fruit” approach will not work for every locus or trait, but it may substantially help move the field from “statistical loci” to “functional genes”.

2182T

Role of *FAM13A* in insulin resistance: *FAM13A* gene expression in human and mice demonstrates tissue specific activity. A. Rao¹, M. Fathzadeh², N. Cook³, J. Li², I. Chennamsetty², M. Gloudemans⁴, N. Abell⁵, S. Montgomery^{6,8}, M. Shin⁷, D. Reilly⁸, J. Yang⁹, T. Quertermous², E. Ingelsson², J. Knowles². 1) Department of Bioengineering, Stanford University, Stanford, CA, USA; 2) Division of Cardiovascular Medicine, Cardiovascular Institute and Diabetes Research Center, Stanford University School of Medicine, Stanford, CA, USA; 3) Department of Medical Sciences, Molecular epidemiology, Uppsala University, Sweden; 4) Department of Biomedical Informatics, Stanford University, Stanford, CA, USA; 5) Department of Genetics, Stanford University, Stanford, CA, USA; 6) Department of Pathology, Stanford University, Stanford, CA, USA; 7) Genetics & Pharmacogenomics, Merck & Co Inc., Boston, MA, USA; 8) Translational Medicine at Merck, Kenilworth, NJ, USA; 9) Department of Comparative Biosciences, College of Veterinary Medicine, University of Illinois at Urbana-Champaign, Urbana, IL, USA.

Insulin resistance (IR) is a major driver of T2D and results in increased cardiovascular risk partly due to IR-associated dyslipidemia (e.g. high plasma triglyceride and low plasma HDL-cholesterol levels). Increases in indices of body fat distribution (e.g. waist-hip-ratio adjusted for BMI, WHRadjBMI) are correlated with IR. Large genome-wide association studies (GWAS) of fasting insulin adjusted for BMI (FladjBMI), a proxy for intravenously assessed IR, and WHRadjBMI showed a significant positive association at the *FAM13A* locus. These findings suggest that *FAM13A* might regulate insulin and fat distribution even in normal-weight individuals, a “lipodystrophy-like” phenotype. We characterized the role of *FAM13A* across human tissues by integrating GWAS data with tissue-specific gene expression, and analyzed tissue specific gene expression from *Fam13a* knockout (KO) mice. **Humans:** We performed colocalization analysis of *FAM13A* gene expression in subcutaneous adipose (SAT), visceral adipose (VAT), liver and skeletal muscle from the Genotype Tissue Expression project (v7) with 7 IR-related GWAS traits using eCAVIAR. Using a modified colocalization posterior probability (mCLPP) computed in eCAVIAR, we observed colocalizations (mCLPP > 0.8) between SAT *FAM13A* expression and FladjBMI, WHRadjBMI, triglycerides and HDL-cholesterol. We did not observe similar colocalizations in the other tissues, suggesting that *FAM13A* plays an active and differential role in SAT. **Mice:** To evaluate the biological effect of *Fam13a* KO, we performed differential expression (DE) analysis in VAT and SAT tissue between wild-type and KO male mice on both chow and high fat diet (total n = 32 tissue samples). After adjustment for the effects of diet and type of adipose tissue, we observed 255 differentially expressed genes at 5% FDR. Pathway analysis showed a strong enrichment of oxidative phosphorylation and a depletion of immunodeficiency pathways in KOs. To understand the differential impact of *Fam13a* KO across tissues, we performed a DE analysis stratified by adipose tissue type adjusted for sex and diet. We observed a much more pervasive effect of the KO on the transcriptome in SAT (96 DE genes) compared to VAT (0 DE genes) at 5% FDR. In conclusion, our results establish *FAM13A* as the causal gene in the locus; indicate that the primary effect is in SAT; and highlight oxidative phosphorylation and immunodeficiency as key biological pathways underlying the association with IR.

2183F

Togetherness of lysinuric protein intolerance and HOIP deficiency in a boy: SLC7A7 and RNF31 gene disruptions. L. Aliyeva¹, S. Erdol², O. Gorukmez³, H. Gurkan⁴, Y. Yerali⁵, B. Baytan⁶, H. Saglam⁷, K. Terali⁸, S. Kilic⁹, S.G. Temel¹. 1) Uludag University, Faculty of Medicine, Department of Medical Genetics, Bursa, Turkey; 2) Uludag University, Faculty of Medicine, Department of Pediatric Metabolism; 3) Sevkett Yilmaz Government Research Hospital, Department of Medical Genetics, Bursa, Turkey; 4) Trakya University, Faculty of Medicine, Department of Medical Genetics, Edirne, Turkey; 5) Uludag University, Faculty of Medicine, Department of Pediatric Immunology, Bursa, Turkey; 6) Uludag University, Faculty of Medicine, Department of Pediatric Haematology, Bursa, Turkey; 7) Uludag University, Faculty of Medicine, Department of Pediatric Endocrinology & Metabolism, Bursa, Turkey; 8) Near East University, Faculty of Medicine, Department of Biochemistry, Nicosia, Cyprus.

Lysinuric protein intolerance (LPI) is a rare autosomal inherited disease caused by defective cationic amino acid at the basolateral membrane of epithelial cells in the intestine and kidney. LPI is a multisystemic disease with a variety of clinical symptoms such as hepatosplenomegaly, osteoporosis, hypotonia, developmental delay, pulmonary insufficiency or end-stage renal disease, anemia, bone marrow may show hemophagocytosis. Inherited, complete deficiency of human HOIP, a component of the linear ubiquitination chain assembly complex (LUBAC), underlies autoinflammation and infections. Here we report on a 8 months old boy with LPI and HOIP deficiency a third child born to a consanguineous couple and delivered via caesarian section. The main clinical findings developmental delay, pulmonary insufficiency, hepatosplenomegaly and hypotonia. Laboratory findings anemia, thrombocytopenia, hypertriglyceridemia, high serum ferritin (35.000) and lactate dehydrogenase levels. Bone marrow morphology is compatible with hemophagocytosis. During follow-up respiratory distress was progressively increased, HSM was not regressed, recurrent infections and autoinflammation findings were observed. Whole exome and targeted sequencing in the proband revealed the presence of one homozygous missense variant c.1573C>T, this variant leads to a p. Arg525Cys change in *RNF31* and deletion of exome 4-11 of *SLC7A7* gene, respectively. Heterozygous *RNF1* variants were confirmed in both mother and father. This variant was found in ExAC (#9) and in CentoMD (#1) databases as heterozygote manner not reported as homozygote manner and regarded as disease causing and damaging based on the in-silico prediction tools (mutation taster, SIFT, Provean). As far as we know our case is the first reported case of homozygous *SLC7A7* deletion and homozygous *RNF31* variation seen together, and the second reported case of HOIP deficiency.

2184W

Prediction of metabolic syndrome in Korean females by a genetic risk score combining SNPs for female specific metabolic syndrome-related quantitative traits. Y. Cho. Hallym University, Chuncheon, Korea, Republic of.

Metabolic Syndrome (MetS) has plagued the many lives of human worldwide and females have been reported to be relatively more vulnerable to this cluster of risks. In order to elucidate genetic variants underlying MetS specifically in females, we performed gender stratified genome-wide association study (GWAS) for metabolic traits, obesity and blood pressure in a total of 13,485 Korean subjects. Our analyses detected female-specific association of rs38796 in *NECTIN2* ($P_{\text{meta}} = 1.95 \times 10^{-9}$) for low density lipoprotein cholesterol (LDLC) and of rs848652 in *MYL10* ($P_{\text{meta}} = 2.19 \times 10^{-9}$) for blood gamma-glutamyltransferase level (GTP). To facilitate the prediction of MetS in females, we calculated a genetic risk score (GRS) combining top 7 SNPs detected from our GWA analyses for MetS-related traits and conducted logistic regression analyses between the GRS and the increased prevalence of MetS in females. A total of 645 MetS cases and 598 controls were grouped in the study populations according to the International Diabetes Federation (IDF) criteria. Our results showed that GRS was associated with the increased prevalence of MetS in females ($P = 1.28 \times 10^{-3}$) but not in males ($P = 0.714$). The area under the curve (AUC) in the receiver operating characteristics (ROC) analysis was higher in GRS prediction of MetS in females (AUC = 0.858) than that in males (AUC = 0.577). In total, this study identified new female-specific genetic variants associated with metabolic traits and suggested that GRS of variants showing compelling association with MetS-related traits is a likely useful predictor of MetS in females.

2185T

Genetic polymorphisms association with the neutrophil-lymphocyte ratio and their clinical implications for metabolic risk factors. *E. Choe¹, B. Park², H. Kang², S. Won².* 1) Seoul National University Hospital Healthcare System Gangnam Center, Seoul, South Korea; 2) Department of Public Health Science, Seoul National University, Seoul, Korea; 3) Department of Internal Medicine, Seoul National University Hospital Healthcare System Gangnam Center, Seoul, Korea.

Background: The neutrophil-lymphocyte ratio (NLR) is valuable as a prognostic or predictive biomarker in various diseases, but the genetic factors underlying the NLR have not been studied. We attempted to investigate the genetic polymorphisms related to the phenotypic NLR and to determine whether they have the power to predict metabolic syndrome or metabolic risks in addition to impacting the phenotypic NLR. **Methods:** A genome-wide association study (GWAS) was carried out for log-transformed NLR among 7,257 healthy Koreans using the Affymetrix Axiom™ KORV1.1-96 Array. Prediction models for metabolic risk status were designed using significant single nucleotide polymorphisms (SNPs). **Results:** We detected 4 SNPs near the TMEM116, NAA25 and PTPN11 genes associated with the NLR. The top SNP associated with the log-transformed NLR was rs76181728 in the TMEM116 gene (P value, discovery set=1.68×10⁻²⁰, validation set=0.00749). A case-control study was performed for the presence of metabolic risks according to the respective SNPs after adjusting for age, sex and body mass index (BMI). All four SNPs displayed significant odds ratios (ORs) with increased blood pressure and increased waist circumference. In the prediction model for metabolic syndrome, rs76181728 showed a significant association (OR=1.465, 95% confidence intervals (CI)=1.091-1.969, P=0.011) after adjustment for the phenotypic NLR. **Conclusions:** This was the first GWAS with which we identified four novel SNPs associated with the NLR in healthy Koreans. Thus, SNPs in the relevant genes at locus 12q24 might be used as biomarkers for diseases that have previously been predicted by the phenotypic NLR.

2186F

Reassessing the significance of the PAH c.158G>A (p.Arg53His) variant in Korean patients with hyperphenylalaninemia. *J. Kim¹, D. Lee², J. Lee².* 1) Department of Pediatrics, Chungbuk National University H, South Korea; 2) Department of Pediatrics, Soonchunhyang University H, South Korea.

Background: The accurate interpretation of sequence variation is critical for successful molecular diagnoses. It is also fundamental to the accurate diagnosis and treatment of phenylketonuria (PKU). This study aims to evaluate the significance of the c.158G>A (p.Arg53His) variant in the PAH gene, which was previously reported to be a pathogenic mutation that results in decreased phenylalanine hydroxylase enzyme activity in hyperphenylalaninemia (HPA) patients. **Methods:** Seven unrelated Korean patients with HPA genotyped with the c.158G>A variant were included in this study. The variant c.158G>A was classified by the standards and guidelines for the interpretation of sequence variants by the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. **Results:** By both directly collecting genetic data and comprehensively reviewing the existing literature, we found that this variant is more appropriately classified as "Likely benign" rather than pathogenic. The allele's frequency is 2.57% in the general Korean population, which was greater than expected for PKU. This variant was observed to be homozygous in healthy subjects and was also observed in *cis* with other pathogenic variants. It is common in East Asian populations (especially in Koreans) compared to Western populations. There is a possibility that it causes decreased enzyme activity without leading to the full pathology of PKU. **Conclusions:** This study expands our understanding of the consequences of variation in PAH and its relationship to HPA.

2187W

Whole-exome sequencing of Emirati population and high-risk genetic susceptibility to type 2 diabetes in Abu Dhabi. D. Simpson^{1,2}, A. Pandya^{1,2}, C. Inman³, M. Vogelsang², A. Abdulle³, R. Ferguson^{1,2}, D. Galani³, M. Haji³, A. Al Hamiz³, A. Al Hosani³, J. Ahn^{1,2}, R. Ramasamy⁴, A. Schmidt⁴, S. Sherman¹, R. Ali³, R. Hayes^{1,2}, H. Alsafar⁵, T. Kirchhoff^{1,2}, UAE Healthy Future Study. 1) Department of Population Health, NYU School of Medicine, New York, NY; 2) NYU Perlmutter Cancer Center, New York, NY; 3) Public Health Research Center, NYUAD, Abu Dhabi, UAE; 4) Diabetes Research Program, Division of Endocrinology, NYU School of Medicine, New York, NY; 5) Center for Biotechnology, Khalifa University, Abu Dhabi, UAE.

In the United Arab Emirates (UAE), the increasing incidence of type 2 diabetes (T2D) and other metabolic traits draws an interest to the role of population-specific germline genetic risk factors. While next-generation sequencing (NGS) initiatives for mapping the genetic risk of metabolic disease have been established in some countries in the Arabian Peninsula, no NGS efforts have yet been expended in UAE. In this study, we present a first whole-exome sequencing (WXS) analysis of 96 subjects from a recently initiated UAE cohort study to elucidate etiological and genetic factors associated with metabolic disease. We sought to generate preliminary data on genetic population structure and exome variation for their relationship to high-risk forms of T2D in the UAE population. We performed WXS on 96 UAE nationals residing in Abu Dhabi, 30 of whom had T2D. By analyzing the structure of the UAE population, we show that there is an admixture from proximal ancestries, in particular European, African and South Asian populations. We furthermore report a first catalogue of exonic variation in the UAE. While our data show no significant differences in the overall distribution of identified rare or common variants compared to other populations, we noted the differences in the distribution of novel indel variants, and an increase of pathogenic deleterious variants. Importantly, we found a significant enrichment of high-risk pathogenic mutations associated with Mendelian forms of T2D. Our study represents a first NGS effort towards the genetic mapping of metabolic disease in the UAE. Using WXS data from 96 UAE nationals in Abu Dhabi, we show the presence of admixture of regional ancestries in the Emirati population. While our study did not detect overall differences in exonic variation distribution compared to other global populations, some differences in structural variation suggest population-specific genetic features to be explored in future analyses. Importantly, enrichment of pathogenic variants associated with Mendelian traits strongly suggests that consanguinity shapes the Emirati genome and, pending future validation on a genome-wide level, may contribute to the genetic susceptibility of metabolic disease in UAE.

2188T

Studies of liver tissue identifies functional gene regulatory elements associated to gene expression, type 2 diabetes and other metabolic diseases. C. Wadelius¹, M. Cavalli¹, N. Baltzer², G. Pan¹, J.R. Bårceñas-Walls¹, C. Kumar^{3,4}, S. Skrtic^{5,6}, J. Komorowski^{2,7}. 1) Science for Life Laboratory, Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden; 2) Science for Life Laboratory, Department of Cell and Molecular Biology, Computational Biology and Bioinformatics, Uppsala University, Uppsala, Sweden; 3) Translational Sciences, Cardiovascular Metabolic and Renal Diseases, IMED Biotech Unit, AstraZeneca, Gothenburg, Sweden; 4) Karolinska Institutet/AstraZeneca Integrated CardioMetabolic Center (KI/AZ ICMC); 5) Translational Medicine Unit CVRM, Early Clinical Development, IMED Biotech Unit, AstraZeneca, Gothenburg, Sweden; 6) Institute of Medicine at Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden; 7) Institute of Computer Science, Polish Academy of Sciences, Warszawa, Poland.

Genome wide studies of diseases and traits have been effective in finding association to gene regions but not the functional SNP or the gene mediating the effect. A fraction of functional variants are coding but the majority are predicted to be located in gene regulatory elements and mediate differential binding of transcription factors (TF) and thereby expression from the alleles. Difference in gene regulatory signals can be detected genome wide using ChIP-seq of TFs or histone modifications by aligning reads to known polymorphisms in individual genomes. Gene regulatory mechanisms are often studied in cell lines derived from cancer cells, which have single nucleotide and copy number variants that drive the cancer and additional aberrations acquired during prolonged culturing. It is important to study gene regulation in native conditions so we obtained normal human liver tissue and sequenced the genome using 10x genomics technology to reliably call all polymorphic positions and found 4 588 678 SNPs out of which 2 331 202 were heterozygous and used further in the analysis. We then performed ChIP-seq for H3K4me3, H3K27ac and CTCF, aligned the reads to the two phased personal genomes, and used an established pipeline to detect sites with a genome-wide significant difference in signal between the alleles. We only considered regions predicted to bind TFs and after removing SNPs in black list regions we detected 2,329 allele specific (AS-)SNPs. From the GWAS catalogue we extracted diseases and traits that can be attributed to liver biology and took the top SNP and those in LD $r^2 > 0.8$ and intersected with the AS-SNPs. We found 26 AS-SNPs that are likely candidates to mediate the effect. One of them is rs655185 with genome-wide corrected binomial test p-value 3.1×10^{-4} is candidate to explain the signal for rs622217 which is associated with type 2 diabetes (T2D). The effect may be mediated by the nearby *LPA* which has previously been associated to familial hyperlipidemia and T2D. Environmental and gene regulatory signals acting on rs655185 may now be studied further to clarify the pathways. We also investigated top SNP and highly associated SNPs from GTEx eQTLs in the liver and found 314 AS-SNPs that are putative candidates to mediate the effect.

2189F

Polygenic susceptibility underlies trajectories of weight and severe obesity from birth to adulthood.

M. Chaffin¹, A.V. Khera^{1,2,3}, K.H. Wade⁴, S. Zahid¹, J. Brancale⁵, M.E. Haas¹, A. Bick¹, K.G. Aragam^{1,2,3}, G.D. Smith⁴, N.J. Timpson⁴, L.M. Kaplan⁵, S. Kathiresan^{1,2,3}. 1) Cardiovascular Disease Initiative of the Broad Institute of Harvard and MIT, Cambridge, MA, USA; 2) Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA; 3) Cardiology Division of the Department of Medicine, Massachusetts General Hospital, Boston, MA, USA; 4) MRC Integrative Epidemiology Unit (IEU), Bristol Population Health Science Institute, University of Bristol, United Kingdom; 5) Obesity, Metabolism, and Nutrition Institute, Massachusetts General Hospital, Boston, MA, USA.

Obesity is an epidemic attributed largely to a rapidly changing environment; yet, obesity is highly heritable, with much of that heritability due to common DNA sequence variants. Here, we use large studies and improved algorithms to define a new polygenic predictor for severe obesity that incorporates a genome-wide set of 2.1 million common DNA variants. Applying this genome-wide polygenic score to 288,016 middle-aged adults in the UK Biobank (mean age 57 years), we find that those in the top decile of polygenic score are 29 pounds heavier than those in the bottom decile. Remarkably, the polygenic score identified both subsets of the population at substantial risk of severe obesity (BMI>40), and those that enjoyed considerable protection. Among 2,177 non-obese young adults (mean age 30 years) in the Framingham Heart Study who were followed for 26 years with serial weight measurements, 40 of 218 (18%) of those in the top decile of the polygenic score distribution went on to develop severe obesity as compared with 0 of 218 (0%) of those in the bottom decile. In the ALSPAC longitudinal birth cohort, we noted minimal gradient in weight at time of birth across polygenic score deciles, but this gradient increased to 8 pounds by age 8 years and 27 pounds by age 18 years. A genome-wide polygenic score identifies subsets of the population at high risk or near complete protection from severe obesity. Weight trajectories across strata of polygenic risk diverged starting at a very young age suggesting that clinical prevention needs to start in early childhood.

2190W

Harnessing longitudinal data to derive a new genetic risk score for childhood obesity.

S.J.C. Craig^{1,2}, J. Lin³, A.M. Kenney², I.M. Paul^{2,4}, L.L. Birch⁵, J.S. Savage^{6,7}, M.E. Marini⁷, F. Chiaromonte^{2,3}, M. Reimherr^{2,3}, K.D. Makova^{1,2}. 1) Biology Department, Penn State University, University Park, PA; 2) Center for Medical Genomics, Penn State University, University Park, PA; 3) Department of Statistics, Penn State University, University Park, PA; 4) Department of Pediatrics, Penn State College of Medicine, Hershey, PA; 5) Department of Foods and Nutrition, University of Georgia, Athens, GA; 6) Department of Nutritional Sciences, Penn State University, University Park, PA; 7) Center for Childhood Obesity Research, University Park, PA.

The childhood obesity epidemic affects one in three children in the United States. Obesity has a complex etiology with 40 to 70% of the interindividual variability attributed to genetics. However, only 5% is explained by known genetic variants. Here we exploit the longitudinal structure of the growth data to estimate the relationship between genetics and infant weight gain patterns during the first three years after birth. We genotyped 225 first-born children from the Intervention Nurses Start Infants Growing on Healthy Trajectories (INSIGHT) study on the Affymetrix Precision Medicine Research Array. Weight and length/height were collected at five time points between birth and one year, and at two and three years, and used to build growth curves from these participants. Utilizing functional data analysis techniques we identified 24 SNPs associated with the growth curves. Rapid infant weight gain genetic risk score (GRS) was calculated as the weighted sum of these SNPs' effect sizes. This risk score is highly predictive in the first three years after birth, after which GRS defined by SNPs identified in GWAS with adult obesity becomes effective. Our results suggest that predisposition to rapid infant weight gain and to adult obesity might be determined, at least in part, by different genetic variants. This information, along with other 'omic' and environmental factors (e.g. microbiome, methylome, diet, etc.) can aid in identifying at-risk children who will benefit most from very early-life obesity intervention programs.

2191T

Association of adiponectin gene variants (+276G/T and -3971A/G) with obesity susceptibility in Punjabi population of North India. *H. Kaur, B. Badaruddoza, V. Bains.* Human Genetics, Guru Nanak Dev University, Amritsar, India.

Human obesity is a worldwide health epidemic which is widely accepted as an important risk factor for metabolic syndrome and other non-communicable diseases. The etiology of obesity is multifactorial and involves complex interactions among genetic and environmental factors. Visceral fat accumulation is associated with dysregulation in the secretion of adipocytokines that participate in the development of obesity and its related disorders. The *ADIPOQ* gene spans 16kb consisting of three exons and two introns on chromosome 3q27 and *ADIPOQ* variants have been proved to be a promising markers for understanding the genetic basis of obesity in different study populations. The present study was aimed to investigate the association *ADIPOQ* variants +276G/T and -3971A/G with obesity in North Indian Punjabi population. The present case control study included a total of 520 unrelated adult subjects comprising of 240 obese (BMI >25 kg/m²) cases and 280 non obese (BMI <23 kg/m²) healthy controls. Biochemical analysis including fasting glucose and lipids was done for all the participants. Anthropometric parameters were measured using standard protocol. The variants were screened by polymerase chain reaction based restriction fragment length polymorphism method and was confirmed using sanger sequencing. It was revealed from the analysis that minor allele frequencies for both the polymorphisms were significantly higher among the obese cases than controls. It was also evidenced that both the variants of *ADIPOQ*-3971GG and +276TT significantly increases the risk towards the development of obesity (OR= 2.23(1.17-4.26), p=0.01, OR= 1.69(1.01-2.82) p=0.04 respectively) in our study population. All anthropometric and biochemical parameters except triglycerides and very low-density lipoproteins cholesterol have shown significant association with both -3971A/G and +276G/T genotypes. The *ADIPOQ* polymorphisms were significantly associated to increased susceptibility to obesity under the recessive genetic model and were shown to provide two fold risk (OR=2.21(1.19-4.12), p=0.011 and OR=1.64(CI: 1.0-2.7), p=0.04 respectively for -3971A/G and +276G/T polymorphism) in developing obesity. This study reveals that the adiponectin gene variants -3971A/G and +276G/T contributes to the genetic risk towards the development of obesity and related traits, and clinically helpful to estimate the disease risk in North Indian Punjabi population.

2192F

Identification of pleiotropic waist-to-hip ratio SNPs using the NHGRI-EBI catalog of published genome-wide association studies. *Y. Kaur^{1,2}, D.X. Wang^{1,3}, H.Y. Liu^{1,3}, D. Meyre^{1,4}.* 1) Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Canada; 2) Faculty of Medicine, University of Toronto, Toronto, Canada; 3) Schulich School of Medicine and Dentistry, University of Western Ontario, London, Canada; 4) Department of Pathology and Molecular Medicine, McMaster University, Hamilton, Canada.

Visceral obesity is associated with a wide array of health conditions, some of which share a genetic overlap with obesity. In order to assess pleiotropic associations, we conducted a hypothesis-free cross-trait analysis for waist-to-hip ratio adjusted for body mass index (WHR_{adjBMI}) loci derived through genome-wide association studies (GWAS). Summary statistics from 12 previously published GWAS were used to capture all WHR_{adjBMI} SNPs, and proxy SNPs were identified. These SNPs were used to extract cross-trait associations between WHR_{adjBMI} SNPs and other traits using the NHGRI-EBI GWAS catalog. Pathway analysis was conducted for pleiotropic WHR_{adjBMI} SNPs. We found 159 WHR_{adjBMI} SNPs and 3,674 proxy SNPs. Cross-trait analysis identified 239 associations, of which 100 were for obesity traits (e.g. BMI). The remaining 139 associations were grouped into 12 phenotypic categories: lipids, red blood cell traits, white blood cell counts, inflammatory markers and autoimmune diseases, type 2 diabetes-related traits, adiponectin, cancers, blood pressure, height, neuropsychiatric disorders, electrocardiography changes, and urea measurement. The highest number of trait associations were found for triglycerides (n=10), HDL-C (high-density lipoprotein cholesterol; n=9), and reticulocyte counts (n=8). Pathway analysis for WHR_{adjBMI} pleiotropic SNPs found immune pathways (e.g. crosstalk between dendritic cells and natural-killer cells, antigen presentation pathway, dendritic cell maturation) as the top canonical pathways. Our results point to a possible genetic association between reticulocyte counts and WHR_{adjBMI}, which has not yet been reported. We also highlight the interplay between immune pathways and body fat distribution, and their relationships with other traits at a genetic level. Our original and hypothesis-free methodology is a powerful tool to decipher genetic correlation between complex human traits and diseases.

2193W

Identification of novel genes whose expression in adipose tissue is causally associated with obesity traits.

S. Konigorski¹, J. Janke¹, G. Patone², D. Drogan³, M.M. Bergmann⁴, J. Hierholzer⁵, N. Hübner^{6,7}, R. Kaaks⁸, H. Boeing¹, T. Pischon^{1,6,7}. 1) Molecular Epidemiology Research Group, Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany; 2) Genetics and Genomics of Cardiovascular Diseases Research Group, Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany; 3) AOK Research Institute (WIdO), AOK Bundesverband, Berlin, Germany; 4) Department of Epidemiology, German Institute of Human Nutrition Potsdam-Rehbrücke (DIfE), Nuthetal, Germany; 5) Department of Diagnostic and Interventional Radiology, Clinic Ernst-von-Bergmann, Potsdam, Germany; 6) DZHK (German Center for Cardiovascular Research) partner site Berlin, Germany; 7) Charité Universitätsmedizin Berlin, Germany; 8) Department of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany.

Obesity has continuously increasing prevalence between 10–30% world-wide and is a major risk factor for chronic disease. Adipose tissue (AT), whose amount is used to define obesity, is stored in different body compartments and acts as an endocrine organ, so understanding the association of adiposity with metabolic imbalance is of interest. In this study, our aim was to identify novel genes whose gene expression in subcutaneous AT (SAT) is causally related to the obesity traits of body fat mass and body fat distribution. The study sample consisted of 160 probands from the population-based European Prospective Investigation into Cancer and Nutrition (EPIC) Potsdam substudy. Gene expression in SAT of 30,917 genes was quantified by sequencing total mRNA from abdominal SAT biopsies. SAT mass and the ratio of SAT and total AT (TAT) were obtained from whole-body MRI scans as measures of fat mass and fat distribution. In the analysis, we first investigated the association of gene expression with SAT and SAT/TAT in joint copula models, and followed-up the results with a gene ontology (GO)-term enrichment analysis, and a Mendelian Randomization (MR) study to investigate the causal association of genetically-determined gene expression with SAT and SAT/TAT, using single nucleotide variants (called from the RNA-Seq data with 6.6X) in the respective gene as instruments. The analysis identified 607 obesity-associated autosomal (430 for SAT, 215 for SAT/TAT, of which 38 overlap) and 18 sex-chromosomal genes after multiple testing corrections. 531 of them encode a known protein and 518 are novel candidate genes for obesity. Enrichment analyses indicated that SAT-associated genes are characterized by their higher than expected involvement in cellular, regulatory and immune system processes, and SAT/TAT-associated genes by their involvement in cellular, metabolic, and regulatory processes. MR analyses suggested that the gene expression of 35 (out of the 430) genes was causally related to SAT and the gene expression of 28 (out of the 215) genes to SAT/TAT. The gene expression of these 35 and 28 genes explained 70% and 69% of the variance of SAT and of SAT/TAT, respectively. 29 and 23 of them are novel. In summary, we identified 52 novel genes, 29 for body fat mass and 23 for body fat distribution, whose up-/down-regulated gene expression might causally contribute to obesity, and might help to explain the role of obesity and metabolic imbalance in chronic disease in more detail.

2194T

Association between SNPs in *FTO* and *IRX3* and metabolic traits in population of Colombian Caribbean Coast.

G.J. Mora-Garcia¹, M.S. Ruiz-Diaz¹, D. Mena-Yé², D. Gomez-Camargo¹. 1) Doctorado en Medicina Tropical, Facultad de Medicina, Universidad de Cartagena, Cartagena de Indias, Colombia; 2) Facultad de Medicina, Universidad de Cartagena.

FTO gen has been widely associated with obesity, type 2 Diabetes Mellitus (DM2) and related metabolic disorders. Recently, single nucleotide polymorphisms (SNPs) in *FTO* have been linked with the *IRX3* gen. Some authors have revealed high linkage disequilibrium (LD) between *FTO* and *IRX3* SNPs. In Latin America *FTO* has been also associated with these conditions, however there is little evidence about *IRX3-FTO* relationship. Since this is an admixed population, LD between both genes could show different behavior. Therein, this study aimed to describe the possible *FTO-IRX3* LD, and to estimate the association between these genes, obesity and related disorders in a Latin-American population. A cross-sectional genetic association study was carried out (n=805). Body weight excess (BWE), obesity, abdominal obesity, hypertension, hyperglycemia, DM2 and dyslipidemia were evaluated. Four SNPs were genotyped: 3 at *FTO* locus (rs17817449, rs8050136, rs9939609) and 1 at *IRX3* locus (rs3751723). LD between these SNPs was estimated. Data were analyzed through generalized regression models, including genotype additive and haplotypic models. Associations were adjusted by genetic stratification assuming a three-hybrid substructure (k=3) using a Bayesian approach (Markov Chain Monte Carlo). SNP rs3751723 (*IRX3*) was not in LD with *FTO* SNPs ($D' \leq 0.03$; $r^2 \leq 0.03$). But it was found for *FTO* SNPs ($D' \geq 0.97$; $r^2 \geq 0.93$); therefore, only one of these three *FTO* polymorphisms (rs9939609) was employed in further single-gene [GM2] association analyses. TT genotype (rs9939609) was found to be associated with waist circumference ($p=0.04$; $adj-p=0.01$); Interestingly, no association between this *FTO* SNP and BWE was found ($adj-p=0.13$). *IRX3* SNP was found to be associated with BWE (OR=1.06, $adj-p=0.03$ [GM3]). Also, it was found one *FTO-IRX3* combination associated with BWE (G-A-A-T, rs17817449-rs8050136-rs9939609-rs3751723; OR=0.67, $p=0.04$). *FTO* was related with waist circumference and *IRX3* was associated to BWE. In spite that these genes were not in LD, findings of a haplotype involving *FTO-IRX3* suggest a gene-gene interaction associated with increased risk of obesity. Results from the current study partially agree previous findings of a long-range functional connection between these genes. Further studies on this phenomenon might contribute to a better understanding of *FTO-IRX3* role in obesity pathogenesis.

2195F

Candidate gene analyses in families with discordance for extreme obesity. *M.H. Preuss¹, C. Schurmann², W.D. Li³, R.A. Price⁴, D.R. Reed⁵, R.J.F. Loos^{1,6,7}.* 1) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Regeneron Genetic Center, Regeneron Pharmaceuticals, Inc., Tarrytown, NY; 3) Department of Genetics, College of Basic Medical Sciences, Tianjin Medical University, Tianjin, China; 4) University of Pennsylvania, Philadelphia, PA; 5) Monell Chemical Senses Center, Philadelphia, PA; 6) The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY; 7) Genetics of Obesity and Related Metabolic Traits Program, Icahn School of Medicine at Mount Sinai, New York, NY.

Obesity in the general population is a common multifactorial disease that results from an intricate interplay of genes and environment. On the other hand, extreme and early-onset forms of obesity are rare, typically caused by a single gene mutation, and environment has little impact. Recent genome-wide association studies have shown that genes that play a role in monogenic and syndromic obesity, also may play a role in common obesity (e.g. *MC4R*, *SH2B1*, *POMC*). Here, we examine the role of rare and low frequency coding variants in 119 candidate genes in 393 families ($n = 1,501$ individuals) discordant for extreme obesity. Families had at least one offspring with extreme obesity ($BMI \geq 40 \text{ kg/m}^2$), who had at least one sibling or parent with no history of overweight or obesity (lifetime $BMI < 25 \text{ kg/m}^2$). DNA samples were genotyped using the Illumina MEGA array and the 1000G reference panel was imputed using a pedigree-phased approach. Analyses focused on 119 genes known to play a "definite" or "likely" role in monogenic forms of extreme and early-onset obesity in humans and/or animal models. We performed single-variant (additive/recessive) and gene-based association analyses of rare ($MAF < 1\%$) and low-frequency ($MAF < 5\%$) missense variants. Case ($BMI \geq 40 \text{ kg/m}^2$) vs control ($BMI < 30 \text{ kg/m}^2$) or extreme control ($BMI < 25 \text{ kg/m}^2$) status were used as outcomes. All analyses were adjusted for age and sex. Bonferroni corrected P -values at $\alpha = 0.05$ correspond to 10^{-4} for single-variant analyses and 4×10^{-4} for gene-based analyses. None of the single-variant or gene-based results reached Bonferroni-corrected significance, but for 15 variants ($P < 0.01$) and 8 genes ($P < 0.05$) associations were promising. For example, missense variants in *ALMS1* (rs45608038, p.Asn1786Asp, $P = 0.002$), *ATXN2* (rs7969300, p.Ser248Asn, $P = 0.0015$), *BBS1* (rs77298332, p.Ser410Phe, $P = 0.004$) and *BDNF* (rs66866077, p.Glu6Lys, $P = 0.003$) were found to be associated with increased risk of extreme obesity. Gene-based analyses highlighted *SH2B1* ($P_{\text{burden}} = 0.004$), *GLUCY2C* ($P_{\text{burden}} = 0.01$), *TLR5* ($P_{\text{burden}} = 0.01$), among others. Despite a family design with discordance in extreme obesity, our study had insufficient statistical power to identify significant associations with rare and low-frequency variation in candidate genes. Nevertheless, our analyses have prioritized a number of genes and variants for further follow up analyses to elucidate the mechanisms through which they affect human obesity.

2196W

Genome-wide association study of body fat distribution identifies novel adiposity loci and sex-specific genetic effects. *M. Rask-Andersen, T. Karlsson, W.E. Ek, A. Johansson.* IGP, Uppsala University, Uppsala, Uppsala, Sweden.

Body mass and body fat composition are of clinical interest due to their links with cardiovascular- and metabolic diseases. Fat stored in the trunk has been suggested as more pathogenic compared to fat stored in other compartments of the body. In this study, we performed genome-wide association studies (GWAS) for the distribution of body fat to the arms, legs and trunk. Proportions of fat, distributed to these compartments were estimated from segmental bio-electrical impedance analysis (sBIA) for 362,499 individuals from the UK Biobank. A total of 97 loci, were identified to be associated with body fat distribution, 40 of which have not previously been associated with an anthropometric trait. A high degree of sex-heterogeneity was observed and associations were primarily observed in females, particularly for distribution of fat to the legs and trunk. Interaction with sex was tested for with GWAMA and showed sex-heterogeneous effects at 42 loci. Enrichment analysis with DEPICT indicated involvement of mesenchyme derived tissues and cell types: the musculoskeletal system, chondrocytes, mesenchymal stem cells and adipocytes; female gonadal and endocrine tissues, as well as arteries. Our findings also highlight several enzymatically active members of the ADAMTS family of metalloproteinases, which are involved in extracellular matrix maintenance and remodeling. In conclusion, GWAS of sBIA-determined body fat distribution revealed a genetic architecture that influences the distribution of body fat throughout the human body. Analyses showed genetic variation to have a stronger influence on the proportion of fat stored in the trunk and legs in females, as compared to males. The capacity for peripheral adipose storage has been highlighted as a potential factor underlying the lower risk of cardiovascular and metabolic disease in women of middle age. Determining the genetic determinants and mechanisms that lead to a favorable distribution of body fat may help in risk assessment and in identifying novel venues for intervention to prevent or treat obesity-related disease.

2197T

Association of longitudinal body mass index (BMI) trajectories with genetic risk score in children with severe early childhood obesity (SECO).

V.V. Thaker¹, K. Qian², M. McDonald³, S. Christensen³, A. Lage³, S. Challa³, B. Weaver⁴, S. Vedantam³, S. Lopez-Pintado², J. Hirschhorn^{3,4,5}. 1) Columbia University Medical Center, New York, NY; 2) Mailman School of Public Health, New York, NY; 3) Boston Children's Hospital, Boston, MA; 4) Harvard Medical School, Boston, MA; 5) The Broad Institute, Boston, MA; 6) Boston University Medical School, Boston, MA.

Background: SECO, BMI > 120% of 95th percentile for children under 6 years of age, is a serious threat to our future health. While childhood BMI has been shown to track into adulthood, limited data are available on the association of childhood BMI trajectory and genetic risk. The objectives of the study were: 1) To identify the differences in the BMI trajectory of children with SECO compared to those with normal BMI. 2) To study the association between the functional BMI trajectories with polygenic risk score for obesity in children with severe obesity. **Methods:** Lifetime height and weight data of children attending outpatient clinics at Boston Children's Hospital were obtained from 2012-2017 for Genetics of Early Childhood Obesity (GECO) study by a validated electronic algorithm from the electronic health records (PMID: 27452794) and from growth charts using Plot Digitizer v3.12. A functional data analysis approach was used to model the BMI data measured at sparse time points using functional principal component decompositions and a mixed model approach. A functional depth-based global envelope test was used to compare the differences in the BMI trajectory patterns. We studied the association between the BMI smoothed functions and the additive polygenic risk score generated with BMI loci identified in population-based studies, in a subset of children. **Results:** A total of 332,250 observations in 28,554 children of mixed ethnicity (31% African Americans, 22% Hispanic); 52% girls, 22% with severe obesity, were obtained, with min. 2 BMI readings/child. Of the children with BMI available at age < 6 years (n=16,683), 7.6% had severe obesity. The global functional envelope test identifies differences in the BMI trajectory of children with SECO compared to those with normal BMI. No adiposity rebound is seen in the children with SECO, while their trajectories cross the percentiles predicted by the extension of the CDC 2000 BMI charts, stratified by gender. The polygenic risk score is associated with the BMI trajectory patterns. **Conclusions:** A novel functional data analysis approach can be used to analyze sparse longitudinal clinical anthropometric data. These methods can test and visualize the differences in BMI trajectories between different groups of children with SECO. The association of polygenic risk score with BMI trajectories of children with obesity may help in profiling those with higher genetic risk for severe obesity with a goal of personalized therapy.

2198F

Obesity and personality are genetically intertwined. U. Vainik^{1,2}, R.

Mõttus^{3,2}, A. Realo^{4,2}, J. Allik^{2,5}, T. Esko^{6,7}, A. Dagher¹. 1) Montreal Neurological Institute, Montreal, Quebec, Canada; 2) Institute of Psychology, University of Tartu, Tartu, Estonia; 3) University of Edinburgh, Edinburgh, UK; 4) University of Warwick, Warwick, UK; 5) Estonian Academy of Sciences, Estonia; 6) Estonian Genome Centre, University of Tartu; 7) Broad Institute, Cambridge, Massachusetts.

Obesity-related genes mostly express in the brain and obesity is related to personality. However, it is unclear how genetic propensity for obesity translates into observable behaviour. Here, we assessed overlap between genetic risk for obesity and personality traits. We relied on genomic and NEO-PI-3 personality data (240 items) in 3445 Estonian adults with a wide age range (18-91). Obesity polygenic score (OPGS) constructed from GWAS weights obtained from GeneAtlas (UK Biobank, 452,264 individuals) comprising 4,759,748 single nucleotide polymorphisms (SNPs) predicted body mass index in this sample ($R^2=0.034$, $p<0.001$). When correlating OPGS with individual personality items, the analysis highlighted items known to be related to obesity from previous meta analysis. Namely, the personality item-OPGS correlation profile tracked with the meta-analytic personality item-BMI profile ($r=0.35$; $p<0.001$). On an individual item level, OPGS associated positively with behaviours, such as overeating, sociability, and joyfulness, and negatively with deliberation, straightforwardness, and compliance. Together, current analysis shows a systematic overlap between personality and genetic propensity for obesity, which may inform intervention efforts.

2199W

Genetic architecture of human obesity traits in the rhesus macaque. A. Vinson^{1,2}, M.J. Raboin¹, J. Letaw¹, A.D. Mitchell¹, D. Toffey¹, C. Roberts², J.E. Curran³. 1) Oregon National Primate Research Center, Primate Genetics Section, Oregon Health & Science Univ., Beaverton, OR; 2) Oregon National Primate Research Center, Div. of Cardiometabolic Health, Oregon Health & Science Univ., Beaverton, OR; 3) South Texas Diabetes & Obesity Institute, Dept. of Human Genetics, Univ. of Texas Rio Grande Valley School of Medicine, Brownsville, TX.

Rhesus macaques display considerable spontaneous variation in adiposity and susceptibility to dietary challenge. While the physiological consequences of obesity have been studied extensively in this species, corollary genetic studies of obesity do not exist. We aimed to assess genetic contributions to spontaneous adiposity in rhesus macaques. We measured phenotypic variation by age-class and sex for 4 well-studied human obesity traits in 584 chow-fed macaques, including body mass index (BMI), waist-to-height ratio (WHtR), waist-to-thigh ratio (WTR), and circulating ghrelin. We estimated total and sex-specific heritability for all traits, including waist-to-thigh ratio adjusted for BMI (WTRadjBMI), as well as genotypic and phenotypic correlation among all traits. We also assessed functional genetic variation segregating at *ADIPOQ*, *BDNF*, *FTO*, *LEP*, *LEPR*, *MC4R*, *POMC*, and *RETN* in 4 animals with extreme values of BMI, WHtR, or WTR. All measures of adiposity increased with age, and were greater for males than females. Trait heritabilities ranged from 0.13 to 0.43 ($P=2.6 \times 10^{-2}$ to 4.0×10^{-7}). When stratified by sex, heritability was greater in females for all traits, with the exception of BMI, for which heritability was greater in males. Sex-stratified heritability estimates for BMI, WHtR, WTR, and ghrelin included confidence intervals that did not overlap between males and females. Phenotypic variation in all traits was robustly correlated, with genetic correlations ranging from 0.63-0.89, indicating substantial pleiotropy. Macaques with extreme adiposity segregated 825 functional genetic variants at 7 of 8 obesity genes, including 6 missense mutations in *ADIPOQ*, *BDNF*, *FTO*, *LEP*, and *LEPR*, and notably, 1 premature stop mutation in *LEPR*. We conclude that there are significant polygenic contributions to adiposity in macaques, and mutations with potentially larger effects in genes known to influence obesity in humans. We find substantial similarity between macaques and humans for traits that are robust indicators of adiposity in humans.

2200T

PheWAS of sub-components of T2D genetic risk in large-scale biobanks reveals clinical outcomes linking to physiological effects. X. Zhong¹, A. Mahajan², J. Torres², A.B. Faucon¹, R. Tao¹, N.J. Cox¹, M.I. McCarthy². 1) Vanderbilt Genetics Institute, Nashville, TN, USA; 2) The Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK.

Development of type 2 diabetes (T2D) involves dysfunction of multiple mechanisms. Our primary goal was to comprehensively interrogate clinical phenotypes associated with T2D risk alleles considered to act through different biology. We constructed three genetic risk scores (GRS) according to the physiological classification of T2D risk variants by Mahajan *et al.* (2018 Nat Genet) that capture the primary effect of insulin secretion (IS), insulin action (IA), and BMI and dyslipidemia (BL). We tested these instruments for association across >1600 clinical endpoints in two independent biobanks, with a total of 362,485 samples of European ancestry (23,294 from BioVU and 339,191 from UK Biobank). Despite different ascertainment of subjects in BioVU and the UKBB, the GRS-phenotypes associations were directionally consistent (concordance rate ranged 60%-100% over a range of significance thresholds). We then meta-analyzed the phenome-wide results over 1099 phenotypes shared by the two sites, and identified 42 significant associations (Bonferroni correction, $p < 4.4 \times 10^{-5}$) that involved 33 phenotypes in addition to diabetes. For most of these associated phenotypes (88%), the effect size estimates significantly differed across the GRS categories (Cochran's Q, $p < 0.05$), including hypercholesterolemia, coronary artery disease, obesity, chronic nonalcoholic liver disease, renal failure, and osteoarthritis etc. We further carried out association analysis stratified by T2D status, and observed strong associations only in T2D patients but not in patients without diabetes for a variety of phenotypes (e.g., Renal failure ($p=6.2 \times 10^{-4}$) and Sleep apnea ($p=0.0006$) in BL, Intestinal obstruction ($p=0.001$) in IA, and Cerebral artery occlusion, with cerebral infarction ($p=0.004$) and Speech disturbance ($p=0.006$) in IS). In conclusion, the GRSs comprised of SNPs associated with different mechanisms of T2D genetic risk were associated at different degrees with clinical outcomes, consistent with the hypothesis that different pathophysiology may contribute to the differentiated T2D comorbidities/complications. Combining results from independent biobanks increased our ability to resolve the patterns of phenome associations with GRS for sub-components of the T2D genetic risk physiology.

2201F

Genome-wide association study identifies novel risk locus for erectile dysfunction and implicates hypothalamic neurobiology and diabetes in etiology. J. Bovijn^{1,2}, L. Jackson³, J. Censin^{1,2}, C.Y. Chen^{4,5}, T. Laisk-Podar^{6,7}, S. Laber^{1,2}, T. Ferreira¹, C.A. Glastonbury¹, J. Smoller⁸, J.W. Harrison⁹, K.S. Ruth⁹, R.N. Beaumont⁶, S.E. Jones⁸, J. Tyrrell⁶, A.R. Wood⁶, M.N. Weedon⁸, R. Mägi⁶, B. Neale^{4,5}, C.M. Lindgren^{1,2,9}, A. Murray⁶, M.V. Holmes^{10,11,12}. 1) Big Data Institute at the Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, OX3 7LF, UK; 2) Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, OX3 7BN, UK; 3) Institute of Biomedical and Clinical Science, University of Exeter Medical School, University of Exeter, Exeter, EX2 5DW, UK; 4) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts 02114 USA; 5) Psychiatric & Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, 02114, USA; 6) Estonian Genome Center, Institute of genomics, University of Tartu, Tartu 51010, Estonia; 7) Department of Obstetrics and Gynecology, Institute of Clinical Medicine, University of Tartu, Tartu, Estonia; 8) Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Exeter, EX2 5DW, UK; 9) Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, 02142, USA; 10) National Institute for Health Research Oxford Biomedical Research Centre, Oxford University Hospital, Old Road, Oxford OX3 7LE, UK; 11) Clinical Trial Service Unit & Epidemiological Studies Unit (CTSU), Nuffield Department of Population Health, Big Data Institute Building, Roosevelt Drive, University of Oxford, Oxford, OX3 7LF, UK; 12) Medical Research Council Population Health Research Unit at the University of Oxford, Nuffield Department of Population Health, University of Oxford, UK.

Erectile dysfunction (ED) is a common disease, affecting 20–40% of men over 60. The genetic architecture of ED remains poorly understood, owing in part to a paucity of well-powered genetic association studies. We conducted a genome-wide association study (GWAS) for ED using data from 6,175 cases among 223,805 European men from UK Biobank (UKBB), Estonian Genome Center of the University of Tartu (EGCUT) and Partners HealthCare Biobank (PHB). GWAS in UKBB revealed a single genome-wide significant ($p < 5 \times 10^{-8}$) locus at 6q16.3 in the intergenic region between *MCHR2* and *SIM1* (lead variant rs57989773, EAF_{UKBB} (C-allele) = 0.24; OR 1.23; $p = 3.0 \times 10^{-11}$). Meta-analysis with estimates from PHB (OR 1.20; $p = 9.84 \times 10^{-6}$) and EGCUT (OR 1.08; $p = 0.16$) yielded a pooled meta-analysis OR 1.20; $p = 5.72 \times 10^{-14}$. Meta-analysis of all variants yielded no further genome-wide loci. In-silico analysis of the locus revealed enhancer marks in several tissues, including embryonic stem cells, mesenchymal stem cells and endothelial cells, indicating that the ED-associated interval lies within a regulatory locus. Chromosome conformation capture (Hi-C) and Capture Hi-C showed that *MCHR2* and *SIM1* were in the same topologically associated domain as the ED-associated variants, with high contact probabilities between the ED-associated interval and the *SIM1* promoter. A phenome-wide association study of rs57989773 found significant associations with 12 phenotypes including adiposity (9 traits), adult height and sleep-related traits. Sex-stratified analyses revealed sexual dimorphism for waist-hip ratio (WHR), systolic and diastolic blood pressure. These associations recapitulate the effects on blood pressure and adiposity seen in patients with rare *SIM1* coding variants. These findings suggest that the ED-associated intergenic region and *SIM1* are functionally connected. *SIM1* encodes a transcription factor that is highly expressed in hypothalamic neurons and *Sim1*-expressing neurons have recently been implicated in murine male sexual function. LD score regression identified ED to be correlated and share genetic architecture with type 2 diabetes (T2D; $rg = 0.40$, nominal p -value = 0.0008; FDR-adjusted p -value = 0.0768). Mendelian randomization identified genetic risk to T2D to be causally implicated in ED: OR 1.11 (95% CI 1.05–1.17, $p = 3.5 \times 10^{-4}$), per 1-log higher genetic risk of T2D, with no causal effects of BMI on ED. Our findings provide novel insights into the biological underpinnings of ED.

2202W

Single nuclei RNA-seq to decipher hypothalamic transcriptional response to chronic topiramate administration. K. Ushakov^{1,2}, M.F. Lopez^{1,3}, A. Nelson⁴, G. Gorski^{1,2}, E.Z. Macosko^{4,5}, J.N. Hirschhorn^{1,2,3}. 1) Center for Basic and Translational Obesity Research and Divisions of Endocrinology and Genetics, Boston Children's Hospital, Boston, MA, USA; 2) Metabolism Initiative and Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA; 3) Department of Medicine, Endocrine Division, Harvard Medical School, Boston, MA, USA; 4) Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA, USA; 5) Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA.

Obesity remains a major global public health problem caused by a combination of environmental and genetic factors. Some progress has been made in our understanding of appetite and body weight regulation, but there are still no highly effective and safe therapies to treat obesity. One recently approved anti-obesity treatment includes topiramate, which can lead to decreased appetite by unknown mechanisms. Genome wide association studies (GWAS) have previously identified numerous common genetic variants that are associated with body mass index (BMI). Computational methods developed by our laboratory and others have strongly implicated these variants with neuronal biology, and specifically glutamate signaling and synaptic plasticity. These findings overlap with one of topiramate's proposed mechanism of action, which is modulation of glutamate signaling. Thus, BMI GWAS loci may contain genes encoding site(s) of action for topiramate's anorectic effect. In the present study, we decided to take an unbiased gene expression approach to dissect the mechanism by which topiramate elicits its effects and also to determine whether there is overlap between genes regulated by topiramate and those implicated by GWAS. Obese mice were injected intraperitoneally with either topiramate (50 mg/kg/day) or vehicle for 14 days. Brain tissue from relevant hypothalamic regions (arcuate and paraventricular nuclei) was collected and single nuclei were isolated. Using a combination of flow cytometry and the 10X Genomics Chromium Platform, we generated gene expression signatures of >8,000 hypothalamic nuclei collected from mice treated with vehicle or topiramate. The data was aligned and clustered to associate nuclei with different cell types. Results show differential expression of several non-coding RNAs such as *Malat1*, *Meg3* and *Nrg3os* in multiple clusters between topiramate or vehicle treatments. Our data also identifies expression differences in genes underlying obesity or related to proposed mechanisms of topiramate action, including *Ksr2*, *Kcnd3*, and *Gria2*. We are performing gene set analysis to find pathways enriched in differentially expressed genes within cell types. Our results begin to describe a hypothalamic transcriptional signature for the anti-obesity drug topiramate. These studies could lead to a better understanding of processes that regulate body weight and help guide development of more precise and effective therapies for weight loss.

2203T

High-resolution characterization of cellular imbalance and dysregulation in obesity-induced type 2 diabetes. J. Vijay^{1,2}, M-F. Gauthier³, ED. Aguilar⁴, M. Pelletier⁵, WA. Cheung⁴, J. Johnston⁴, H. Djambazian², A. Staffa², A. Pramatarova², M-C. Vohl⁵, G. Bourque^{1,2}, A. Tchernof³, E. Grundberg^{1,2,4}. 1) Department of Human Genetics, McGill University, Montreal, QC, Canada; 2) McGill University and Genome Quebec Innovation Centre, Montreal, Québec, Canada; 3) Québec Heart and Lung Institute, Université Laval, Québec, QC, Canada; 4) Children's Mercy Kansas City, MO, US; 5) Institute of Nutrition and Functional Foods (INAF), Université Laval, Québec, Québec, Canada.

Adipose tissue (AT) is a complex endocrine organ found throughout the body at specific locations including under the skin (subcutaneous, SAT) or around internal organs (visceral, VAT). Visceral adiposity is an important determinant of obesity-linked metabolic complications, but the precise mechanism of association remains unknown. Dysfunctional adipocytes and infiltration of immune cells are proposed mechanisms but the complete cellular landscape remains to be elucidated. To this end, we aimed to characterize the non-adipocyte fraction of AT (i.e. stromal vascular fraction - SVF) derived from both SAT and intra-abdominal VAT from obese (BMI > 35 kg/m²) individuals with or without type 2 diabetes (T2D). In total, we obtained 26 tissue samples (T2D = 11) and generated mRNA profiles at single-cell resolution applying the 10x Genomics Chromium platform. On average we sequenced 1,527 cells/sample at a depth that resulted in on average 1,175 genes/cell. We used RNA-Seq from matching bulk samples for validation purpose. After initial QC and filtering we identified 26,350 high quality cells for clustering and further analysis. The majority of the cells (~60%) were stem/progenitor cells, found clustered as six different subpopulations. Among these, three clusters constituted of VAT cells only, with two clusters showing enrichment for the mesothelial marker *MSLN* whereas the other had pronounced expression of the pro-adipogenic factor *CFD* similar to progenitor clusters from SAT. We further noted sub-populations among *MSLN*-expressing cells being derived mainly from T2D patients. We were also able to profile the large number of immune cell subtypes residing in AT in a depot and T2D-specific manner. For instance, we confirmed the notion that AT macrophages are residing to larger extent in VAT compared to SAT including cells at different stages of polarization – a pattern that was further enhanced when analysis was restricted to cells derived from T2D subjects. Finally, we identified the presence of multiple endothelial cell types including AT-microvascular endothelial cells expressing *FABP4* and *CD36* which were recently shown to be involved in obesity-linked AT dysfunction. In conclusion, we have shown that high-throughput single-cell approaches applied to deeply phenotyped cohorts and carefully prepared tissues samples allow the identification of dysfunctional cellular and molecular machineries underpinning complex diseases such as obesity and type 2 diabetes.

2204F

Identification of death associated protein kinase 2 as a novel obesity-induced driver of non-alcoholic fatty liver disease. A. Ko^{1,2}, J.N. Benhammour^{1,3}, K. Garske¹, D. Kaminska^{1,4,5}, E. Nikkola¹, J.R. Pisegna^{1,3}, S.T. Hui⁶, A.J. Lusis^{1,6}, V. Männistö⁷, V. Kärjä⁸, R.M. Cantor¹, J.S. Sinheimer^{1,8}, K.L. Mohlke⁹, J. Pihlajamäki^{4,10}, P. Pajukanta^{1,2,11}. 1) Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, CA; 2) Molecular Biology Institute at UCLA, Los Angeles, CA; 3) Division of Gastroenterology, Hepatology and Parenteral Nutrition, Department of Medicine, VA Greater Los Angeles Healthcare System, Los Angeles, CA; 4) Institute of Public Health and Clinical Nutrition University of Eastern Finland, Kuopio, Finland; 5) Turku PET Centre, Turku University Hospital, Turku, Finland; 6) Department of Medicine/Division of Cardiology, David Geffen School of Medicine at UCLA, Los Angeles, CA; 7) Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland; 8) Department of Biomathematics, David Geffen School of Medicine at UCLA, Los Angeles, CA; 9) Department of Genetics, University of North Carolina, Chapel Hill, NC; 10) Clinical Nutrition and Obesity Center, Kuopio University Hospital, Kuopio, Finland; 11) Bioinformatics Interdepartmental Program, UCLA, Los Angeles, CA.

The rising prevalence of obesity has led to a pandemic of non-alcoholic fatty liver disease (NAFLD), affecting almost 25% of adults worldwide. Without intervention or treatment, NAFLD can progress into nonalcoholic steatohepatitis (NASH), cirrhosis or hepatocellular carcinoma, which are often irreversible. To improve our understanding and treatment of NAFLD, we sought to delineate the molecular mechanisms of NAFLD by testing the hypothesis that obesity impairs the function of adipose tissue, which can lead to ectopic fat deposition in the liver. We RNA-sequenced subcutaneous adipose tissue samples collected from 259 morbidly obese individuals during their bariatric surgery and re-profiled their adipose transcriptomes one-year post surgery to understand the effect of weight loss on gene expression. In parallel, liver histology was performed to accurately stage NAFLD and NASH. We discovered that the expression of 17 genes is significantly downregulated in NASH (adjusted $P < 0.05$) and negatively correlated with liver steatosis score and serum triglycerides (TGs). Furthermore, adipose expression of five of the 17 genes was also correlated with diet-induced liver steatosis in a large panel of mouse strains. Notably, the adipose expression of one of the five genes, *DAPK2*, recovered after the weight loss at the one-year follow-up. To infer a causal pathway among obesity, serum TGs, *DAPK2* adipose expression, and NAFLD, we leveraged longitudinal transcriptome information and carried out mediation analyses. The results of our mediation analyses support the causal effect of obesity on NAFLD, mediated by adipose expression of *DAPK2* and serum TGs. To gain insight into the functional consequences of reduced *DAPK2* expression in the adipose tissue, we knocked down *DAPK2* in human primary preadipocytes via siRNA. Five key genes involved in autophagy were significantly downregulated, including *TGM2* and *ULK2*, which also regulate adipocyte differentiation. These data suggest impaired autophagy and adipocyte differentiation due to *DAPK2* reduction in the adipose tissue as a novel pathogenic mechanism of NAFLD. In summary, we present genomic and experimental data to support the causal role of *DAPK2* in NAFLD pathogenesis.

2205W

Association between hepatic CYP7A1 activity and PGC-1 α polymorphism.

T. Inamine¹, A. Makimoto¹, K. Ohyama², K. Tsukamoto¹. 1) Department of Pharmacotherapeutics, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan; 2) Department of Pharmacy Practice, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan.

Background Hepatic cholesterol 7 α -hydroxylase (CYP7A1) is the rate-limiting enzyme in the classical bile acid biosynthesis pathway and transcriptionally induced by HNF4 α and PGC-1 α . We evaluated the effects of CYP7A1, HNF4 α , and PGC-1 α polymorphisms on CYP7A1 enzyme activity in healthy Japanese individuals. **Methods** A study cohort consisted of 42 healthy Japanese students (31 female and 11 male) in School of Pharmacy, Nagasaki University. Blood was drawn from 1 to 3 PM. Serum 7 α -hydroxycholesterol-4-en-3-one (C4), which is a serum marker of hepatic CYP7A1 activity, and total cholesterol were measured by high performance liquid chromatography-ultraviolet detection and enzymatic colorimetric method, respectively. Four single nucleotide polymorphisms of the 3 genes (rs3808607 within CYP7A1, rs6017340 and rs6031587 within HNF4A, and rs8192678 within PPARGC1A) were genotyped by polymerase chain reaction (PCR)-restriction fragment length polymorphism and PCR-high resolution melting curve analysis. Serum C4 concentration as well as C4/cholesterol ratio were compared among three genotypes of each SNP by Jonckheere-Terpstra trend test. **Results** In our healthy Japanese cohort, serum C4 concentration (35.53 vs 34.92 μ g/L) and total cholesterol concentration (13.71 vs 14.25 mmol/L) and C4/cholesterol ratio (7.10 vs 6.61 μ g/mmol) were comparable between female and male subjects. In regards to rs3808607, rs6017340, and rs6031587, there were no significant trend in C4, total cholesterol, and C4/cholesterol ratio among three genotypes of each SNP. With respect to rs8192678 of PPARGC1A, C4 level (GG 28.31 vs GA 34.36 vs AA 43.03 μ g/L) as well as C4/cholesterol ratio (GG 12.25 vs GA 12.32 vs AA 16.99 μ g/mmol) were increased in A-allele dependent manner ($P = 0.007$ and $P = 0.015$, respectively). **Conclusion** This study demonstrated association between hepatic CYP7A1 activity and a PPARGC1A polymorphism for the first time. rs8192678 that is a missense variant of PGC-1 α (p.Gly482Ser) might influence on CYP7A1 activity more than rs3808607 which is a functional promoter SNP of CYP7A1, at least in a healthy Japanese.

2206T

Common variants in NPC1L1 are not associated with response to ezetimibe treatment in the IMPROVE-IT trial.

Y. Zhang¹, G. Kosova², C.T. Ruff³, Z. Guo¹, D.F. Reilly², D.V. Mehrotra¹, Y. Mitchell¹, S.H. Shah⁵, M.A. Blazing⁵, E.A. Bohula³, R.P. Guigliano³, C.S. Fox², M.S. Sabatine³, P.M. Shaw². 1) Biostatistics and Research Decision Sciences, Merck Sharpe and Dohme, Kenilworth, NJ; 2) Genetics and Pharmacogenomics, Merck Sharpe and Dohme, Kenilworth, NJ; 3) TIMI Study Group, Brigham and Women's Hospital, Boston, MA; 4) Clinical Research, Merck Sharpe and Dohme, Kenilworth, NJ; 5) Duke Clinical Research Institute, Durham, NC.

IMProved Reduction of Outcomes: Vytorin Efficacy International Trial (IMPROVE-IT) demonstrated a significant benefit of adding ezetimibe to a statin therapy in lowering LDLc levels and reducing the risk of cardiovascular (CV) events. Ezetimibe lowers absorption of dietary cholesterol by inhibiting Niemann-Pick C1-like 1 (NPC1L1), a cholesterol transporter found in the intestinal epithelial cells. Previous studies have shown genetic variation in NPC1L1 to be associated with NPC1L1 levels, LDLc levels and risk of CV events. Therefore, we hypothesized that variants in NPC1L1 are likely candidates to explain a portion of inter-individual variability observed in LDLc reduction in response to ezetimibe treatment. We examined the associations between five common, independent NPC1L1 variants previously shown to be associated with LDLc levels in Global Lipid Genetics Consortium (GLGC) analyses and LDLc levels in the ~6,500 participants of the IMPROVE-IT trial, who were randomized to receive simvastatin monotherapy or simvastatin+ezetimibe combination treatment. There were no significant associations between baseline LDLc levels or percent change in LDLc after 1 month of either treatment and any of the SNPs, tested either as a single variant or as a genetic risk score ($P > 0.1$). Association between rs10234070 (MAF=0.11) and risk of CV events in patients treated with simvastatin+ezetimibe remained significant after correcting for multiple testing ($P_{corrected}=0.015$). However, the observed direction of effect is opposite to what we predicted, such that the minor ("T") allele associated with increased LDLc levels in GLGC is associated with reduced risk of CV events in our study. While we cannot rule out a potential role for rare, inactivating mutations in NPC1L1 in response to ezetimibe treatment, we did not find significant associations between common variants and baseline LDLc, LDLc reduction or CV events in the IMPROVE-IT trial in a directionally consistent manner. Therefore, our data suggests that the reduction in LDLc levels by adding ezetimibe to a statin regimen remains consistent regardless of the common variation in NPC1L1.

2207F

Genetic susceptibility to low HDL-C among Filipinos. *E. Cutiongco de la Paz*^{1,2}, *R. Sy*², *J. Nevado Jr.*¹, *P. Reganit*², *F. Punzalan*³, *L. Santos*³, *R. Tiongco II*², *E. Llanes*³, *D. Ona*³, *J. Magno*³, *J. Aherrera*³, *L. Abrahan IV*³, *C. Agustin*³, *A. Aman*², *A. Bejarin*². 1) Institute of Human Genetics, National Institutes of Health, University of the Philippines Manila, Philippines; 2) Philippine Genome Center, University of the Philippines; 3) Section of Cardiology, Department of Medicine, College of Medicine, University of the Philippines Manila.

High density lipoprotein cholesterol (HDL-C) is strongly correlated with atherosclerotic cardiovascular disease (ASCVD) risk. Differences in HDL-C levels are seen across ethnic groups with Filipinos having high prevalence of low HDL-C from national nutrition surveys in the Philippines over a ten year period (2003 to 2013). Studies on HDL-C metabolism and heritability estimates suggest that 60-70% of variance in HDL may be due to genetic factors. HDL-C is influenced by genetic variants across entire spectrum of allele frequencies in genes involved in its metabolism. The study aims to validate the important role of common variations at different loci influencing HDL-C levels in the Filipino adult population. This is an observational study with 62 participants having low HDL-C profile and 123 have normal HDL-C profiles. Patients' DNA samples were genotyped using a customized Illumina GoldenGate microarray chips. Polymorphisms in CETP, ABCG8, APOB, GALNT2 and GCKR genes have been found to be associated with low HDL-C profile among Filipinos. This study has provided an opportunity to reveal population-specific genetic determinants of HDL-C.

2208W

GWAS of serum lipids in Africans identifies 4 novel loci. *A.R. Bentley, G. Chen, A.P. Doumatey, A.A. Adeyemo, C.N. Rotimi, the AADM Investigators.* National Human Genome Research Institute, Bethesda, MD.

Serum lipids are biomarkers of cardiometabolic disease risk, and understanding the genomic factors contributing to their distribution has been of considerable interest. Large meta-analyses have identified over 150 lipids loci; however, there has not yet been a GWAS of Africans. Given the genomic diversity among those of African ancestry, it is expected that a GWAS in Africans could yield novel lipids loci. While GWAS have been conducted in African Americans (AA), because of the drastically different environmental context, these studies are not proxies for studies in continental Africans. Therefore, we conducted a GWAS of 4,116 Africans enrolled in the AADM study. We used linear mixed models of the inverse normal transformations of covariate-adjusted residuals of lipids traits with adjustment for 3 principal components and the random effect of relatedness. Replication of loci associated at $p < 10^{-6}$ was attempted in 9,542 AA from the HUGS, ARIC, CFS, MESA, and JHS studies. We also evaluated known lipids loci in Africans using both exact replication and "local" replication, which accounts for interethnic differences in linkage disequilibrium. We identified 3 novel loci: rs4908578 (*CAMTA1*) with TG (MAF 0.06, $p = 2 \times 10^{-8}$) and TG/HDL ($p = 9 \times 10^{-9}$), rs1003611624 (intergenic) with TG/HDL (MAF 0.04, $p = 1 \times 10^{-8}$), and rs142840854 (*TRDN*) with LDL (MAF 0.07, $p = 3 \times 10^{-9}$). None of these replicated in AA despite similar allele frequencies, however, meta-analysis of Africans and AA identified novel locus rs1199578 (intergenic) associated with TG (MAF 0.42, $p = 4 \times 10^{-8}$). Only 14% of previously identified SNP-Trait associations replicated in our data. With additional replications using the local method, the replication rate was 17% (19% for SNPs with $MAF \geq 0.05$ in our data). In this first GWAS of Africans, we identify 3 novel loci, and discover an additional novel locus upon meta-analysis with AA. Identified variants alter regulatory motifs (rs142840854, rs1199578) and occur in enhancer histone marks (rs142840854, rs1199578) and DNase hypersensitivity sites (rs1199578), and rs1199578 is an eQTL for *SLC22A23* and *ST8SIA2*. Replication of these loci is being sought in a comparable sample of Africans, to reduce the effect of divergent environmental backgrounds observed between Africans and AA. None of these loci have been previously identified in traits or pathways known to influence serum lipids, and if these associations are replicated functional work will be needed to determine their role.

2209T

Genetics of hypertriglyceridemia: An assortment of polygenic effects.

J.S. Dron^{1,2}, J. Wang¹, M.A. Iacocca^{1,2}, J.R. Menard³, H. Cao¹, A.D. McIntyre¹, I.G. Movsesyan⁴, M.J. Malloy⁴, C.R. Pullinger⁴, J.P. Kane⁴, R.A. Hegele^{1,2,5}. 1) Robarts Research Institute, Schulich School of Medicine and Dentistry, Western University, London ON, Canada; 2) Department of Biochemistry, Schulich School of Medicine and Dentistry, Western University, London ON, Canada; 3) Department of Physics and Astronomy, University of Waterloo, Waterloo ON, Canada; 4) Cardiovascular Research Institute, University of California San Francisco, San Francisco CA, United States; 5) Department of Medicine, Schulich School of Medicine and Dentistry, Western University, London ON, Canada.

Hypertriglyceridemia (HTG) is a complex trait influenced by genetic and non-genetic factors. Genetic determinants include both rare and common single nucleotide variants (SNVs) in triglyceride (TG)-associated genes. However, the individual genetic influences on HTG have so far been examined only in a piecemeal manner. Furthermore, copy-number variants (CNVs) in HTG patients have not been systematically evaluated. Here, we concurrently assessed rare SNVs and CNVs, and the accumulation of common SNPs in a discovery and replication cohort of 377 and 312 Caucasian patients with severe HTG (TG $\geq$ 10 mmol/L), respectively. As a contrast, we studied 503 healthy Caucasians from the 1000 Genomes Project. Patient DNA was subjected to targeted next-generation sequencing (the LipidSeq panel) of >70 genes and 185 SNPs associated with dyslipidemia. We first screened for rare SNVs and CNVs in TG-associated genes. For rare variants with likely large phenotypic effects, 2.7% and 1.3% of subjects had bi-allelic SNVs, 16.7% and 15.0% had heterozygous SNVs, and 0.8% and 0.7% had heterozygous CNVs in the HTG discovery and replication cohorts, respectively. In the normolipidemic controls, 4.0% of subjects had heterozygous SNVs. We then assessed patients for an accumulation of common SNPs using a polygenic risk score. We identified an extreme score (defined as >90th percentile of the general population) in 29.4% and 29.3% of discovery and replication patients, respectively, compared to 10.1% of normolipidemic controls. Taken together, 49.6% and 45.3% of the HTG discovery and replication patients had either a rare variant or high polygenic burden, compared to only 14.1% of controls. Since there was no significant difference in proportions of genetic factors between patient cohorts, we combined them into a single affected HTG cohort to derive global risk ratios. Compared to normolipidemic controls, HTG patients are 5.25-fold (CI 95% 3.22-8.55; $P<0.0001$) more likely to carry a rare variant, 5.11-fold (CI 95% 3.62-7.20; $P<0.0001$) more likely to have a polygenic burden, and 5.78-fold (CI 95% 4.30-7.77; $P<0.0001$) more likely to carry any TG-related genetic factor. We thus report the most in-depth, systematic evaluation of genetic contributors of severe HTG to date. Next steps include: 1) identification of novel genetic determinants in patients negative for genetic determinants studied here; and 2) evaluations of genotype differences in clinical outcomes and intervention response.

2210F

Low density lipoprotein cholesterol and sepsis: A clinical and genetic approach.

Q. Feng¹, W.Q. Wei², S. Chaugai¹, G. Carranza-Leon³, J.D. Mosley^{1,2}, L. Jiang¹, A. Ihegword¹, C. Shaffer¹, C. Chung^{1,4}, C.M. Stein^{1,5}. 1) Division of Clinical Pharmacology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN; 2) Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN; 3) Division of Diabetes, Endocrinology and Metabolism, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN; 4) Division of Rheumatology & Immunology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN; 5) Department of Pharmacology, Vanderbilt University, Nashville, TN.

Background: Sepsis is a major global health concern. Lipoproteins, including low-density lipoprotein cholesterol (LDL-C), bind bacterial toxins and facilitate their clearance. In animal models, high LDL-C levels were protective against sepsis, and in patients with sepsis, low LDL-C levels were a marker for poor prognosis. An study of community-dwelling adults showed that low LDL-C levels were associated with increased risk of sepsis, but residual confounding remained a concern. We used a clinical and genetic approach to test the hypothesis that low LDL-C levels are directly associated with increased risk of sepsis and poorer outcomes. **Methods:** Using de-identified electronic health records (EHR) we defined a cohort of patients admitted to Vanderbilt University Medical Center with infection. Patients were older than 18 yrs, had an ICD code indicating infection, and received an antibiotic within 2 days of hospital admission. We used 3 approaches: 1) clinically measured LDL-C levels (excluding measurements <1 year before hospital admission and those associated with acute illness) (N=3961), 2) a genetic risk score (GRS) for LDL-C derived from 81 SNPs reported by the Global Lipid Genetics Consortium (GLGC) (N=7804), 3) a Bayesian sparse linear mixed modeling (BSLMM) GRS for LDL-C (N= 4397) derived from ARIC. For each approach we tested the association with 3 outcomes: (1) development of sepsis (defined using a validated algorithm), (2) admission to the intensive care unit (ICU) and (3) in-hospital death. Analyses were adjusted for age, sex, Charlson/Deyo comorbidity index, HDL-C, triglycerides, EHR length, and BMI. **Results:** In unadjusted analyses, low measured LDL-C levels were strongly associated with increased risk of sepsis ($P=0.001$) and ICU admission ($P=0.008$), and weakly with in-hospital mortality ($P=0.055$). None of these associations was statistically significant after adjusting for covariates (all P values > 0.05). Both the GLGC-based LDL GRS and the BSLMM-based LDL GRS significantly predicted LDL-C levels (all $P<2\times 10^{-16}$), but neither GRS was significantly associated with any of the outcomes. **Conclusion:** Low measured LDL-C levels were significantly associated with increased risk of sepsis in patients admitted to hospital with infection but this was due to confounding by co-morbidities since adjusted clinical models and genetic models showed no increased risk. LDL-C levels do not directly alter the risk of sepsis or poor outcomes.

2211W

Genetic risk determination of hypercholesterolemia at the individual patient level. C.C.F. Ma, M.M Roth, L. Shen. Genben Lifesciences, San Diego, CA.

Background: Severe hypercholesterolemia (SHC; >190mg/dL) is a major risk factor for atherothrombotic diseases (ASCVD). The genetic architecture is complex and heterogeneous with monogenic and polygenic causes. Studies show the clinical utility and cost-effectiveness of genetic analysis in assessing SHC risk. We have developed and are testing a comprehensive NGS panel for genetic risk assessment of ASCVD. Our panel analyzes coding and selected non-coding regions identified in GWAS to comprehensively quantify risk. We have developed a 0-100 risk score within major classes of dyslipidemia/vascular disease that weights and adds/subtracts based on the known or predicted effect of a variant and grouped by patient's ancestry. **Methods:** 48 patients with primary cause of SHC (ICD-10 codes: E78.01 or E78.5) from 3 academic-based lipid clinics were analyzed with our GBinsight NGS panel on an Illumina instrument. Laboratory work is performed at CLIA-certified Otagenetics Inc. Salivary DNA was acquired from patients after providing signed consent. **Results:** We identified 5 patients with heterozygous pathogenic/likely pathogenic familial hypercholesterolemia variants: p.Arg3527Gln in APOB and p.Gly219del, p.Cys681Ter, p.Pro685Leu and p.Met652Glyfs within LDLR. We identified 4 patients with heterozygous pathogenic/likely pathogenic variants in LPL that cause combined hyperlipidemia. Separately, 3 family members carried heterozygous ApoE4-Freiburg subtype. We identified several very rare variants (MAF<0.1%) of uncertain significance (VUS) including: p.Asn76Lys and p.Leu127Phe in ANGPTL3, p.Ser934Cysfs and p.Arg989His in ABCA7, p.Ala161Ser and p.Glu185Lys in APOA4, p.Val442Gly in LPL and p.Met2Thr in STAP1. Common variants enriched in our cohort include: p.Ser19Trp APOA5, p.Leu446Pro in GCKR, p.Ile1891Met in LPA, p.Asp19His in ABCG8, p.Asp-2702Gly in SVEP1, APOE4 and an intergenic variant near the ASGR1 gene (rs314253). The median polygenic risk score for hypercholesterolemia in our cohort is 74 (25.7 SD) with a range of 4-100. Median risk score for combined hyperlipidemia is 66.5 (34.7 SD) with a range of 0-100. **Conclusion:** At the individual patient level, a complex mix of monogenic and polygenic variants affect SHC. Combinatorial genetic epistasis may explain the heterogeneity in clinical phenotypes observed at the individual patient SHC level. Risk stratification may help to apply more precise medical management and reduce morbidity, mortality and cost.

2212T

Replication of association between triglyceride level and SLC25A40 in a multi-ancestry sample. E.A. Rosenthal¹, A.S. Gordon¹, I.B. Stanaway², D.R. Crosslin², L. Albertson-Junkans³, E.B. Larson³, D. Carrell³, G.P. Jarvik¹ on behalf of the eMERGE Network. 1) Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA; 2) Department of Biomedical Informatics and Medical Education, University of Washington, Seattle, WA, USA; 3) Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA.

Elevated triglycerides (TG) is associated with, and may cause, cardiovascular disease (CVD). Hypo- and hypertriglyceridemia (HTG) have major genetic components (e.g., APOC3 and APOA5, respectively). Other genes (e.g., APOE) have smaller effect on TG. Knowledge of genetic causes for high TG may lead to early intervention and targeted treatment for CVD. We previously identified linkage and association of a rare, highly conserved missense SNV in SLC25A40 with HTG in a large family, as well as an association with rare missense SNVs at highly conserved sites in SLC25A40 in a European American (EA) and African American (AA) sample (PMID: 24268658). SLC25A40 encodes a mitochondrial transmembrane protein. In *Drosophila*, SLC25A40 loss-of-function and missense SNVs were shown to contribute to mitochondrial damage due to excessive oxidation. Here, we replicate the association in a multi-ancestry population (Kaiser Permanente Washington) from the Electronic Medical Records and Genomics (eMERGE) Network phase 3, freeze 1 study. As TG vary widely within individuals, we calculated the median TG (TGmed) over all medical records prior to onset of lipid medication (LRx) (N=4909, TGmed range: 30-1840 mg/dL; median age range: 24-90). For individuals with records only post-LRx (N=397), we imputed pre-LRx TGmed based on post-LRx TG, sex, fibrate use, and onset age of LRx (TGmed range: 61-2532 mg/dL, age range: 41-81). Of these, 963 participants (755 EA, 35 AA, 123 Asian American (AS), 50 mixed ancestry; 54% male) were sequenced on a custom platform, consisting of study-wide nominated genes and SNVs, and the original 56 ACMG medically actionable genes. Five EA individuals were removed due to having a known TG lowering SNV in APOC3 (N=4; TGmed range: 35-78 mg/dL) or a TG elevating SNV in APOA5 (N=1; TGmed=178 mg/dL). Our trait of interest (TGres) was the residuals from the log-transformed TGmed regressed on median age, sex, APOE, and the 1st 10 components of ancestry. At SLC25A40, we observed 4 rare (maxpopMAF < 0.001) missense SNVs and 1 rare frameshift. Heterozygotes at the missense SNVs (3 EA, 1 AS) had lower mean TGres than the rest of the sample (β =-0.61, p=0.02). The individual with the frameshift was in the upper 2nd percentile of TGmed and of mixed EA/AS ancestry. We replicate an association between rare missense SNVs in SLC25A40 and TG in a multi-ancestry sample and show that the direction of effect may be variable, warranting further study of how SLC25A40 affects TG.

2213F

Best practices for polygenic risk score modeling of laboratory values from biobank data: Application to blood lipid levels. J. Dennis^{1,2}, G. Chen³, P. Straub^{1,2}, D. Hucks^{1,2}, J.D. Mosley⁴, R. Tao^{2,5}, W.L. Song⁶, M.F. Linton⁶, D. Ruderfer^{1,2}, N.J. Cox^{1,2}, L.K. Davis^{1,2}. 1) Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN; 2) Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN; 3) Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI; 4) Department of Medicine, Vanderbilt University Medical Center, Nashville, TN; 5) Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN; 6) Division of Cardiology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN.

Background: Polygenic risk score (PRS) modeling of laboratory values from electronic health records could provide insights into trait architecture, improve trait prediction, and identify novel biomarkers. Optimal strategies for analyzing these data, however, are unclear. Our objectives were to determine how best-fit p-value thresholds (pT) for SNP inclusion in a PRS, as well as the proportion of variability explained (R^2) by PRS varied across: (i) a two-stage strategy for PRS validation and testing; (ii) strategies for control of medication and disease; and (iii) models of repeated measures, with and without accounting for trajectories over time. **Methods:** We developed our approach in 25,461 patients (9,259 validation, 16,202 testing) in the Vanderbilt University Medical Center biobank (BioVU) with at least one lipid measurement. Median values were calculated. Discovery GWAS results for PRS weighting and thresholding were obtained from the Global Lipids Genetics Consortium. We constructed PRS for high- and low- density lipoprotein cholesterol (HDL and LDL) and for triglycerides (TG) at discovery GWAS p-value thresholds from 5×10^{-4} to 0.5 in intervals of 5×10^{-4} , and tested these PRS for association with the same trait in BioVU. We defined pT as the threshold that maximized the R^2 of the PRS in the validation set. **Results:** (i) The PRS at pT explained 0.059, 0.019, and 0.043 of testing set variance in HDL, LDL, and TG, and the pT of HDL and LDL were remarkably close to the pT fit independently in the testing set. (ii) Restricting to observations before lipid-lowering medication use (~30% of all observations) lowered the pT, in accordance with the selection of patients in the discovery GWAS, and improved the R^2 for HDL, LDL, and TG by 7.4, 76.8, and 4.3%. Excluding patients with coronary artery disease or type 2 diabetes had marginal consequences. (iii) Preliminary analyses of repeated measures confirmed that only pre-medication LDL levels were strongly associated with a PRS of genome-wide significant SNPs. **Conclusions:** PRS are associated with lipid levels measured in a biobank. However, no tools exist to calculate pT using repeated measures. Therefore, we will next adapt the pT framework to longitudinal data and test whether PRS calculated at different p-value thresholds explain trajectories over time. As biobank data continue to accrue, methods that leverage repeated measures could shed light on the role of genetics in disease progression.

2214W

Identifying genetic determinants of osteoporosis and obesity in post-menopausal women. S. Zhou^{1,2}, L. Laurent³, V. Forgetta³, J.A. Morris^{2,3}, J. Vijay⁴, A. Pramatarova⁴, E. Grundberg^{4,5}, J.B. Richards^{1,2,3,6}. 1) Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Quebec, Canada; 2) Lady Davis Institute, Jewish General Hospital, McGill University, Montreal, Quebec, Canada; 3) Department of Human Genetics, McGill University, Montreal, Quebec, Canada; 4) McGill University and Genome Quebec Innovation Centre, Montreal, Quebec, Canada; 5) Children's Mercy Hospitals and Clinics, Kansas City, Missouri, USA; 6) Department of Twin Research and Genetic Epidemiology, King's College London, London, UK.

Background: Osteoporosis is an age-related disease that leads to increased fracture risk, with an increase in waist to hip ratio (WHR) - a clinically relevant obesity marker - are two common public health issues of women after menopause. WHR and bone mineral density (BMD), the major risk factor for osteoporosis have high genetic heritability, yet no shared pathophysiological mechanisms have been identified. A recent finding identified that inhibiting follicle-stimulating hormone (FSH) leads to reduced body fat and bone loss in ovariectomized mice, however, its relevance in humans has not been tested. If this were true, pharmacological manipulation of such biomarkers may be possible. **Methods:** We performed genome-wide association studies (GWAS) of BMD and WHR in the UK Biobank, which include 434,985 individuals and deeply imputed to ~18.6 million genetic variants after QC, stratified by sex and menopausal status of women. The sex and menopausal status heterogeneity were also tested in a large meta-analysis comprised of GIANT and UK Biobank (N=694,649) to determine the genetic etiology differences of BMD and WHR between sex and age. Two-sample Mendelian Randomization (MR) was performed using 1,538 postmenopausal women from TwinsUK and 9,900 individuals from Canadian Longitudinal Study on Aging to identify the genetic instruments for FSH level and its association with BMD and WHR. To further understand the function of FSH, the expression of FSH genes in different stages of mesenchymal stem cell differentiation were measured and the effect of FSH on adipocytes differentiation will also be assessed. **Results:** We found that there was a strong evidence for variants at the *LHCGR-FSHR* locus to be associated with WHR in postmenopausal women (N=193,400, $p=1 \times 10^{-16}$), and this association was attenuated in pre-menopausal women (N=30,493, $p=1 \times 10^{-4}$) and men (N=211,092, $p=5 \times 10^{-6}$). The MR results also showed that a decrease in FSH level was associated with increased BMD in postmenopausal women ($p=2.99 \times 10^{-6}$). In adipocyte culture, *FSHR* expression was absent in un-differentiated and matured adipocytes but reached peak level of expression at one week after adipocyte differentiation. **Conclusion:** These results show that the FSH signaling pathway has an effect on WHR and BMD, which is pronounced in postmenopausal women. We provided further evidence of a role for FSH in osteoporosis and obesity, which could present a tractable drug target to treat these common and costly diseases.

2215T

Genetic polymorphism of LDLR (rs688) and correlation of inflammatory marker with genetic damage in patients with hypercholesterolemia. *M. Monu, G. Gandhi.* Department of Human Genetics, Guru Nanak Dev University, Amritsar, Punjab, India.

The low density lipoprotein receptor is a cell surface receptor that plays an important role in cholesterol homeostasis by binding and uptaking the plasma LDL particles. LDLR gene variations can lead to unbinding of receptor with LDL particles, thus resulting in high levels of circulating LDL-C which may induce an inflammatory response and its atherogenic effects can cause DNA damage. Alongwith genetic damage, this inflammatory response can induce production of C-reactive protein (CRP) by hepatocytes. Methodology: In this study, the venous blood samples of patients (n=100, 20-40y) from local hospitals and age-, sex- and socioeconomic status- matched healthy controls from general population (n=100, 20-40y) were withdrawn after informed consent and approval by the Institutional Ethical Committee(IEC). Genomic DNA was isolated from venous blood samples and *LDLR*-rs688C > T gene polymorphism was detected using RFLP-PCR. DNA damage using the alkaline Single Cell Gel Electrophoresis(SCGE) assay was investigated in peripheral blood leukocytes of the subjects for damage index, damage frequency and per cent DNA in tail. using comet assay/. Isolated serum was used for assessment of CRP levels. Results: Statistical analyses revealed that CRP levels were significantly higher ($p=0.000$) in patients ($17.506\pm0.706\text{mg/L}$) as compared to that in controls ($2.986\pm0.257\text{mg/L}$). Damage frequency (95.705 ± 0.461 , $p=0.050$), damage index(144.66 ± 1.117 ; $p=0.001$) and per cent DNA in tail(45.836 ± 1.031 , $p=0.000$) were also significantly increased in patients. Positive association of LDL-C with damage index($p=0.050$) and with CRP($p=0.000$) while negative association of HDL-C with CRP($p=0.001$) was observed in correlation analysis. The observed frequencies of genotypes CC,CT and TT in patient group were 30%, 65% and 5%, respectively, while all healthy control samples were homozygous wild type. The genetic damage as a function of *LDLR*(rs688) genetic polymorphism revealed no significant differences in patients and controls. Conclusion: In the absence of co-morbidities and/or exposures, the results of this study imply that the hypercholesterolemic condition acts as causal factor for increased inflammation and DNA damage. However dietary and lifestyle management can assist in reducing risk to cardiovascular and hepatocellular diseases. .

2216F

Genome-wide association studies of ezetimibe response in IMPROVE-IT trial. *G. Kosova¹, C.T. Ruff², Y. Zhang³, Z. Guo³, D.F. Reilly¹, D.V. Mehrotra², Y. Mitchell⁴, S.H. Shah², M.A. Blazing⁵, E.A. Bohula², R.P. Giugliano², C.S. Fox¹, M.S. Sabatine², P.M. Shaw¹.* 1) Genetics and Pharmacogenomics, Merck, Kenilworth, NJ; 2) TIMI Study Group, Brigham and Women's Hospital, Boston, MA; 3) Biostatistics and Research Decision Sciences, Merck, Kenilworth, NJ; 4) Clinical Research, Merck, Kenilworth, NJ; 5) Duke Clinical Research Center, Durham, NC.

IMProved Reduction of Outcomes: Vytorin Efficacy International Trial (IMPROVE-IT) demonstrated a significant benefit of Vytorin (ezetimibe [EZ]+simvastatin [simva]) vs. simva alone in lowering LDLc levels and reducing the risk of cardiovascular events. However, inter-individual variability in drug response, as measured by reduction in LDLc levels, has been observed with statin therapy. Previous studies of statin monotherapy have demonstrated common pharmacogenetic variants contribute to variation in drug response. Therefore, we performed a GWAS in ~6,500 participants of the IMPROVE-IT trial to test the hypothesis that genetic variation is associated with EZ efficacy. In total, five independent loci were associated with fold-change in LDLc levels after one month of lipid lowering therapy both in the simva monotherapy arm and in the simva+EZ combination arm at $P<5\times10^{-8}$. Four of these loci (*APOE*, *LPA*, *CELSR2/SORT1* and *SLCO1B1*) have previously been implicated in statin response. However, our study showed a trend for larger effect sizes of these genetic variants in the EZ+simva arm, compared with simva monotherapy. For example, patients receiving simva+EZ showed on average 15.9% additional reduction in LDLc per number of *APOE*- $\epsilon 2$ allele (e.g., if a patient carrying no copies of $\epsilon 2$ and has a LDLc of 100mg/dL at baseline had a reduction to ~55mg/dL at 1 month of treatment, then a patient carrying 1 copy of $\epsilon 2$ would have a reduction to ~46mg/dL; $P=1.2\times10^{-28}$), compared to 8.6% additional reduction per allele in patients treated with simva alone (e.g. from 100mg/dL to 70mg/dL if carrying no $\epsilon 2$, and to 64 mg/dL if carrying 1 copy; $P=9.0\times10^{-9}$, $P_{\text{tag}}=2.7\times10^{-4}$). When tested as a composite score combining LDLc lowering alleles at these four loci, there was 9.2% additional reduction in LDLc per LDLc reducing allele in the simva+EZ arm ($P=1.5\times10^{-37}$), compared to 5.2% in the simva alone ($P=1.2\times10^{-12}$; $P_{\text{tag}}=2.5\times10^{-5}$). We also identified one novel association with an intronic variant in *NCK1* (MAF=0.02) only in patients receiving simva+EZ, however this association requires a further validation. In conclusion, we replicated the four previously reported "statin-response" associations with in patients receiving simva monotherapy, and simva+EZ combination therapy. The larger effect sizes that we observed in the combination arm suggest that these variants are not statin-specific, and may have distinct roles in influencing the magnitude of LDLc reduction in response to ezetimibe.

2217W

Linking genome to phenome through lipidomic profiles identifies disease susceptibility genes.

R. Tabassum¹, J.T. Rämö¹, P. Ripatti¹, J.T. Koskela¹, M. Kurki^{2,3}, J. Karjalainen^{1,3,4}, S. Hassan¹, J. Nunez-Fontana¹, S. Söderlund⁵, N. Matikainen^{5,6}, M.J. Gerl⁷, C. Klose⁷, N.O. Stitzel^{8,9,10}, A.S. Havulinna^{1,11}, S.K. Service¹², V. Salomaa¹¹, N.B. Freimer¹², M. Pirinen^{1,13,14}, M. Jauhiainen^{11,15}, A. Palotie^{1,16,17}, M.R. Taskiran⁵, K. Simons^{7,18}, S. Ripatti^{1,17,19}. 1) Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland; 2) Medical and Population Genetics, Broad Institute, Cambridge, MA, USA; 3) Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; 4) Broad Institute of the Massachusetts Institute of Technology and Harvard University, Cambridge, MA, USA; 5) Research Programs Unit, Diabetes & Obesity, University of Helsinki and Department of Internal Medicine, Helsinki University Hospital, Helsinki, Finland; 6) Endocrinology, Abdominal Center, Helsinki University Hospital, Helsinki, Finland; 7) Lipotype GmbH, Dresden, Germany; 8) Cardiovascular Division, Department of Medicine, Washington University School of Medicine, Saint Louis, MO, USA; 9) Department of Genetics, Washington University School of Medicine, Saint Louis, MO, USA; 10) McDonnell Genome Institute, Washington University School of Medicine, Saint Louis, MO, USA; 11) National Institute for Health and Welfare, Helsinki, Finland; 12) Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, California, USA; 13) Department of Public Health, University of Helsinki, Helsinki, Finland; 14) Helsinki Institute for Information Technology HIIT and Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland; 15) Minerva Foundation Institute for Medical Research, Biomedicum, Helsinki, Finland; 16) Program in Medical and Population Genetics and The Stanley Center for Psychiatric Research, The Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; 17) Psychiatric & Neurodevelopmental Genetics Unit, Department of Psychiatry, Analytic and Translational Genetics Unit, Department of Medicine, and the Department of Neurology, Massachusetts General Hospital, Boston, Massachusetts, USA; 18) Max Planck Institute of Cell Biology and Genetics, Dresden, Germany; 19) Department of Public Health, Clinicum, Faculty of Medicine, University of Helsinki, Helsinki, Finland.

Background: Genetic investigation of the plasma lipidome, composed of a large number of structurally and functionally diverse lipids species, may elucidate lipid metabolism and its link to diseases beyond traditional lipid measures. **Methods:** We carried out a genome-wide association study of 141 lipid species in 2,045 Finnish individuals with ~9.3 million genetic variants, followed by phenome-wide scans (PheWAS) with 58 clinical end-points related to cardiovascular, metabolic, psychiatric and gastro-intestinal disorders in >383,000 individuals from the Finnish and UK Biobanks. **Results:** We found that: 1. *Lipid species are quite heritable (h^2_{SNP} ranged 10-54%).* Lipids with ω -3 and ω -6 fatty acids showed highest heritability and shared genetic component, suggesting considerable genetic regulation at acyl chains. 2. *35 genomic loci, including 24 novel, were associated with at least one lipid species ($P < 5 \times 10^{-8}$).* 3. *Lipidomic profiles could help to identify new disease risk-modifying variants.* The PheWAS showed associations for eight of 35 identified loci (APOA1, ABCG5/G8, BLK, LPL, LIPC, FADS2, MIR100HG, DHRS12) with various diseases (false discovery rate <0.05), including novel disease associations (Table 1). 4. *Lipidomic profiles provide clues to new biology, both in new and known lipid loci.* The association patterns of novel variants at SYNGR1, MIR100HG, PTPRN2, VWA3B and ABLIM2 suggested their possible involvement in desaturation and/or elongation of fatty acids. At many known lipid loci including ABCG5/G8, MBOAT7, FADS2 and LPL, our genetic associations provided better understanding of their roles in lipid biology and complex diseases. 5. *Traditional lipids used in routine clinical work fail to capture association with lipid species such as LPCs and CEs that are potential disease risk factors.*

Conclusions: Our study reveals genetic regulation of detailed plasma lipids and highlights the importance of lipidomic profiling in disease gene mapping.

Table 1: Selected novel variants associated with plasma lipid species and diseases (*reported previously)

Locus (Gene)	Effect on transcriptome	Effect on lipidome	Effect on phenotype	Known gene functions
rs1478898-A(BLK)	↑ BLK	↑ PC (16:0;0-16:0;0)	↓ Hypertension* ↓ Venous thromboembolism ↓ Obesity ↑ Gallstones	Immune response, Insulin synthesis and secretion
rs512948-C(DHRS12)	NA	↓ LPE-18:2;0	↓ Heart failure	Metabolism of lipid compounds
rs10790495-G(MIR100HG)	↓ HSPA8	↓ TAG-56:4;0	↑ Venous thromboembolism	Regulation of cell growth

2218T

Family-based heritability of statin low-density lipoprotein cholesterol response.

A. Oni-Orisan^{1,2}, T.J. Hoffmann^{2,3}, D. Ranatunga⁴, M.W. Medina⁵, E. Jorgenson⁴, C. Schaefer⁴, R.M. Krauss^{5,6}, C. Iribarren^{3,4}, N. Risch^{2,3,4}. 1) Department of Clinical Pharmacy, University of California, San Francisco, CA; 2) Institute for Human Genetics, University of California, San Francisco, CA; 3) Department of Epidemiology and Biostatistics, University of California, San Francisco, CA; 4) Kaiser Permanente Northern California Division of Research, Oakland, CA; 5) Children's Hospital Oakland Research Institute, Oakland, CA; 6) Department of Medicine, University of California, San Francisco, CA.

The genetic basis of low-density lipoprotein cholesterol (LDL-C) response to statin therapy has not been fully elucidated. In particular, there are no reports describing the heritable component of variability in statin-induced changes of LDL-C. The primary objective of this study was to estimate heritability of statin LDL-C response in a large, multi-ethnic population of statin users. We calculated familial correlations in first-degree relatives from 33,139 statin users of the Kaiser Permanente Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort, which also includes genome-wide genotype data. The effect of statin dosing on LDL-C was determined from electronic health records (EHRs). We first demonstrated the validity of the EHR-derived LDL-response by showing a statin dose-response relationship comparable to that seen in randomized clinical trials. We also identified age, sex, smoking, diabetes and race/ethnicity as significant covariates. Familial first-degree relationships were determined genetically using KING v1.4. First-degree relatives included parent-offspring and sibling pairs. To estimate heritability, we first applied a square-root transformation to the covariate-adjusted statin LDL-C response residuals for trait normalization and calculated an interclass correlation for the parent-child pairs and intraclass correlation for the sib pairs. Heritability was determined from the doubled correlations. Among the 33,139 statin users, we identified 229 parent-offspring pairs and 296 sib pairs. The parent-offspring correlation of statin LDL-C response was 0.060 (SE = 0.067, P = 0.365) and the sibling correlation was 0.054 (SE = 0.059, P = 0.357). The heritability estimate was 0.114 (95% confidence interval -0.06 to 0.286), which was below the threshold of statistical significance. Thus, statin-induced changes in LDL-C were found to be modestly inherited. These results provide novel information about the contribution of genetic factors to statin LDL-C response variation.

2219F

Causal association of lipid fractions with estimated glomerular filtration rate. A multivariable Mendelian randomization analysis of the HUNT Study, Norway. H. Rasheed^{1,2,3}, J. Rees⁴, B. Brumpton^{1,2,5}, J. Zheng⁶, G.D. Smith⁷, S. Burgess^{4,6}, B.O. Asvold^{1,7}. 1) K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, Norwegian University of Science and Technology (NTNU), Trondheim, Norway; 2) MRC Integrative Epidemiology Unit, University of Bristol, UK; 3) Department of Chemistry, University of Engineering and Technology, Lahore, Pakistan; 4) Cardiovascular Epidemiology Unit, University of Cambridge, Cambridge, UK; 5) Department of Thoracic Medicine, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway; 6) MRC Biostatistics Unit, University of Cambridge, Cambridge, UK; 7) Department of Endocrinology, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway.

Background: In observational studies, higher low-density lipoprotein cholesterol (LDL-C), lower high-density lipoprotein cholesterol (HDL-C) and higher triglyceride (TG) levels have been associated with lower estimated glomerular filtration rate (eGFR). We aimed to investigate whether these associations are likely to be causal in the Nord-Trøndelag Health Study (HUNT), using multivariable Mendelian randomization (MR), being advantageous then conventional MR methods, to disentangle the effects of the separate lipid fractions. **Methods:** In a two-sample MR approach, we obtained information on associations of 185 independent lipid-associated genetic variants with serum levels of LDL-C, HDL-C and TG from the Global Lipids Genetics Consortium (n=188,577). Associations of these variants with eGFR were estimated in 69,570 participants from the HUNT. Causal associations of serum lipid levels with eGFR were estimated both using standard inverse-variance weighted (IVW) MR and multivariable MR. Evidence of pleiotropy was evaluated using MR Egger regression. All analyses were adjusted for age and sex. The estimates (b [95% CI]) represent the difference in eGFR (mL/min/1.73 m²) per 1 standard deviation higher lipid levels. **Results:** In observational analyses, higher measured LDL-C (-0.0033 [-0.0039, -0.0028], P<10⁻¹⁶) and TG (-0.0015 [-0.0018, -0.0013], P<10⁻¹⁶) were associated with lower eGFR, and higher HDL-C was associated with higher eGFR (0.0033 [0.0028, 0.0039], P<10⁻¹⁶). The IVW method showed no evidence of a causal association of LDL-C (0.441 [-0.100, 0.982], P=0.11) or HDL-C (-0.214 [-0.800, 0.373], P=0.47) with eGFR, but suggestive evidence that TG could influence eGFR (0.663 [-0.057, 1.383], P=0.07). However, using multivariable MR, there was no strong evidence for causal associations of either LDL-C (0.346 [-0.209, 0.901], P=0.22), HDL-C (0.015 [-0.627, 0.657], P=0.96) or TG (0.566 [-0.245, 1.377], P=0.17) with eGFR. MR Egger regression showed no evidence of pleiotropy in any of the analyses. **Conclusion:** Our findings do not support a causal effect of serum lipid concentrations on eGFR. Further research should use Mendelian randomization in larger well-phenotyped cohorts to replicate and extend our findings.

2220W

Regional genomic heritability analysis of waist circumference in a Japanese population. O. Gervais¹, K. Misawa¹, Y. Harada¹, T. Mimori¹, Y. Kawai², K. Tokunaga², M. Nagasaki¹. 1) Tohoku University, Sendai, Miyagi, Japan; 2) The University of Tokyo, Tokyo, Japan.

The obesity epidemic, which was initially mostly conspicuous in high-income countries, has become a worldwide health issue; for instance, it has been reported recently that in Bangladesh, a low-income country, the prevalence of overweight/obesity has reached one-fourth of the population and is higher in younger age groups. Nevertheless, and in spite of fact that regional genetic differences are known to affect obesity, the majority of studies on the underlying genetic factors of obesity are conducted in populations of European descent. To better understand the regional genetic differences of obesity, this study primarily aims to quantify the effect of each genomic region on variation in obesity in a Japanese population of almost 10,000 individuals by using SNP array genotyping data. The trait considered in this study was waist circumference (WC), as it has been suggested in several studies that it may be a better indicator of cardiovascular disease risk than BMI. In this study, several polygenic models were applied to estimate the genetic components of the population considered at three different levels: at the genome-wide level to estimate trait SNP heritability, at the chromosomal level, as well as at the regional level to estimate regional genetic variance. The analysis of each genomic region (a region being represented by a given number of adjacent SNPs) was conducted by using the Regional Heritability Mapping approach. The models were adjusted to account for relevant environmental factors and population structure.

2221T

Genome-wide association study identifies 4 novel loci with metabolites in East Asians. J.F. Chai¹, I.W. Khor², F. Torta³, D.R. Herr⁴, C.M. Khoo², M.R. Wenk³, E.S. Tai^{1,2,5}, X. Sim¹. 1) Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore; 2) Department of Medicine, Yong Loo Lin School of Medicine, National University Health System, Singapore; 3) Singapore Lipidomics Incubator (SLING), Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore; 4) Department of Pharmacology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore; 5) Duke-NUS Medical School, Singapore.

Metabolites are small intermediate products of cellular metabolism, which can serve as biomarkers for understanding complex disorders and to aid in drug development. Here, we genotyped 2,434 Singapore Chinese across three Illumina arrays, and profiled 14 amino acids, 45 acylcarnitines, and 103 sphingolipids in plasma samples to understand the role of genetic variants on these metabolites. We conducted single variant association tests using an additive linear mixed model between 47 million genetic dosages and 232 metabolites (including 70 derived sphingolipid ratios). Association test statistics across arrays were combined using fixed-effects sample-size weighted meta-analysis. At a conservative genome-wide threshold of 2.16×10^{-10} ($5 \times 10^{-8}/232$), we report associations at 17 genetic loci with 64 metabolites. Four of these loci represented novel associations with metabolites. The strongest association among these four loci was between dihexosylceramide/lactosylceramide (DHCer) d18:1/C24:1 and a missense variant at *ABCA7* that has been linked to Alzheimer's disease (rs3752246, MAF=0.37; $P=3.21 \times 10^{-30}$), suggesting a potential mechanism linking glycosylceramides and Alzheimer's disease. Other novel loci included *GSTM2* (rs3587154; associated with acylcarnitine C18:1-DC, MAF=0.24; $P=8.52 \times 10^{-13}$), *CHCHD2* (rs57427081; associated with serine, MAF=0.30; $P=8.67 \times 10^{-13}$), and *A4GALT* (rs77076498; associated with the ratio of monohexosylceramide (MHCer) d18:1/C24:0 to DHCer d18:1/C24:0, MAF=0.43; $P=4.63 \times 10^{-28}$). At the remaining 13 loci, after conditioning on previously reported metabolite associations in Europeans, we identified independent (conditional $P < 6.49 \times 10^{-6}$) signals in each metabolite class. For example, the association between intron *CPT2* variant (rs77466051) and acylcarnitine C5-DC (MAF=0.22; $P=1.73 \times 10^{-18}$; conditional $P=2.19 \times 10^{-14}$), was independent of previously reported European GWAS variant rs13375749. We also observed selectivity for the d16 sphingosine backbone of sphingosine-1-phosphate (S1P) for *SPTLC3* and *SGPP1* variants, which were significantly associated with d16:1 S1P (*SPTLC3*: rs438568, MAF=0.44, $P=7.84 \times 10^{-27}$; *SGPP1*: rs34817779, MAF=0.11, $P=1.41 \times 10^{-16}$), but not d18 S1P. This is consistent with the role of *SPTLC3*, which promotes use of myristoyl-CoA in the production of d16:0 sphingoid bases. Altogether, we replicated 13 known metabolic loci and identified 4 new loci in Chinese that are candidate biomarkers for understanding metabolic and other disorders.

2222F

Differential gene expression and large-scale lipid molecular species quantitation during human macrophage activation. M.K. Lin^{1,2}, A. Moreno-Moral³, S. Muralidharan^{2,4}, M. Bagnati⁵, M.R. Wenk^{1,2}, E. Petretto⁴, J. Behmoaras⁵, F. Torta^{1,2}. 1) National University of Singapore, Yong Loo Lin School of Medicine, Department of Biochemistry, Singapore; 2) National University of Singapore, Singapore Lipidomics Incubator, Singapore; 3) DUKE-NUS Medical School, Centre for Computational Biology, Singapore; 4) National University of Singapore, Faculty of Science, Department of Biological Sciences, Singapore; 5) Imperial College London, Centre for Complement and Inflammation Research, London, United Kingdom.

Macrophage activation induced by lipopolysaccharide (LPS) mimics the immune response to bacterial infections. Genes regulating the immune, apoptotic and lipid metabolism pathways have been shown previously to be involved in macrophage activation. In particular, lipid metabolism has been appreciated to be perturbed in a number of clinical pathologies and during inflammation. In this study, we explored gene and lipid molecular species interaction during macrophage activation. Activation of human primary macrophages from 11 individuals was induced with LPS. Cells were harvested before and at 0.5, 3, 8, and 16 hours after activation. Lipids were extracted and quantified using a targeted approach by liquid chromatography-mass spectrometry. Approximately 300 lipid molecular species were measured. RNA of the macrophages across the five time-points were sequenced in four individuals. Gene expression and lipid molecular species levels were corrected for inter-individual variation and correlated. Time series clustering using Gaussian processes was employed to cluster both gene expression and lipid molecular species abundance, followed by functional enrichment annotation for each cluster. The results showed that the sphingolipids and the cholesterol esters pathways are significantly affected after LPS-induced human primary macrophage activation. The sphingolipid pathway can be altered in a number of clinical pathologies, namely cancer progression, cardiovascular and metabolic diseases. However, the clarification of the influence of gene expression on particular lipid molecular species abundance and vice versa, is still an ongoing effort. This study identifies differentially expressed genes previously known to play a role in, and also genes not typically associated with, lipid metabolism during LPS-induced macrophage activation.

2223W

Effects of delivering *SLCO1B1* genotype-informed statin therapy on cholesterol levels: Analysis of commercial laboratory data. D. Voora^{1,2}, E. Perry³, M. Mehan⁴, R. Gill⁴, Y.-J. Li³. 1) Department of Medicine, Duke University, Durham, NC; 2) Center for Applied Genomics & Precision Medicine, Duke University, Durham, NC; 3) Department of Biostatistics and Bioinformatics, Duke University, Durham, NC; 4) Boston Heart Diagnostics, Boston, MA.

Background: HMG-CoA reductase inhibitors ("statins") lower cholesterol and prevent cardiovascular disease (CVD). A common, reduced function variant (rs4149056, *5) in *SLCO1B1* that encodes a hepatic statin transporter is a known risk factor for statin side effects, premature statin discontinuation, and higher cholesterol levels and CVD risk. The *5 variant increases the risk of side effects for simvastatin and possibly atorvastatin but not for rosuvastatin, pravastatin, or fluvastatin – forming the basis of *SLCO1B1* genotype-informed statin therapy (GIST). Commercial laboratory *SLCO1B1* testing and GIST is currently in use. However, outcomes of GIST in a general population are unknown. **Methods:** Based on prior randomized clinical trial findings, we hypothesized that GIST was associated with cholesterol levels. We accessed *SLCO1B1* genetic and longitudinal cholesterol testing data from a single commercial laboratory vendor. We included all patients with >1 cholesterol measurement and excluded patients who received *SLCO1B1* testing after their first cholesterol test. Baseline clinical data including age, sex, lipid-lowering therapy, CVD, and diabetes were used to build a 1:1 propensity score matched dataset for *SLCO1B1* testing. A mixed effects regression model adjusting for sex was used to estimate the effect of *SLCO1B1* testing on cholesterol levels over time. **Results:** Among 143757 patients that met our criteria, 92271 received *SLCO1B1* genetic testing that had, on average, 2.9±1.5 cholesterol measurements over a period of 14.6±10.1 months. Tested patients had more CVD and pre-diabetes and were younger, more likely female, and not using lipid-lowering therapy ($p < 0.01$). After matching, 1907 patients who received *SLCO1B1* testing were similar to those who did not with the exception of sex. Compared to matched, untested controls, patients who received *SLCO1B1* testing had a larger decrease in low-density lipoprotein (-1.8 ± 0.8 mg/dl/year, $p = 0.035$), non-high density lipoprotein (-2.7 ± 1 mg/dl/year, $p = < 0.001$), and total cholesterol (-2.7 ± 1.0 mg/dl/year, $p = 0.006$). In patients receiving GIST, there were no differences in cholesterol levels across *SLCO1B1**5 genotypes. **Conclusions:** In a general population, delivery of *SLCO1B1* genotype-informed statin therapy was associated with modest cholesterol reductions that were smaller than reported in prior randomized trials.

2224T

eQTL mapping in CD4+ T cells from the Barbados asthma genetics study (BAGS). S. Chavan¹, N.M. Rafaels¹, M.A. Taub², M. Daya¹, A. Shetty¹, T.M. Brunetti¹, M. Boorgula¹, M. Campbell¹, P. Maul³, T. Maul³, A.R. Greenidge⁴, D. Walcott⁵, A. Franklin¹, R.C. Landis¹, H. Watson⁵, I. Ruczinski⁶, T.H. Beaty⁶, R.A. Mathias¹, K. Barnes¹. 1) Department of Medicine, University of Colorado, Aurora, CO, USA; 2) Department of Biostatistics, Bloomberg School of Public Health, JHU, Baltimore, MD, USA; 3) University of West Indies at Cavehill, Barbados, WI; 4) Edmund Cohen Laboratory for Vascular Research, The George Alleyne Chronic Disease Research Centre, The University of the West Indies at Cave Hill, Barbados W.I.; 5) Faculty of Medical Sciences, The University of the West Indies, Queen Elizabeth Hospital, Bridgetown, St. Michael, Barbados; 6) Department of Epidemiology, Bloomberg School of Public Health, JHU, Baltimore, MD, USA.

RATIONALE: Asthma is a complex disease partly attributable to genetic factors and environmental triggers. CD4+ T helper cells are known to play a role in the pathogenesis of asthma and are therefore good target tissues for elucidating disease mechanisms. One of the main goals of the Barbados asthma genetic study (BAGS) expression quantitative trait loci (eQTL) analysis is to identify cis and trans-effects of variants identified in the transcriptome of CD4+ T cells from atopic asthmatics. **METHODS:** Paired-end RNA-Seq on RNA from CD4+ T-cells isolated from 84 samples was performed on the Illumina HiSeq4000, using the TruSeq Total RNA Ribozero kit for library preparation. Ten samples were removed from further analysis due to high percentages of ribosomal bases (>20%). Low quality RNA-Seq read removal and adapter trimming were performed using BBDUK. Trimmed reads for all samples were mapped and assembled using the HISAT2 graph aligner and STRINGTIE using the UCSC hg38 reference. Assembled transcripts from all samples were merged using STRINGTIE to create a uniform set of transcripts. Gene/transcript expression tables were created with the Ballgown R package, and genes with median FPKM (fragments per kilobase of transcript per million mapped reads) ≤ 1 across all samples were filtered out. eQTL analysis was then performed on 43 asthmatic cases with Illumina OMNI2.5 GWAS data using a linear model implemented in the MatrixEQTL R package to detect cis and trans eQTLs. FPKM values were log10 transformed, batch and the first 6 principal components derived from GWAS data were included as covariates. **RESULTS:** There are 10,940 genes with median FPKM > 1 in the 43 samples analyzed. eQTL analysis in total identified 275 genes (405 variants) with cis-eQTLs and 137 genes (307 variants) with trans-eQTLs. There were 5 significant trans-eQTLs with false discovery rate (FDR) adjusted $p < 1 \times 10^{-5}$. SNPs rs9272219 and rs9273012 appeared as trans-eQTLs for HLA-DQA1 gene, a locus previously associated with asthma, and were the most significant with $p = 4.9 \times 10^{-17}$ (FDR $p = 1.8 \times 10^{-7}$), while SNPs rs2708377, rs2597974 and rs1669424 were identified as trans-eQTLs for the SIM1OL1 gene with $p = 3.4 \times 10^{-15}$ (FDR $p = 5.0 \times 10^{-6}$). **CONCLUSIONS:** We identified multiple trans-eQTLs that may play a role in asthma pathogenesis. Future work includes incorporating additional subjects into these analyses and identifying eQTLs with effects that differ between atopic asthmatics and non-atopic non-asthmatics.

2225F

Gene co-expression networks are associated with asthma and reveal diverse molecular signatures and disease processes. *K.M. Magnaye¹, M.M. Soliai¹, D.K. Hogarth², E.T. Naureckas², S.R. White², D.L. Nicolae^{1,2,3}, C. Ober¹.* 1) Human Genetics, University of Chicago, Chicago, IL; 2) Medicine, University of Chicago, Chicago, IL; 3) Statistics, University of Chicago, Chicago, IL.

Asthma is a common chronic inflammatory disease of the airways, affecting over 235 million people worldwide. The etiology of asthma is complex, and its diagnosis likely reflects the effects of different molecular pathways. To further dissect the molecular heterogeneity of asthma, we characterized the transcriptome in primary bronchial epithelial cells (BECs) by RNA sequencing from 82 asthmatics and 42 non-asthmatic adults. Technical sources of variation were identified and adjusted using principal components analysis and limma, respectively. Using 16,003 autosomal genes, weighted gene co-expression network analysis (WGCNA) assigned 14,974 genes into 38 co-expression modules (module size ranged from 20-2,924 genes). We tested for associations between the eigengenes in each module and asthma using a linear model, correcting for age, sex, current smoking status, and 3 ancestry PCs. Five of the 38 modules were significantly associated with asthma after adjusting for multiple testing ($p < 1.32 \times 10^{-3}$). The eigengenes for each asthma-associated module were then correlated with clinical phenotypes among the asthmatic cases, and Ingenuity Pathway Analysis (IPA, Qiagen) was used for enrichment studies of genes within each module. One module was positively correlated with both the "Type 2" asthma gene signature ($p = 5.90 \times 10^{-25}$) and fractional exhaled NO ($p = 7.92 \times 10^{-06}$). IL-13 was a significant upstream regulator of genes in this module (by IPA) further suggesting that this module represents the Type 2 asthma endotype. A second module that was negatively associated with asthma was negatively correlated with blood eosinophil count ($p = 1.83 \times 10^{-04}$) and enriched for genes in estrogen-signaling pathways. Two other asthma-associated modules were enriched for genes involved in iCOS-iCOSL signaling and IL-17 signaling but were not associated with any of the measured phenotypes. Overall, these results revealed 5 bronchial epithelial cell molecular signatures that correlated with distinct phenotypes and/or diverse immune and endocrine profiles, potentially reflecting underlying asthma endotypes and discrete mechanistic pathways. Supported by U19 AI095230; KMM and MMS were supported in part by T32GM007197..

2226W

Shared and distinct genetic risk factors for childhood onset and adult onset asthma. *M. Pividori¹, N. Schoettler^{1,2}, D. Nicolae^{1,2,3}, C. Ober², H. Im^{1,2}.* 1) Department of Medicine, The University of Chicago, Chicago, IL; 2) Department of Human Genetics, The University of Chicago, Chicago, IL; 3) Department of Statistics, The University of Chicago, Chicago, IL.

Asthma is the most prevalent chronic respiratory disease worldwide. Childhood onset asthma and adult onset asthma are distinct diseases with respect to sex ratios, triggers of exacerbation, associated co-morbidities, and severity, and potentially also for genetic risk factors. However, it remains unknown as to whether asthma risk loci are associated with age of onset or contribute specifically to childhood onset or adult onset asthma. Our goal was to provide insights into the molecular mechanisms contributing to the different clinical manifestations of asthma that is diagnosed at different ages. Toward that goal, we first determined shared and distinct genetic risk loci for childhood and adult onset asthma to assess the effects of associated loci on the age of onset of asthma. Then we identified candidate genes that may be mediating the effects of associations at shared and age of onset specific asthma loci. Leveraging the data available in the U.K. Biobank, this study is the first GWAS of age of onset in asthma that includes both childhood (9,433 cases, onset prior to 12 years of age) and adult onset asthma (23,395 cases, onset after 25 years), which were compared against participants without an asthma diagnosis (323,376 controls). Individuals with COPD were excluded, and all disease status, age of onset asthma and gender were extracted from self reported questionnaires and hospital records. Stratifying by age of onset revealed 63 independent asthma loci: 17 specific to childhood onset, 3 specific to adult onset, 5 shared but with significantly larger effect sizes in the childhood onset GWAS, and 38 additional shared loci. Among the 63 loci were 32 not reported in previous GWAS. Using a gene-based test, PrediXcan, we identified 120 unique genes associated with asthma, including 33 at 6 childhood onset asthma loci, one gene at an adult onset locus, 76 genes at 19 shared loci, and 10 genes outside the loci identified in the GWASs. Our findings of more childhood onset asthma loci and larger effect sizes of risk alleles in the childhood onset cases are particularly striking given that there were nearly 2.5-times more adult onset than childhood onset cases in this study. Overall, our study indicates that genetic risk for adult onset asthma is largely a subset of risk loci for childhood onset asthma, but with overall smaller effect sizes..

2227T

QTLs detected in an airway smooth muscle cell model of gene-environment interactions identify functional candidates from bronchial responsiveness and contractile response GWAS. E.E. Thompson¹, Q. Dang², B. Mitchell-Handley¹, K. Rajendran², S. Ram-Mohan², J. Solway³, R. Krishnan², C. Ober¹. 1) Human Genetics, The University of Chicago, Chicago, IL; 2) Center for Vascular Biology Research, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA; 3) Department of Medicine, The University of Chicago, Chicago, IL.

Airway hyperresponsiveness (AHR) to inhaled stimulants is a key feature of asthma. We used cell models of AHR to discover functional single nucleotide polymorphisms (SNPs) with direct effects on transcriptional and epigenetic responses to two asthma and AHR-promoting cytokines, IL-13 and/or IL-17, in cultured airway smooth muscle cells (ASMCs) from 71 individuals. We hypothesized that some of these variants will be enriched for small p-values among SNPs included in bronchial responsiveness, contractile responsiveness, and asthma GWAS. Expression and DNA methylation data were obtained using Illumina Human HT-12 v4 arrays and Infinium MethylationEPIC BeadChip, respectively. A GWAS of bronchial responsiveness index (BRI) was performed in 964 Hutterite individuals and a GWAS of contractile response was performed in ASMCs using quantitative measurements of contractile responses to methacholine. We conducted expression quantitative trait locus (e)QTL and methylation (me)QTL mapping using Matrix eQTL in cytokine-treated samples. To maximize our power to identify signals shared across exposures, we used mashr with a local false sign rate (lfsr) <0.01; enrichment testing among GWAS SNPs was performed using GARFIELD. We identified 6,390 eQTLs (± 250 Kb of the transcription start site) and 61,207 meQTLs (± 5 Kb of the nearest CpG) that are either unique to or shared among IL-13, IL-17, IL-13+IL-17, and vehicle-exposed ASMCs. Both eQTLs (3.91 fold enrichment (FE)) and meQTLs (7.55 FE) were over-represented among asthma GWAS SNPs with small p-values ($P < 0.01$), and eQTLs (3.13 FE) and meQTLs (6.94 FE) were over-represented among BRI GWAS SNPs. Genes with eQTLs shared among exposures are enriched for upregulated expression in smooth muscle cells (Human Gene Atlas; $P = 0.006$). To look for overlap between our *in vitro* cellular assay of ASM contraction and bronchial responsiveness in human subjects, we compared the results of our GWAS of contractile response in ASMCs to the results of a GWAS of BRI in Hutterites. We observed 2,476 unique SNPs across all exposures, which are enriched for positive regulation of calcium ion transport (GO:0010524; $P = 0.0028$) and H3K9ac (active enhancer) marks in smooth muscle ($P = 0.04$). These results indicate that IL-13 and IL-17-exposed cultured ASMCs serve as a model for elucidating the function of asthma-associated SNPs identified in GWAS, and for identifying additional candidates that did not meet stringent criteria for genome-wide significance.

2228F

Detection of new genetic variants associated with eosinophil-derived neurotoxin and eosinophil cationic protein levels using bivariate genome-wide association study. R. Verne^{1,2}, P. Margaritte-Jeannin^{1,2}, R. Matran³, F. Zerimech⁴, V. Siroux⁵, M-H. Dizier^{1,2}, R. Nadi^{6,7}, F. Demenais^{1,2}, E. Bouzigon^{1,2}. 1) Inserm, UMR-946, Genetic Variation and Human Diseases Unit, Paris, France; 2) Université Paris Diderot, Paris, France; 3) Université Lille and CHU de Lille, Lille, France; 4) Pôle de Biologie Pathologie Génétique, Laboratoire de Biochimie et Biologie Moléculaire, CHU de Lille, Lille, France; 5) Inserm, CNRS, University Grenoble Alpes, IAB, Grenoble, France; 6) INSERM U1168, VIMA (Aging and Chronic Diseases: Epidemiological and Public Health Approaches), Villejuif, France; 7) Université Versailles St-Quentin-en-Yvelines, UMRS 1168, Montigny-le-Bretonneux, France.

The number of genetic factors identified for asthma remains limited. The study of biological phenotypes involved in the pathophysiological mechanisms of asthma may help understanding the genetic architecture of the disease. To characterize the genetic factors influencing two correlated immunomodulatory enzymes released by eosinophils in the allergic response in asthma: ECP (eosinophil cationic protein) and EDN (eosinophil-derived neurotoxin), we conducted univariate and bivariate genome-wide association (GWA) analyses of these phenotypes in 1,018 subjects from the French Epidemiological study on the Genetics and Environment of Asthma (EGEA) using 1000-Genomes imputed SNPs. The univariate GWA analyses detected two genome-wide significant loci associated with ECP on chromosomes 14q11 (rs7141958, located 36kb downstream of *RNASE3* and 27kb upstream of *RNASE2*, $P = 1.2 \times 10^{-10}$) and 18q21, near *MC4R* (rs8097644, $P = 2.7 \times 10^{-9}$). We also identified one significant locus associated with EDN at 6.5kb downstream of *RNASE2* (rs67049014, $P = 1.1 \times 10^{-11}$). The two 14q11 lead SNPs were not correlated ($r^2 = 0.03$). Overall, the 14q11 EDN-associated SNPs were not associated with ECP and conversely ($P > 0.16$). Moreover, the effect of SNPs detected at the genome-wide significant level did not differ according to the allergy and/or asthma status ($P > 0.12$). The bivariate GWA analysis confirmed the associations with the 14q11 (rs67049014, $P = 3.8 \times 10^{-14}$) and 18q21 regions (rs8097644, $P = 6.9 \times 10^{-9}$) and detected a third region on 1p31 in *AK4* gene (rs12091953, $P = 3.7 \times 10^{-8}$). Fine-mapping of these three regions, based on bivariate conditional analysis, showed that the 14q11 region contained two distinct signals located 76kb apart: the lead SNP rs67049014 (previously associated with EDN) and another SNP, rs76185655 located in the 5' region of *RNASE3*. This latter SNP was in moderate LD with the ECP-associated SNP rs7141958 ($r^2 = 0.5$). Interrogation of eQTL databases showed that the two distinct 14q11 SNPs were strongly associated with expression of *RNASE2* and *RNASE3* in the blood ($P < 10^{-19}$). Functional annotations indicated colocalization of these genetic variants with DNA regulatory elements in blood cells. This study shows that bivariate analysis can increase power to detect new loci and may help refining the associated regions. Further replication of these results are ongoing. Funded: Région Nord-Pas-de-Calais.

2229W

FUT2 variants confer susceptibility to familial otitis media. *R.L.P. Santos-Cortez^{1,2,3}, C.M. Chiong^{3,4,5}, L. Hafren⁶, D.N. Frank⁷, A.P.J. Giese⁸, K.A. Daly⁹, W. Szeremeta¹⁰, M.A. Scholes¹¹, H. Pine¹⁰, N. Friedman¹¹, E.G.d.V. Llanes^{3,5}, T.M. Wine¹¹, M.L.C. Tantoco^{3,5}, J.D. Prager¹¹, J. Kere^{12,13}, S. Riazuddin⁸, A.L. Chan^{3,5}, P.J. Yoon¹¹, E.M. Cutiongco-de la Paz^{14,15}, S.O. Streube¹¹, M.R.T. Reyes-Quintos^{3,4,5,14}, H.A. Jenkins¹, P. Mattila⁶, K.H. Chan^{1,11}, K.L. Mohlke¹⁶, A.F. Ryan¹⁷, S.M. Leal¹⁸, T. Chonmaitree¹⁰, Z.M. Ahmed⁸, M.M. Sale^{19,20,21}*, University of Washington Center for Mendelian Genomics. 1) Department of Otolaryngology, University of Colorado School of Medicine (CUSOM), Aurora, CO; 2) Center for Children's Surgery, Children's Hospital Colorado (CHCO), Aurora, CO; 3) Philippine National Ear Institute, University of the Philippines (UP) Manila - National Institutes of Health (NIH), Manila, Philippines; 4) Newborn Hearing Screening Reference Center, UP Manila - NIH, Manila, Philippines; 5) Department of Otorhinolaryngology, UP College of Medicine - Philippine General Hospital (PGH), Manila, Philippines; 6) Department of Otorhinolaryngology, Head & Neck Surgery, University of Helsinki & Helsinki University Hospital, Finland; 7) Division of Infectious Diseases, Department of Medicine, CUSOM, Aurora, CO; 8) Department of Otorhinolaryngology, Head and Neck Surgery, School of Medicine, University of Maryland, Baltimore, MD; 9) Department of Otolaryngology, Head and Neck Surgery, University of Minnesota, Minneapolis, MN; 10) Department of Otolaryngology, University of Texas Medical Branch, Galveston, TX; 11) Department of Pediatric Otolaryngology, CHCO, Aurora, CO; 12) Folkhälsans forskningscentrum, Biomedicum Helsinki, Helsinki, Finland; 13) Karolinska Institutet, Sweden; 14) UP Manila - NIH, Manila, Philippines; 15) Section of Genetics, Department of Pediatrics, UP College of Medicine - PGH, Manila, Philippines; 16) Department of Genetics, University of North Carolina, Chapel Hill, NC; 17) Division of Otolaryngology, Department of Surgery, University of California San Diego School of Medicine, La Jolla, CA; 18) Center for Statistical Genetics, Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 19) Center for Public Health Genomics, University of Virginia School of Medicine, Charlottesville, VA; 20) Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, VA; 21) Department of Public Health Sciences, University of Virginia, Charlottesville, VA.

Otitis media (OM) is a very common disease in early childhood that can lead to complications such as hearing loss. It remains one of the most common reasons for health care visits and antibiotic use. Despite maximization of public health interventions, the worldwide incidence of OM is increasing, pointing to other risk factors such as genetic predisposition. Previously a synonymous *FUT2* variant that is in linkage disequilibrium with a common *FUT2* variant c.461G>A (p.Trp154*) was identified as a genome-wide-significant protective locus against childhood ear infections. Non-secretor status or loss of cell surface expression of ABO blood group antigens due to homozygosity for the p.Trp154* variant is known to be associated with either risk for autoimmune diseases or protection against viral diarrhea and HIV. To determine the role of *FUT2* in familial OM, DNA samples were obtained from 853 multi-ethnic families and patients with OM. Exome and Sanger sequencing, linkage analysis, Fisher exact and transmission disequilibrium tests (TDT) were performed. The common *FUT2* c.604C>T (p.Arg202*) variant co-segregates with OM in a Filipino pedigree (LOD=4.0), while heterozygosity for a rare variant c.412C>T (p.Arg138Cys) is associated with recurrent/chronic OM in European-American patients (OR=3.3, $p=8.5 \times 10^{-6}$) and US trios (TDT $p=0.005$). A second missense variant c.311C>T (p.Ala104Val) was identified in a European-American OM patient and also co-segregates with OM in a Pakistani family. Moreover 30 middle ear samples from OM patients were submitted for microbiome sequencing and analyses. The p.Trp154* variant was associated with shifts in middle ear microbiota composition (PERMANOVA $p<10^{-7}$) and increased biodiversity. *Fut2* is transiently upregulated after bacterial inoculation of mouse middle ear, with peak levels (10x) at 24 hours post-infection. To further elucidate the functional effects of *FUT2* variants, COS-7 cells were transfected with human GFP-FUT2 variant constructs then GFP expression was determined by flow cytometry. GFP levels were decreased in cells with nonsense variants p.Trp154* and p.Arg202*, suggesting lower expression or quicker degradation of FUT2. On the other hand, expression of A antigen is decreased by all four *FUT2* variants p.Ala104Val, p.Arg138Cys, p.Trp154* and p.Arg202*. In conclusion, common and rare *FUT2* coding variants confer OM risk, potentially by modulating the middle ear microbiome through decreased mucosal expression of A antigen.

2230T

Multiple sclerosis genetic risk score distribution and utility across diverse populations. *A. Beecham^{1,2}, L. Amezcua³, A. Chinea⁴, G. Beecham^{1,2}, S. Delgado⁵, D. Conti⁶, J. Oksenberg⁷, J.L. McCauley^{1,2}*. 1) John P. Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, FL; 2) Dr. John T. Macdonald Department of Human Genetics, Miller School of Medicine, University of Miami, Miami, FL; 3) Department of Neurology, Keck School of Medicine, University of Southern California, Los Angeles, CA; 4) San Juan MS Center, Guaynabo, Puerto Rico; 5) Multiple Sclerosis Division, Department of Neurology, Miller School of Medicine, University of Miami, Miami, FL; 6) Department of Preventative Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA; 7) Department of Neurology, University of California at San Francisco, San Francisco, CA.

Observed differences in multiple sclerosis (MS) prevalence and clinical presentation across populations provide evidence that allele and locus heterogeneity may exist. Currently, 200 non-MHC autosomal risk variants have been identified as conferring risk for MS in >100,000 ancestral Europeans. We computed a genetic risk score of the 200 variants in 1000 Genomes populations of Europe (N=441), Africa (N=323), Asia (N=166), and the Americas (N=258) in order to evaluate the distribution of risk alleles across populations. The mean risk score ranges from 18.8 in Asians to 19.8 in Africans (~13 more risk alleles in Africans), with Europeans and people of the Americas demonstrating intermediate scores of 19.4 and 19.3 respectively. When comparing the mean risk score across the four populations, differences ($p<0.05$) were observed for each of the pairwise combinations except Europeans and people of the Americas; indicating differences in the distribution of MS risk variants across ancestral groups. We further evaluated the genetic risk score in admixed (3 or 2-way European, African, and Native American ancestry) Hispanics (1398 cases; 1386 controls) and African Americans (1305 cases; 1155 controls); in addition to Europeans (1811 cases; 477 controls) for comparison. The mean risk score was 19.4, 20.0 ($p=8.12 \times 10^{-58}$); 19.7, 20.03 ($p=3.51 \times 10^{-23}$); and 19.4, 20.0 ($p=2.88 \times 10^{-39}$) in Hispanic, African American, and European controls, cases respectively. The increase in Hispanic and African American cases versus controls indicates the utility of the risk score in characterizing risk among diverse populations. In each admixed data set, we compared risk scores in the top 5th percentile of the distribution to those in the interquartile range and found that being in the top 5th percentile indicates greater odds of disease in Hispanics than in African Americans (3.16 vs 2.39) despite higher mean risk scores observed in African Americans. Utilizing local ancestry to compute the odds of MS for those with 0, 1, and 2 alleles of European, African, and Native American ancestry (derived from Karitiana, Maya, Pima, and Surui peoples of HGDP) for each risk variant; we find evidence that for some risk loci, haplotype structures differ across populations and the identified risk variant may not be the best tag for the true causal variant in all populations.

2231F

Proteomic profiling of T cells from multiple sclerosis patients and healthy controls. T. Berge^{1,2}, A. Eriksson^{2,3}, I.S. Brorson^{2,3}, E.A. Høgestøl^{2,3}, H.F. Harbo^{2,3}, S.D. Bos^{2,3}, F. Berven⁴. 1) Institute of Mechanical, Electronics and Chemical Engineering, Oslo Metropolitan University, Oslo, Norway; 2) Neuroscience Research Unit, Oslo University Hospital, Oslo, Norway; 3) Institute of Clinical Medicine, University of Oslo, Norway; 4) Proteomics Unit, Department of Biomedicine, University of Bergen, Norway.

Multiple sclerosis (MS) is an autoimmune disorder affecting the central nervous system. It is one of the most common neurological conditions among young adults, leading to both physical and cognitive impairments. The disease develops in genetic susceptible individuals, triggered by common environmental factors such as reduced serum levels of vitamin D, virus infection and smoking. The exact underlying pathogenesis remains unclear, but T lymphocytes, both CD4⁺ and CD8⁺ T cells, have long been considered to play pivotal roles in MS. In addition, the genetic architecture of MS susceptibility emerged from genome-wide association studies, indicates an important role for the adaptive immune system, in particular T cells, for MS. To get an insight into immune-cell processes in MS, we evaluated immune dysregulation at the protein level using liquid chromatography combined with mass-spectrometry analyses. We have analyzed the proteomic profile of purified immune-cell subsets, i.e. CD4⁺ and CD8⁺ T cells, which allows us to disentangle potential cell-subtype specific differences that could not be detected in a heterogeneous cell material such as whole blood, allowing a comprehensive understanding of disease mechanisms. CD4⁺ and CD8⁺ T cells were purified from blood by magnetic separation from 13 treatment-naïve female patients with relapsing-remitting MS and 14 age and gender-matched healthy controls. More than 2000 proteins were identified above detection threshold, of which 195 and 230 proteins were differentially expressed ($p < 0.05$) in CD4⁺ and CD8⁺ T cells respectively. In order to identify the true candidate proteins differentially expressed between MS cases and healthy controls, we used a more stringent filter for selection (i.e. fulfilling two of the following criteria; fold change > 2 , area under the curve > 0.8 , $p < 0.01$). This resulted in 91 and 62 proteins from CD4⁺ and CD8⁺ T cells, respectively, that were analyzed using the Ingenuity Pathway Analysis software. We were also able to identify two MS risk loci, which genotypes showed correlation with protein expression from neighboring genes. It should be highlighted that a larger verification study is needed to replicate our current findings.

2232W

Generation of humanized MHC knock-in mice to study immune-related phenotypes. P. Chen^{1,2,3}, K. Chen¹. 1) Graduate Institute of Medical Genomics and Proteomics, National Taiwan Univ., Taipei, Taiwan, Taiwan; 2) Department of Medical Genetics, National Taiwan University Hospital, Taipei, Taiwan; 3) Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan.

Most immune-related human phenotypes actually lack good animal models. Technology advances (such as recombineering, pronuclear injection-based mouse targeted transgenesis (PITT) and CRISPR/Cas9) in recent years would open a new era of excellent humanized animal models. Previously animal models involved in major histocompatibility complex (MHC) were mostly simple transgenic lines. There are syntenic regions between human and mouse MHC loci/genes, but the correlations are far from perfect. More sophisticated humanized MHC knock-in design holds the promise to shed light on the mechanism of human immune-related diseases. In this study, we aimed at generating mouse lines carrying various *HLA-DRB1* alleles. We started with the embryonic stem cells from the C57BL/6N line, thus future backcrossing to the B6 background will not be necessary. We successfully achieved first targeting; the mouse *H2-Aa* was replaced with human *HLA-DRA* gene, which was retrieved from human BAC RF11-379F19. We then accomplished the second targeting by replacing mouse *H2-Ab1* with the human *HLA-DRB1* gene, which was retrieved from human BAC CTD-2510D15. We are introducing various HLA alleles (such as *HLA-DRB1*03:01* and *HLA-DRB1*08:03*) into the mouse to better model various human phenotypes (such as Graves' disease and antithyroid drug-induced agranulocytosis). A companion humanized CD4 mouse will also be available soon. Humanized MHC mice could provide important help to study immune-related human diseases/phenotypes.

2233T

DNA methylation profiles separate clinical subtypes of Sjögren's syndrome.

C. Chi^{1,2}, H. Quach¹, D. Quach¹, K.E. Taylor², L.F. Barcellos¹, L.A. Criswell³. 1) University of California, Berkeley, Berkeley, CA; 2) Computational Biology Graduate Group, University of California, Berkeley, Berkeley, CA; 3) Russell/Engleman Rheumatology Research Center, University of California, San Francisco.

Sjögren's syndrome (SS [MIM 270150]) is an autoimmune disorder resulting in glandular dysfunction. Clinical heterogeneity of SS presents a challenge in clinical care, therefore it is important to identify disease subtypes and corresponding biomarkers of disease. We performed a genome-wide DNA methylation study of human labial salivary gland biopsy samples obtained from 131 white female subjects from the Sjögren's International Collaborative Clinical Alliance (SICCA; HHSN268201300057C) Registry (67 cases and 64 controls). A total of 402 significant differentially methylated regions associated with SS (DMR-SS) were identified using Bumphunter. Principal component analysis (PCA) of CpGs in the DMR-SS revealed that cases and controls are linearly separable, and a support vector machine (SVM) classifier using the top 10 principal components (PC) was able to achieve 82% ten-fold cross validation accuracy for distinguishing cases from controls. Chromosome 6 accounted for the majority of DMR-SSs and neighborhood genes included *HLA-E* and *TAP*. *TAP* is known to transport MHC class I molecules to form complexes with *HLA-E* for natural killer cell recognition. We next searched for clinical subtypes of SS and their associated DMR (DMR-CS). Since SS is clinically heterogeneous, a large number of phenotypes including antibody serotypes, dry eyes, salivary gland function, and focus score (a quantitative assessment of the severity of focal lymphocytic sialadenitis) were collected for all subjects. PCA together with k-means identified two clusters of SS cases, with one exhibiting severe and the other exhibiting mild phenotypic outcomes. Bumphunter identified 82 significant DMR-CSs associated with severe and mild clinical subtypes, with chromosome 6 accounting for the largest share of DMR-CSs. Using the top 10 PCs from DMR-CS CpGs as predictors, the SVM SS subtype classifier was able to achieve a 92% ten-fold cross validation accuracy. Nearest genes to the DMR-CSs included *HLA-DPA1*, *TAP1*, *HLA-DQB1*, *HLA-DPA1*, *HLA-B*, and *HLA-E*; and pathway analysis with Panther on chromosome 6 genes identified antigen processing and presentation as the most significantly enriched pathway. This work demonstrates that genome-wide DNA methylation can be used to distinguish cases from controls and severe from mild clinical subtypes. Many DMRs are located near immune genes and offers evidence that epigenetic changes affect autoimmunity in SS.

2234F

Whole exome sequencing of more than 16,500 hypothyroidism cases from a clinical cohort implicates a role for protein altering variants in novel and known autoimmunity genes.

J. Freudenberg, C. Schumann, L. Gurski, C. O'Dushlaine, A. Li, G. Della Gatta, J. Horowitz, N. Gosalia, J. Staples, A. Marcketta, X. Bai, J. Penn, L. Habegger, K. Praveen, N. Abul-Husn, T. Teslovich, F. Dewey, O. Gottesman, M. Jones, M. Sleeman, A. Shuldiner, J. Overton, J. Reid, S. Haxhinasto, A. Baras, Geisinger-Regeneron DiscovEHR Collaboration. Regeneron Genetics Center, Tarrytown, NY.

Hypothyroidism (HT) is a very common trait among the elderly. In the present study, we analyzed exome sequencing data from 16,819 HT cases and 63,627 controls from the GHS DiscovEHR cohort. 69,604 individuals also had at least one measurement of thyroid stimulating hormone (TSH). In total more than two millions protein altering variants (PAVs) were ascertained, of which more than 300 thousands with a minor allele count greater than eight were separately evaluated. When variants were tested for association under an additive genetic model, twelve independent non-MHC loci were returned under a significance threshold of $P < 1e-7$. Implicated novel genes include a common missense variant in *TLR3* (rs3775291, L412F, MAF=29.6%, HT:OR=0.92; $P=1.71e-8$, TSH:beta=-0.027; $P=4.70e-6$), which alters a position that is conserved across human *TLR3*, *TLR5*, *TLR7*, *TLR8* and *TLR9* protein sequences. *TLR3* encodes an intracellular endolysosomal dsRNA sensor. Furthermore, we observed novel associations for several rare variants and an infrequent missense variant in *B4GALNT3* (rs149363012, A967V, MAF=1.99% HT:OR=1.12; $P=8.30e-3$; TSH:beta=0.11; $P=1.15e-8$). *B4GALNT3* encodes a Golgi-localized glycosylation enzyme expressed in the thyroid and other tissues. Previously recognized autoimmunity genes with common missense variants displaying significant associations comprise *CTLA4* (rs231775, T17A), *PTPN22* (rs2476601, W620R), *SH2B3* (rs3184504, W262R), and *C1QTNF6* (rs229526, P42R). Additionally, the aggregate burden test of all rare (MAF<1%) PAVs in the known autoimmunity gene *IFIH1* displays evidence for protective effects (HT:OR=0.85; $P=2.8e-6$, TSH:beta=-0.045; $P=1.5e-3$). Interestingly, *IFIH1* also encodes an intracellular dsRNA sensor, suggesting a shared mechanism with *TLR3*. In conclusion, our whole exome sequencing study finds novel and known associations with hypothyroidism both in the common and the rare variant spectrum. The results support autoimmunity as its major determinant in general, and a role for innate immune responses and type I interferon signaling in particular. The results further point to targets for potential interventions in thyroiditis itself as well as functionally related immunological traits.

2235W

Characterizing oxysterols as a moderator of inflammation and mediator of remyelination in multiple sclerosis. S.G. Gregory¹, S.F. Arvai¹, S.K. Siecinski¹, K. Pegram², M. Kanayama³, M.L. Shonihara³, E.J. Benner². 1) Duke University School of Medicine, Department of Neurology, Durham, NC; 2) Duke University School of Medicine, Department of Pediatrics, Durham, NC; 3) Duke University School of Medicine, Department of Immunology, Durham, NC.

Multiple sclerosis (MS) is the most common neurological disease of young adulthood, affecting an estimated 1 million individuals in the U.S. and 2.3 million worldwide. MS is an autoimmune disease mediated, in part, by T-cell components that trigger demyelination and neuronal damage of the central nervous system (CNS), resulting in debilitating neurological symptoms. While disease-modifying therapies have proven to be efficacious, they only prolong remission, they do not change disease course, and the majority of individuals with MS will likely experience worsening of clinical symptoms during the course of their disease. There is a significant gap in our knowledge with respect to curative therapies for MS that modulate autoimmune response and promote remyelination via mechanisms such as differentiation of oligodendrocyte precursor cells into mature oligodendrocytes (ODs) within the adult CNS. We present data that shows treatment with the oxysterol, 20-HC, modulates CD4⁺ activated T Helper 17 (Th17) cell differentiation *and* attenuates the clinical severity of the animal model of MS, experimental autoimmune encephalomyelitis (EAE). We show that 20-HC functions as an agonist of the Liver X Receptor (LXR) pathway and modulates inflammation by preventing polarization in T cells, which suggests that 20-HC blocks the differentiation of Th17 cells *in vitro*. We also establish in a model system that 20-HC treatment triggers remyelination via the differentiation of new ODs from the quiescent pool of oligodendrocyte progenitor cells in the CNS, using *in vivo* and single cell expression data, beyond the limited spontaneous regeneration that occurs during disease course. Our data has significant potential to improve the lives of individuals with MS and establishes the mechanistic foundation for the repair of other neurodegenerative phenotypes in which oxysterol therapy may be efficacious. .

2236T

Epigenomic dissection of multiple auto-immune disorders using H3K-27ac ChIP-seq across 200 individuals. L. Hou^{1,2}, L. He^{1,2}, Y. Li^{1,2}, L. Wang^{1,2}, S. Steelman², C. Nusbaum², M. Kellis^{1,2}. 1) Computer Science and Artificial Intelligence Lab, Massachusetts Institute of Technology, 32 Vassar St, Cambridge, Massachusetts 02139, USA; 2) The Broad Institute of Harvard and MIT, 415 Main Street, Cambridge, Massachusetts 02142, USA.

Understanding the molecular basis of complex traits and human disease requires integration of functional genomics data and genetic information from genome wide association studies (GWAS). An increasing number of genome wide enrichment analyses have revealed that >90% of genetic loci associated with complex traits impact noncoding regions. These enrichments make an implicit assumption that changes in gene regulation underlie complex traits, but these changes have thus far been rarely observed directly. To address this challenge, we directly profiled epigenomic alterations in a large cohort of individuals from the MGH biobank, consisting of patients with Crohn's Disease (CD), Ulcerative Colitis (UC), Rheumatoid Arthritis (RA), and healthy controls, according to their electronic healthy records. We focused on autoimmune disorders, given the readily accessible nature of blood enabling us to observe changes directly in disease-relevant cell types. We carried out ChIP-seq for H3K27ac, a mark associated with activated enhancers, and sought to understand whether changes in enhancer activity are linked to genetic variants associated with disease. Given the diversity of blood cell types, we reasoned that the epigenomic alterations may more manifest strongly in specific immune cell types that may not be visible at the bulk level. To recognize those, we developed a deconvolution pipeline that separates bulk samples into each of 8 components, based on epigenomic profiles for sorted cells from ROAD-MAP and BLUEPRINT epigenomic projects, including memory CD4⁺ T cells, naïve CD4⁺ T cells, memory CD8⁺ T cells, naïve CD8⁺ T cells, monocytes, B cells, Natural killer (NK) cells, and neutrophils. We applied our algorithm to compare individuals between each disease group and healthy group, resulting in differential H3K27ac peaks for each cell type in each disease relative to health controls. Based on Gene Ontologies analysis for genes near differential peaks, we found disease-related terms, including "T helper cell differentiation", "Inflammatory bowel disease (IBD)", "Autophagy", confirming the biological relevance of our findings. Lastly, we used RiVIERA to study the enrichment of GWAS signals in these alterations across traits, and to recognize the most relevant cell types for each. This predict NK cells for CD and RA, memory CD8⁺ T cells for UC are those with epigenomic alterations more strongly enriched with relevant genetic alterations for further investigation.

2237F

Shared underlying genetic susceptibility to Crohn's disease and leprosy in East Asians. S. Jung¹, H.S. Lee¹, M. Hong¹, Y. Li², L. Wang², A. Irwanto², B.D. Ye³, S.K. Yang³, F. Zhang⁴, J.J. Liu², K. Song¹. 1) Department of Biochemistry and Molecular Biology, University of Ulsan College of Medicine, Seoul, Korea; 2) Human Genetics, Genome Institute of Singapore, Singapore; 3) Department of Gastroenterology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea; 4) Shandong Provincial Institute of Dermatology and Venereology, Shandong Academy of Medical Sciences, Jinan, China.

Genome-wide association studies in Chinese patients with leprosy and European patients with Crohn's disease (CD) have suggested that CD and leprosy might share a common underlying genetic susceptibility related to nonspecific innate immunity. Few studies have investigated the relationship between Asian CD and leprosy susceptibility loci at a genome-wide level. The object of the study was to identify shared susceptibility loci to both diseases in East Asians. We first performed meta-analyses of three Immunochips and a GWAS on CD (3,566 cases and 8,885 controls of Korean, Chinese and Japanese ancestry) and three published GWAS on leprosy (2,960 cases and 3,747 controls of Chinese ancestry) and followed up susceptibility loci at genome-wide significance for shared disease association. In addition, we tested established Asian CD risk SNPs for association with leprosy and vice versa. By testing non-HLA 18 CD risk loci for association with leprosy and 19 leprosy risk loci for association with CD, we identified nine shared susceptibility loci: *IL23R* at 1p31.3, *IL18RAP* at 2q12.1, *IL12B* at 5q33.3, *RIPK2* at 8q21.3, *TNFSF15* at 9q32, *ZNF365* at 10q21.2, *CCDC88B* at 11q13.1, *LACC1* at 13q14.1, and *NOD2* at 16q12.1. Of the nine shared loci, only two loci (*LACC1* at 13q14.1 and *RIPK2* at 8q21.3) showed concordance in allelic effects between CD and leprosy. Our study shows that CD and leprosy share about half of their susceptibility genes in East Asians. The majority of shared susceptibility loci between CD and leprosy showed allelic effects in opposite directions, suggesting complex genetic susceptibility between inflammatory diseases and infection.

2238W

Phenome-wide scan of genetic liability of multiple sclerosis using UK Biobank. R.E. Mitchell¹, G. Davey Smith¹, S.J. Sawcer², L. Paternoster¹. 1) Medical Research Council Integrative Epidemiology Unit, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK; 2) University of Cambridge, Department of Clinical Neurosciences, Box 165, Cambridge Biomedical Campus, Hills Road, Cambridge, CB2 0QQ.

Multiple sclerosis (MS) is an inflammatory demyelinating disorder of the central nervous system with diverse manifestations and comorbid conditions. Genome-wide association studies (GWAS) have identified MS as a polygenic disease with a large number of genetic variants, contributing to susceptibility. However, limited evidence explains how genetic risk for MS is manifest in a general population. Therefore, we investigated the association of the genetic variants contributing to the susceptibility of MS with a comprehensive set of phenotypic traits within the UK Biobank cohort. A weighted genetic risk score (GRS) was generated on an individual basis using the latest International Multiple Sclerosis Genetics Consortium (IMSGC) GWAS meta-analysis. Regression analyses were used to assess the association between the GRS and phenotypic traits. This analysis was undertaken using PHESANT (Phenome Scan ANalysis Tool), a software package for performing comprehensive phenome scans in UK Biobank. Due to the major contribution of the human leukocyte antigen (HLA) in the genetics of multiple sclerosis, this analysis has been performed excluding and including genetic variants in the HLA region as well as with the HLA variants alone. The non-HLA GRS is associated 15 traits below a stringent Bonferroni p value threshold with a positive association with hypothyroidism (OR per SD increase in GRS 1.07, 95% CI 1.06-1.09, $p=2.94 \times 10^{-20}$) and eosinophil counts (10^9 cells/Litre) (β per SD increase in GRS 0.018, 95% CI 0.015-0.021, $p=2.49 \times 10^{-29}$). The pheWAS using the GRS including independent variants in the HLA region identified 57 associations below a Bonferroni corrected p value, including several autoimmune diseases. Interestingly, hypothyroidism (OR per SD increase in GRS 0.89, 95% CI 0.88-0.91, $p=4.71 \times 10^{-44}$) and eosinophil counts (10^9 cells/Litre) (β per SD increase in GRS -0.014, 95% CI -0.017 - -0.011, $p=1.55 \times 10^{-17}$) are negatively associated with the HLA GRS. These findings suggest polygenic overlap between genetic variants associated with MS and these phenotypes that are a manifestation of MS genetic risk. Increased understanding of the genetic overlap between different conditions helps to understand the underlying mechanisms of these diseases and may also provide useful information into repositioning and off-target effects of current therapies.

2239T

A comprehensive analysis of 94 SLE loci identifies multiple novel functional SNPs and their target genes. J.E. Molineros¹, B. Singh¹, M. Alasti¹, Y. Uzun², K. Tan², S.K. Nath¹. 1) Oklahoma Medical Research Foundation, Oklahoma City, OK; 2) Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, 19104,.

Systemic Lupus Erythematosus (SLE) is a complex autoimmune disease with a strong genetic component. Multiple genome-wide association and candidate gene studies have identified several susceptibility loci. However, the largest challenge has been to identify causal variants, their cell-specific target genes and associated pathways that explain those association signals. The objectives of this study were to (1) re-assess each of the published association signals; (2) identify functional variants and their target genes within the identified associated regions, and (3) characterize the effect of those genes in SLE involved pathways. We analyzed each of the 308 previously reported genome-wide significant (GWS) lead SNPs and all their correlated ($r^2 > 0.7$ and $D' > 0.9$) variants within 1Mb. We incorporated summary SLE GWAS information as well as cell specific epigenetic information to each annotated variant. We used two recently published methods for functional SNP identification and gene prioritization using epigenetic annotations, cell and tissue-specific eQTLs, and network based analyses. SNPs were considered functional if they were in an exon (CADD score > 12.37), had enhancer (eSNPs) or promoter function, affected expression of a target gene or were inside anchor points for promoter-capture Hi-C assays. We used epigenetic information from blood cell lines to identify target genes. We performed network and submodule analysis to further link functional SNPs to target genes within pathways as well as to identify core target genes. Finally, we validated some candidate eSNPs using Luciferase assays. We curated previously reported SLE GWS SNPs were located in 94 independent loci/regions. We identified associated functional SNPs in 63 of those loci. There were 14 loci that included coding variants (from 45 genes), 13 of which also included eSNPs or promoter linked SNPs. There were 49 additional SLE regions that had eSNPs and SNPs in chromatin interaction anchors. Target genes extended outside associated regions, yielding 416 target genes implicated through eQTLs, eSNPs or promoter-enhancer chromatin interaction targets. Target genes were enriched in interferon I signaling and response to interferon I stimulus pathways. Luciferase assays confirmed the functional potential and differential allele-specific effect of rs10239000 in *IKZF1*. In conclusion, out of 94 independent SLE associated regions, we identified functional SNPs and their cell-specific gene targets at 63 SLE loci.

2240F

HLA-DQA1*05 is associated with the development of antibodies to anti-TNF therapy. A. Sazonovs¹, N.A. Kennedy², L. Moutsianas¹, C. Bews-Hea², G.J. Walker², K. de Lange¹, J. Goodhand², J.C. Barrett¹, T. Ahmad², C.A. Anderson¹, PANTS Investigator Consortium. 1) Human Genetics, Wellcome Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, United Kingdom; 2) The Institute of Biomedical & Clinical Science, Exeter Medical School, University of Exeter, EX1 2LU, United Kingdom.

Anti-TNF therapies are the most common biologic therapies for treating immune-mediated diseases, such as rheumatoid arthritis, psoriasis, and Crohn's disease. Crohn's Disease (CD) is a chronic condition associated with inflammation of the gastrointestinal tract. While there is no cure for it, anti-TNF drugs are effective in treating patients resistant to conventional therapies. However, 10-40% of patients fail to respond to 10-12 weeks of therapy, and a further 25-45% lose response within 12 months. Immunogenicity to anti-TNF therapy is a major cause of loss of response, hypersensitivity reactions, and discontinuation of treatment, and it currently cannot be predicted prior to treatment. A number of factors have been associated with the risk of immunogenicity, but knowledge of the cellular and molecular mechanisms remains limited. Our aim was to investigate genetic susceptibility to immunogenicity. The Personalised Anti-TNF Therapy in Crohn's Disease study is a 3-year prospective UK-wide study of response to anti-TNF drugs. Patients' anti-drug antibodies (ADAs) were serially measured over a period of at least 365 days. Using a Cox proportional hazard model, we carried out a GWAS for time to immunogenicity. We identified a strong association in the HLA region on chromosome 6 (rs2097432; $P = 4.2 \times 10^{-13}$; HR=1.69; 95% CI, 1.48 to 1.94). We imputed the HLA alleles and demonstrated that this signal was driven by a dominant effect at HLA-DQA1*05 ($P = 5.9 \times 10^{-13}$; HR=1.90; 95% CI, 1.60 to 2.25). The effect of HLA-DQA1*05 carriage on immunogenicity is the same for patients treated with infliximab versus adalimumab, on anti-TNF monotherapy versus combination therapy with an immunosuppressant, and for smokers versus non-smokers (heterogeneity of effect $P > 0.05$). We replicated the association in an independent retrospective cohort ($N = 178$, $P = 4.31 \times 10^{-4}$; HR=2.00; 95% CI, 1.36 to 2.96). Around 40% of Europeans carry HLA-DQA1*05 and our data suggest that 95% of these would develop immunogenicity within the first year of infliximab monotherapy treatment (a common anti-TNF treatment regime). Genetic testing prior to treatment might thus allow targeted use of immunomodulatory therapies to deliver more durable, safe and cost-effective anti-TNF therapy. We are currently investigating the impact of HLA-DQA1*05 carriage on other clinical outcomes following anti-TNF therapy – loss of response and adverse drug reactions, and these results will also be presented at the meeting.

2241W

Functional study of Peptidylarginine Deiminase genes in arthritis model mice. A. Suzuki, Y. Kochi, K. Yamamoto. RIKEN, Yokohama, Kanagawa, Japan.

Previously, peptidylarginine deiminase type 4 (PADI4) was identified as a susceptibility gene for Rheumatoid arthritis (RA) by genome-wide association studies. Peptidyl citrulline is a target antigen of anti-citrullinated peptide antibodies (ACPAs), and only PADs (translated protein from PADI genes) can provide peptidyl citrulline via modification of protein substrates. PADI4 is highly expressed in bone marrow, macrophages, neutrophils, and monocytes and PADI2 has overlapped distribution with PADI4 in immune cells. The aim of this study was to investigate the relationship between PADI4 gene and PADI2 gene in the progression of RA. Padi4 $-/-$ DBA1J and wild-type mice were immunized with bovine type II collagen (CII) to develop collagen-induced arthritis (CIA). Expression of various inflammatory cytokines and Padi genes in immune cells was detected by real-time TaqMan assay. Cytokine concentration in sera was measured by enzyme-linked immunosorbent assay. Localization of PAD4 and PAD2 protein was indicated by immunohistochemistry. We also generate Padi2 $-/-$ mice and performed experimental arthritis. We demonstrated that the clinical disease score was significantly decreased in Padi4 $-/-$ mice and Padi4 expression was induced by CII immunization. In Padi4 $-/-$ mice sera, serum anti-type II collagen (CII) IgM, IgG, and inflammatory cytokine levels were also significantly decreased compared with those in wild-type mice sera. Interestingly, Padi2 expression was compensationally induced in CD11b+ cells of Padi4 $-/-$ mice. We also demonstrated that the clinical disease score was significantly decreased in Padi2 $-/-$ CIA mice. It appears that Padi4 and Padi2 enhance collagen-initiated inflammatory responses. Our results revealed that PAD4 affected on expression of various cytokines and also controlled Padi genes.

2242T

Molecular analysis of a skin equivalent tissue culture model system of systemic sclerosis using RNA sequencing, epigenetic assays, histology, and immunoassays. D.M. Toledo¹, M. Huang¹, Y. Wang¹, B.K. Mehta¹, T.A. Wood², A. Smith², Y. Nesbeth², I. Ivanovska³, B.C. Christensen¹, P.A. Pioli¹, J. Garlick², M.L. Whitfield⁴. 1) Molecular & Systems Biology, Dartmouth College, Hanover, NH; 2) Tufts University School of Medicine, Tufts University, Medford, MA; 3) Celdara Medical LLC, Lebanon, NH; 4) Biomedical Data Science, Dartmouth College, Hanover, NH.

Background: The molecular mechanisms of systemic sclerosis (SSc) have been difficult to study outside of patient samples. Mouse models lack key features of the disease, and fibroblast cultures show inconsistent results. We have developed an innovative skin-like tissue of SSc, self-assembled skin equivalents (sSE), where we study fibroblast behavior in a 3D microenvironment with epithelial-dermal crosstalk. Here we investigate the molecular changes in SSc sSE and show that it is molecularly similar to SSc skin biopsies. **Methods:** Isolated fibroblasts from SSc patient skin (SScDF) and normal skin (NDF) were seeded into transwell chambers +/- monocytes. SSc patient plasma and healthy control (HC) plasma were added into the media during the polarization period. Keratinocytes were seeded at 3 weeks for epithelialization, and tissues were harvested after 5 weeks followed by IHC, RNA-seq, DNA methylation, ATAC-seq, and ELISA analysis. **Results:** We created 136 samples of sSE from one SScDF line and one NDF line. H&E staining of SScDF sSE showed increased dermal thickness compared to NDF. Differential expression of SScDF sSE and NDF sSE with autologous plasma and monocytes showed increased expression of the following pathways in SScDF: immune system response, myeloid-mediated immunity/activation, and leukocyte differentiation. The upregulated pathways in NDF sSE showed typical processes involved in epithelial proliferation and cell growth. Differential expression of SScDF sSE +/- monocytes found that tissues with monocytes had increased immune response, immune cell proliferation/activation, and macrophage migration. Samples without monocytes showed upregulated collagen processes, but the strong immune signal was missing. IL6 and IL13 production increased in SScDF supernatant during tissue development. Also, gene expression in SScDF sSE by RNA-seq data showed molecular similarity to human SSc patient skin samples and NDF clustered with HC skin samples. Lastly, 3D sSE tissues and 2D monolayer fibroblast cultures have distinct DNA methylation patterns. **Conclusion:** There is a hierarchy of drivers in the creation of 3D tissues, with the origin of fibroblasts being the biggest modifier of disease morphology and the addition of monocytes being the next biggest factor in developing the immune response. These 3D sSE tissues consistently replicate the molecular pathways found in SSc skin and allow for a controlled model system of SSc to manipulate and test drug therapy responses.

2243F

Pathway-analysis using datasets of GWAS and mRNA expression array identified IFNG as the most significant upstream-regulator in primary biliary cholangitis in the Japanese population. K. Ueno¹, Y. Aiba², O. Gervais³, Y. Kawai¹, Y. Hitomi¹, K. Kojima³, M. Kawashima¹, N. Nishida¹, S. Shimoda¹, M. Nagasaki³, K. Tokunaga¹, M. Nakamura^{2,5}. 1) Department of Human Genetics, Graduate School of Medicine, University of Tokyo, Japan; 2) Clinical Research Center, National Hospital Organization (NHO) Nagasaki Medical Center, Japan; 3) Tohoku Medical Megabank Organization, Tohoku University, Japan; 4) Department of Medicine and Biosystemic Science, Kyushu University Graduate School of Medical Sciences, Fukuoka, Japan; 5) Department of Hepatology, Graduate School of Biomedical Sciences, Nagasaki University, Japan.

Primary Biliary Cholangitis (PBC) is an autoimmune liver disease characterized by chronic inflammation and destruction of intrahepatic small bile ducts. So far, genome-wide association studies (GWAS) and the following pathway-analysis have been conducted worldwide and some disease-pathways such as IL12/JAK-STAT, phosphatidylinositol, and hedgehog signaling have been reported operative. However, disease-pathways and their upstream-regulators are still largely unknown in PBC. In the present study, to better understand the disease-pathways and their upstream-regulator, pathway analysis was performed by Ingenuity Pathway Analysis (IPA) using the dataset 1 of GWAS (1920 PBC cases and 1770 controls performed by Affymetrix ASI and Japonica array) and dataset 2 of mRNA microarray of liver biopsy specimens (15 PBC cases and 5 normal controls performed by Agilent Sure Print G3 Human GE 8x60K v3) in the Japanese population. The dataset 1 included 262 annotated genes derived from 6760 SNPs showing $p\text{-value} < 0.00001$ in GWAS. The dataset 2 included 2293 genes which showed x2 more or less expression as compared to controls in the microarray. There were 41 genes (HLA, IKZF3, PRKCB, etc.) that were overlapped between the two datasets, while there were more than 22 overlapped pathways (antigen presentation, Th1, Th2 pathways, etc.). Furthermore, upstream regulator-analysis by IPA revealed that IFNG is the most significant upstream regulator shared between the significant pathways identified in each dataset. We propose that causal genes identified by GWAS could affect the genes identified by mRNA array via the common upstream-regulator. Our strategy should be useful for identifying upstream-molecular targets in various complex diseases.

2244W

Subphenotype analysis of systemic lupus erythematosus and identification for loci with lupus nephritis predisposition. H. Zhang¹, Y.-F. Wang¹, Y. Zhang², J. Yang¹, Y.L. Lau¹, W. Yang¹. 1) Department of Paediatrics and Adolescent Medicine, Queen Mary Hospital, The University of Hong Kong, Hong Kong SAR, China; 2) Department of Pediatric Surgery, Guangzhou Institute of Pediatrics, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangdong, China.

Systemic lupus erythematosus (SLE) is a chronic autoimmune disease with heterogeneous clinical manifestations. Among the 11 SLE subphenotypes, lupus nephritis (LN) is one of the major causes of end-stage organ failure. Previous studies have suggested higher rates of renal involvement and more severe nephritis in Asians compared to populations of European origin. In this study, we have performed a subphenotype analysis in two Asian SLE cohorts, with a total of 2661 cases and 3588 controls, for which detailed clinical data have been documented. The cumulative genetic risk scores (GRSs) was calculated based on the known susceptibility loci. GRSs for LN cases are significantly higher than non-LN cases. Preliminary analysis also found that several known genes reported as associated with SLE, such as *STAT4*, *SLC15A4* and *TNFSF4*, showed a higher effect size in LN cases than non-LN cases, suggesting different genetic predisposition in patients with and without renal involvement. Identification for novel loci showed promising results. So far, we have selected 44 novel loci with suggestive disease association signal ($P < 1 \times 10^{-3}$) and lupus nephritis predisposition ($P < 1 \times 10^{-4}$), which awaits further confirmation in the replication cohorts.

2245T

Development of a genetic testing method by next-generation sequencing for screening of Thalassemia mutations. Y. Cao¹, S. Ha², C. So², J. Yang², I. Ho², W. Yang¹. 1) Department of Paediatrics and Adolescent Medicine, Queen Mary Hospital, The University of Hong Kong, Hong Kong SAR, China; 2) Department of Pathology, Queen Mary Hospital, The University of Hong Kong, Hong Kong SAR, China.

Thalassemia is a major health threat in Hong Kong due to its high disease prevalence. The current mutation detection methods for thalassemia rely heavily on a combination of different tests that target various types of mutations. These tests are extremely cumbersome and make it impossible to establish efficient screening solutions that are applicable to the general population. Thus, a more efficient, accurate and universal platform is badly needed for population thalassemia screening. To achieve this goal, we first designed a set of probes to capture the entire regions of alpha and beta hemoglobin gene clusters. To secure continuous coverage, intervals between adjacent probes are minimized to no more than 500bps whenever possible, so as to ensure continuous coverage of sequences between these gaps by off-target sequencing. In the next step, next-generation sequencing (NGS) was performed in large volume to ensure in-depth coverage. We have applied the above method on the first batch of 96 samples with known common thalassemia mutations. Data quality check showed that 85.5% of the sequences in the alpha and beta hemoglobin gene loci were covered with an average depth of 1000 folds. The uncovered regions are mainly interspersed repeats (such as Alu sequence) that contribute little to mutation detection. Due to the complex nature of the sequences in the alpha gene cluster, poor mapping in homologous sequences is the major hurdle for variant calling. To increase sensitivity and specificity of mutation detection, currently, we are developing a tailor-made bioinformatics pipeline by taking sequence specificity in each region into consideration. Furthermore, we are trying to define haplotypes associated with founder thalassemia mutations to help aid mutation detection. Preliminary results showed that by making use of a combination of available tools and reference sequences built in house, we are able to detect the breakpoints for Southeast Asia (SEA) deletion with certainty. We are also able to discover and genotype $-a^{3.7}$ and $-a^{4.2}$ deletions with high confidence level through coverage depth analysis. Future work will involve dealing with multiple alignment issues existing within homologous sequences in *HBA1*, *HBA2* and pseudogenes and development of a comprehensive NGS data analysis tool for thalassemia mutation detection.

2246F

Exploration of disease susceptibility HLA alleles for cold medicine-related Stevens-Johnson syndrome with high resolution NGS-based HLA typing. K. Nakatani¹, S. Khor¹, Y. Hitomi¹, M. Ueta², Y. Okudaira³, A. Masuya³, Y. Wada⁴, C. Sotozono⁵, S. Kinoshita², H. Inoko³, K. Tokunaga¹. 1) Department of Human Genetics, The University of Tokyo, Tokyo, Japan; 2) Department of Frontier Medical Science and Technology for Ophthalmology, Kyoto Prefectural University of Medicine, Kyoto, Japan; 3) GenoDive Pharma Inc., Kanagawa, Japan; 4) The Center of Medical Innovation and Translational Research, Graduate school of medicine, Osaka university, Osaka, Japan; 5) Department of Ophthalmology, Kyoto Prefectural University of Medicine, Kyoto, Japan.

[Purpose];Stevens-Johnson syndrome (SJS) and its more severe form, toxic epidermal necrolysis (TEN), are diseases characterized by fever and acute inflammatory vesiculobullous reactions of the skin and mucous membranes, often caused by inciting drugs and/or infectious agents. We previously reported that in the Japanese population, cold medicine-related SJS/TEN was associated with *HLA-A*02:06* and *HLA-B*44:03* using Luminex HLA typing method. In this study, we performed high resolution typing of *HLA-A*, *HLA-B* and *HLA-C* at the 3 or 4-field level in order to elucidate disease related susceptibility HLA alleles at 3 or 4-field level. [Method];High resolution HLA typing of 123 Japanese SJS/TEN patients and 213 Japanese healthy controls were determined by using NXType HLA typing kit (One Lambda) and Ion PGM System (Thermo Fisher Scientific). Genotype determination was performed automatically using Typestream Visual software. For the ambiguous HLA alleles, Sanger sequencing were used to determine the key SNP genotypes and final HLA alleles were finally assigned manually. Finally, we performed the case-control association test by comparing the allele frequency of individual HLA alleles in the patients and controls with the χ^2 -test. [Result□Discussion];Among the HLA alleles in HLA class I genes, *HLA-A*02:06:01* at the 3-field resolution showed significant associations in the patients group ($P = 4.11 \times 10^{-11}$, OR=4.39). *HLA-A*02:06* is previously reported to be associated with cold medicine related SJS/TEN and consistent with previous study, *HLA-A*02:06* is the single 3-field level HLA alleles that exhibited significant association. We are currently in the phase of conducting analysis with more samples, and in this presentation, we will also report new findings utilizing whole genome sequencing data in the HLA region.

2247W

Genome-wide association study of shellfish allergy in Japanese population. K. Saito¹, H. Jia², S. Nogawa¹, K. Kawafune¹, S. Takahashi¹, H. Kato². 1) Genequest Inc. , 5-29-11. Siba, Minato-ku, Tokyo 108-0014, Japan; 2) Laboratory of Health Nutrition, Department of Applied Biological Chemistry, Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1, Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan.

[Background and Objective] Food allergy is caused by an antigen-specific immune reaction to the ingestion of food, and the number of affected patients is increasing, particularly in developed countries. Its diagnosis relies on the detection of immunoglobulin E (IgE) antibodies specific to food allergens using invasive blood and skin tests. In recent years, genome-wide association studies (GWAS) advanced the analysis of single-nucleotide polymorphisms (SNPs) of peanut allergy-related genes commonly occurring in European populations. On the other hand, there has been no report on the shrimp and crab allergies often found in Asian populations, particularly among the Japanese. Therefore, this study conducted a comprehensive search for SNPs related to shrimp and crab allergies using a GWAS of a Japanese population. **[Method]** We administered a Web-based questionnaire to assess the presence or absence of shrimp and crab allergies in 10,000 healthy men and women living in Japan. In addition, respondent DNA was extracted from saliva samples, collected through the direct-to-consumer genetic testing service of Japan, and then subjected to genome-wide genotyping with SNP arrays. A GWAS was performed on the acquired data using PLINK, and statistically significant SNPs were extracted with the genome-wide significance level ($p < 5.0 \times 10^{-8}$). **[Results and Discussion]** In the shrimp allergy group, we were able to identify genetic loci in the region of an antigen-recognition gene (*HLA-DOB*) encoding human leukocyte antigen (HLA), a product of major histocompatibility complex (MHC), and in the region of MHC class II-related genes (*HCG23*, *BTNL2*). On the other hand, in the crab allergy group, although we were able to detect a genetic locus showing significant genome-wide associations, we were unable to elucidate the function of the neighboring gene (*C6orf10*). These findings suggest that the genetic backgrounds of shrimp and crab allergies may differ. In the future, we believe that it will be necessary to confirm the reproducibility of these results and carry out more detailed studies towards their diagnostic application.

2248T

Disease-specific regulation of gene expression in a comparative analysis of juvenile idiopathic arthritis and inflammatory bowel disease. A. Mo¹, UM. Marigorta¹, D. Arafat¹, LHK. Chan², L. Ponder², SR. Jang², J. Prince², S. Kugathasan², S. Prahalad², G. Gibson¹. 1) Georgia Institute of Technology, Atlanta, GA; 2) Emory University School of Medicine and Children's Healthcare of Atlanta, Atlanta, GA.

The genetic and immunological factors that contribute to differences in susceptibility and progression between sub-types of inflammatory and auto-immune diseases continue to be elucidated. Inflammatory bowel disease and juvenile idiopathic arthritis are both clinically heterogeneous and known to be due in part to abnormal regulation of gene activity in diverse immune cell types. Comparative genomic analysis of these conditions is expected to reveal differences in underlying genetic mechanisms of disease. We performed RNA-Seq on whole blood samples from 202 patients with oligoarticular, polyarticular, or systemic JIA, or with Crohn's disease or ulcerative colitis, as well as healthy controls, to characterize differences in gene expression. Gene ontology analysis combined with Blood Transcript Module and Blood Informative Transcript analysis was used to infer immunological differences. Comparative eQTL analysis was used to quantify disease-specific regulation of transcript abundance. A pattern of differentially expressed genes and pathways reveals a gradient of disease spanning from healthy controls to oligoarticular, polyarticular, and systemic JIA, CD, and UC. Transcriptional Risk Scores also provide good discrimination of controls, JIA, and IBD. Most eQTL are found to have similar effects across disease sub-types, but we also identify disease-specific eQTL at loci associated with disease by GWAS.

2249F

A novel locus near *CXCL12* is associated with HIV-related non-Hodgkin lymphoma. C. W. Thorball^{1,2}, T. Oudot³, C. Hammer⁴, D. Costagliola⁵, L. Meyer⁶, I. Theodorou⁷, P.J. McLaren^{7,8}, C. Besson^{9,10,11}, J. Fellay^{1,2}. 1) School of Life Sciences, EPFL, Lausanne, Switzerland; 2) Swiss Institute of Bioinformatics, Lausanne, Switzerland; 3) Centre de génétique moléculaire et chromosomique, GH La Pitié Salpêtrière, Paris, France; 4) Department of Cancer Immunology, Genentech, South San Francisco, CA, USA; 5) Sorbonne Universités, INSERM, UPMC Université Paris 06, Institut Pierre Louis d'épidémiologie et de Santé Publique (IPLESP UMRS 1136), Paris, France; 6) INSERM U1018, Centre de recherche en Épidémiologie et Santé des Populations, Le Kremlin-Bicêtre, France; 7) JC Wilt Infectious Diseases Research Centre, National Microbiology Laboratory, Public Health Agency of Canada, Winnipeg, Canada; 8) Department of Medical Microbiology and Infectious Diseases, University of Manitoba, Winnipeg, Canada; 9) Service d'Hématologie-Oncologie, Centre Hospitalier de Versailles, Le Chesnay, France; 10) Université Versailles Saint Quentin en Yvelines, Université Paris-Saclay, Communauté Paris-Saclay, Paris, France; 11) INSERM U1018, Centre pour la Recherche en Epidémiologie et Santé des Populations (CESP), Equipe Generations et Santé, Gustave Roussy, Villejuif, France.

Background. Human immunodeficiency virus (HIV) infection is associated with a substantially increased risk of developing both Hodgkin lymphoma and non-Hodgkin lymphoma (NHL), with NHL being the most common malignancy. High HIV plasma viral load, low CD4+ T cell counts and absence of antiretroviral treatment (ART) are known predictive factors for NHL, but even in the era of suppressive ART, individuals infected with HIV still remain at increased risk of developing NHL compared to the general population. **Methods.** To search for human genetic determinants of HIV-associated NHL, we performed case-control genome-wide association studies (GWAS) in two cohorts of HIV+ patients of European ancestry (Swiss HIV Cohort Study + French Primo ANRS and ANRS CO16 Lymphovir cohorts). In total data were collected on 206 cases (NHL+/HIV+) and 1652 matched controls (NHL-/HIV+). To reduce the confounding effect of severe immunosuppression due to lack of suppressive ART before 1996, we restricted the analysis to cases with an NHL diagnosis post-2000. After genotype imputation, ~7 Million variants were tested for association with development of NHL using linear mixed models implemented in GCTA (mlma-loco) including sex as a covariate. Results were combined across cohorts using a weighted Z-score-based meta-analysis. Potential physical interactions between variant loci and promoters were assessed through analysis of publicly available promoter capture Hi-C data in GM12878 lymphoblastoid cells. **Results.** After meta-analysis, we identified 2 significantly associated variants; rs7919208 ($p=1.04e-9$) and rs12264354 ($p=2.27e-8$) ($r^2=0.4$). Both variants are on chromosome 10, downstream of *CXCL12*. Analysis of promoter capture Hi-C data revealed a significant interaction between the region of rs7919208 and the active TSS of *CXCL12*, suggesting a possible functional impact of rs7919208 on transcription of *CXCL12*. No variants reported in previous GWAS of NHL in the general population were replicated, even at nominal significance ($p<0.05$), indicating that distinct genetic mechanisms might be at play in HIV-related NHL. **Discussion.** We have identified a novel NHL risk locus in the HIV+ population with a possible link to the chemokine *CXCL12*. *CXCL12* is highly relevant in HIV+ patients, as it is known to modulate infection by X4 tropic HIV strains. Furthermore, increased *CXCL12* mRNA levels has previously been correlated to NHL development in HIV+ kids, underlining its relevance in HIV-related NHL.

2250W

Development and validation of *NUDT15* and *TPMT* genotyping kit based on real-time PCR method for dosing recommendation of thiopurine drugs. W. Lee, H.-K. Kim, D. Chu, S. Ahn, S. Chun, W.-K. Min. Laboratory Medicine, Asan Medical Center, Seoul, Korea.

Background: Thiopurines, widely prescribed anti-cancer and immunosuppressive drug in acute lymphoblastic leukemia and inflammatory bowel disease, have severe myelosuppressive toxicity caused by active thioguanine nucleotide (TGN) metabolites. Thiopurine methyltransferase (TPMT) methylates mercaptopurine and thioguanine, causing an inverse relationship between TPMT activity and concentrations of active TGN metabolites. Nudix hydrolase 15 (*NUDT15*) is hypothesized to dephosphorylate activate TGN metabolites, thus preventing their incorporation into DNA and negatively affecting the cytotoxic effects of thiopurines. Variants in these genes result in decreased activities, are associated with thiopurine-induced side effects. In Korean population, *NUDT15* variant allele frequency is about 10-15%, and *NUDT15* genotyping, in addition to *TPMT* genotyping, could be very informative. **Methods:** We established genotyping kit (SolGent Co., Ltd., Daejeon, Korea) to identify three types of *NUDT15* alleles including *2 or *3 (c.415C>T), *4 (c.416G>A), and *5 (c.52G>A), and three types of *TPMT* alleles, *3B (c.460G>A), *3C (c.719A>G), and *6 (c.539A>T) using hydrolysis probe based real-time PCR method based on differential amplification by allele-specific primers for each variant. Testing was performed in cobas z480 (Roche Diagnostics Ltd. Rotkreuz, Switzerland). In *TPMT* genotyping, *3A allele was designated, when c.460G>A and c.719A>G variants were both detected. We tested one hundred specimens using the real-time PCR method and compared its results with those from Sanger sequencing method for coding region and its flanking regions of two genes. **Results:** All the samples genotyped by the real-time PCR method showed successful amplification patterns for each variant. In comparison with Sanger sequencing method, the real-time PCR method showed 100% concordance in all six targeted variants. There was one case with *TPMT* *1/*38 genotype. Of course, *38 (c.514T>C, p.Ser172Pro; unknown functional status) allele detected by Sanger sequencing method was not detected by real-time PCR method. **Conclusions:** *NUDT15* and *TPMT* genotyping by the real-time PCR method successfully detected all the six targeted variants. The advantage of the real-time PCR method is short turn-around time (TAT), but it could miss rare alleles which are not tested for. A clinical laboratory should choose a method for *NUDT15* and *TPMT* genotyping, considering TAT required and allele frequency in a population.

2251T

Non-HLA celiac disease susceptibility loci are associated with early disease onset and the severity of clinical symptoms. J.X.M. Cerqueira¹, K. Lindfors¹, K. Kurppa², P. Laurikka¹, L. Koskinen³, D. Yohannes³, K. Kaukinen^{1,4}, P. Saavalainen³. 1) Celiac Disease Research Center, Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland; 2) Tampere Center for Child Health Research, University of Tampere and Tampere University Hospital, Tampere, Finland; 3) Research Programs Unit, Immunobiology, Haartman Institute, Department of Medical and Clinical Genetics, University of Helsinki, Helsinki, Finland; 4) Department of Internal Medicine, Tampere University Hospital, Tampere, Finland.

Background: In celiac disease (CeD), dietary gluten leads to the development of small bowel mucosal damage of varying degree and the appearance of serum disease-specific autoantibodies. The clinical picture of CeD is highly heterogeneous. The age of onset varies from early childhood to older age and patients may suffer from multitude of gastrointestinal or extraintestinal symptoms. Furthermore, there are many associated conditions, including Type 1 diabetes (T1D). We assessed whether the 39 non-HLA loci previously associated with CeD susceptibility are implicated in specific CeD phenotypes. **Methods:** The studied population comprised of 625 non-related Finnish CeD patients from whom detailed clinical data had been collected. The FINRISK population (n=1817) was included as a control group. The genotypes of 39 celiac-associated non-HLA SNPs (PMID:20190752) were available for all individuals. Case-control and case-only analysis were performed to investigate their MAF differences across dichotomous demographic, clinical, serological and histological celiac outcomes at diagnosis. After corrections for multiple tests, associations were significant at a 10^{-4} permutation threshold ($P_{\text{EMP2}} < 0.05$). **Results:** *SH2B3/ATXN2* locus increased the risk of concomitant T1D in the case-control analysis (OR=4.39, $P_{\text{EMP2}} = 0.01$), remaining marginally significant in cases (OR~3.6, $P_{\text{EMP2}} = 0.09$). In the case-control analysis, *OLIG3/TNFAIP3* locus increased the risk of anemia (OR~1.70, $P_{\text{EMP2}} < 0.006$) and malabsorption ($P_{\text{EMP2}} = 0.02$), and *ITGA4/UBE2E3* and *IL2/IL21* loci were protective (OR=0.57 and OR=0.71, respectively) to the risk of a more severe small bowel mucosal damage. Risk conferred by *CCR1/CCR3* locus ($P_{\text{EMP2}} < 0.04$) was specific to milder small bowel mucosal damage (OR~1.6), negative serum CeD-autoantibodies (OR~3.1) and CeD-related symptoms in childhood (OR~1.6). *OLIG3/TNFAIP3*, *ITGA4/UBE2E3* and *CCR1/CCR3* loci reached nominal significance ($P < 0.05$) also in the case-only analysis. The Xp22 *TLR7/TLR8* loci was associated (OR~0.30, $P_{\text{EMP2}} < 0.04$) with CeD onset before ≤ 7 years of age, strongly reinforced in cases and with sex-dependent effects in the female strata (Breslow-Day test, $P < 0.001$). **Conclusions:** This study provides novel evidence that distinct genome-wide non-HLA variants modulate specific CeD phenotypes. Given the established roles of *SH2B3* and *TLR7/TLR8* genes in bacterial/virus infections, it is plausible that environmental factors, including infections, also modulate CeD phenotype.

2252F

Meta-analysis of genome-wide association studies for inflammatory bowel diseases in Korean population. B.M. Kim¹, B.D. Ye², S. Jung¹, J.W. Moon¹, J.W. Baek¹, M. Hong¹, H.S. Lee¹, S.H. Park², S.K. Yang², K. Song¹. 1) Department of Biochemistry and Molecular Biology, Ulsan University, Seoul, Korea, Republic of; 2) Department of Gastroenterology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea.

Although genome-wide association studies (GWAS) have identified over 200 loci associated with IBD, studies in non-European populations are limited. To identify novel IBD susceptibility loci in Korean population, we performed meta-analysis comprising 3,481 IBD cases (1,967 Crohn's disease [CD], 1,514 Ulcerative colitis [UC]) and 5,166 controls. The three IBD datasets used include previously reported Korean Immunochip v1.0 data (672 IBD [330 CD, 342 UC] and 898 controls), Korean GWAS results (1,467 IBD [896 CD, 571 UC] and 3,783 controls), unpublished Korean Immunochip v2.0 data (1,342 IBD [741 CD, 601 UC] and 485 controls). Meta-analysis of the three datasets (888,449 common SNPs) did not reveal any novel IBD susceptibility loci at genome-wide significance. Of the previously reported 276 Caucasian IBD susceptibility SNPs, we replicated associations of 12 IBD loci, 14 CD loci, and 7 UC loci at genome-wide significance. rs59789950 in *CD28* at 2q33 (odds ratio = 1.21, 95% confidence interval = 1.14-1.30, $P_{\text{meta}} = 1.14 \times 10^{-8}$) showed association with IBD at genome-wide significance for the first time in an Asian population. Additional 12 IBD loci, 10 CD loci, and 8 UC loci were replicated in Koreans beyond Bonferroni-adjusted significance threshold of $P = 1.81 \times 10^{-4}$ (0.05/276). A further 52 known IBD loci, 40 CD loci, and 39 UC loci showed associations at $P < 0.05$. Although our study does not identify any novel IBD susceptible loci, successful replication of previously reported loci supports the shared genetic characteristics of IBD across populations.

2253W

Genetic analysis of inflammatory markers reveals novel genetic associations and highlights the power of multivariate GWAS. J. Partanen¹, S. Ruotsalainen¹, P. Häppölä¹, J. Nunez-Fontarnau¹, V. Salomaa², M. Salmi^{3,4}, S. Jalkanen^{3,4}, P. Palta¹, M. Pirinen^{1,5,6}, J. Koskela¹, S. Ripatti^{1,5,7}. 1) Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland; 2) National Institute for Health and Welfare, Helsinki, Finland; 3) MediCity Research Laboratory, University of Turku, Turku, Finland; 4) Institute of Biomedicine, University of Turku, Turku, Finland; 5) Department of Public Health, Clinicum, Faculty of Medicine, University of Helsinki, Helsinki, Finland; 6) Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland; 7) Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA.

Introduction: Genetic association studies are predominantly performed by univariate tests between genotype-phenotype pairs. However, multivariate methods combining multiple response variables can reveal associations missed in univariate tests, especially related to more complex biological processes. We illustrate the increased power of multivariate association testing using metaCCA software which implements canonical correlation analysis for multiple sets of univariate GWAS summary statistics. **Materials and Methods:** We studied genetic associations of 60 quantitative measurements of cardiovascular or immunologic relevance (Lim et al. 2014 PLoS Genetics). We first performed separate genome-wide associations studies (GWAS) for 24725 Finnish samples imputed with a population-specific reference panel of 2690 individuals, resulting in 11.7 million variants. Thereafter we performed metaCCA analysis on the summary statistics from the GWASes for phenotypically correlated clusters. **Results:** Within the 60 measures, we identified 5 phenotypically correlated clusters. Here we focus on one cluster comprising mainly inflammatory markers: 12 cytokines and growth factors (5 interleukins [IL-4, IL-6, IL-10, IL-12_P70, IL-17], 1 interferon [IFN γ], 1 chemokine [SDF 1A], 1 colony stimulating factor [G-CSF], 1 tumor necrosis factor [TNF β] and 3 growth factors [FGF, PDGF BB, VEGF]). 6404 samples had complete data on the cluster measures. Within the cluster we identified 50 genome-wide significant loci by metaCCA ($p < 5 \times 10^{-8}$, p -values are adjusted for the number of traits). We replicated 15 previously reported associations found in GWAS Catalog ($p < 1 \times 10^{-5}$) within a 500kB distance from our top hits. Additionally, we identified 35 novel associations of which 12 have been previously associated with other cytokines or growth factors and an additional 14 with other immunological traits such as autoimmune diseases. The 12 univariate GWASes on the cluster traits revealed only 9 independent genome-wide significant loci, all of which were detected by metaCCA. The results are available at: http://35.195.196.173/pheno/NORM_CLUSTER. We are currently analyzing the four remaining clusters. **Conclusions:** By analyzing phenotypically correlated traits together we dramatically increase the number of genome-wide significant loci. When analyzed together with disease endpoints, these results can aid in the discovery of disease-related associations by drawing evidence from biologically relevant traits.

2254T

Molecular prediction in inflammatory bowel disease. R. Liu^{1,2}, H. Huang^{1,2}, D. McGovern³, M. Daly^{1,2}. 1) The Broad Institute, Cambridge, MA; 2) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston MA; 3) F. Widjaja Family Foundation Inflammatory Bowel and Immunobiology Research Institute, Cedars-Sinai Medical Center, Los Angeles, CA.

The inflammatory bowel diseases (IBD) are polygenic autoimmune disorders. Two major types of IBD are Ulcerative colitis (UC) and Crohn's disease (CD). Due to the heterogeneous presentation of IBD, diagnosis can be challenging in up to 20% of patients using the conventional classification. Diagnostic uncertainty and subsequent misclassification in IBD are associated with higher rates of complicated disease and relapse. Here we propose a molecular-based risk model as a novel clinical tool to discriminate CD from UC by aggregating genetics, serology and tobacco smoking information. We performed GWAS on the International IBD Genetics Consortium (IIBDGC) Immunochip cohort using phenotypes coded as CD (20,124) v.s. UC (15,160). After LD-clumping using the study cohort as reference, 7,816 SNPs remained and were used to build the genetic component of the model. We tested this model using a) 5-fold cross-validation on the IIBDGC data, and b) model trained on IIBDGC but tested on the CEDAR cohort (an independent cohort with 2,044 CD and 3,462 UC patients). We achieved an area under the curve (AUC) of 0.80 and 0.78 for discriminating CD from UC respectively. We next assessed the ability of serologies (ASCA (IgA/IgG), ANCA, anti-Cbir1, anti-OMPC and anti-I2) and smoking status (never, ex-smoker and current) to discriminate IBD subtype in the CEDAR cohort. The serology-only model and smoking-only model had AUC=0.82 and 0.53 respectively. Serology alone showed the best AUC but is uninformative in more than 10% of patients. Importantly genetics performs well when serology is uninformative (AUC=0.72), establishing that genetics and serology provide independent information. Furthermore, combining with genetic, serologic, and smoking status, we developed overall risk models designed to discriminate CD from UC with AUC=0.88. In conclusion, our analysis clearly suggests that molecular-based model combining genetics, serology and smoking information could complement current diagnostic strategies and help classify patients based on biologic state rather than imperfect clinical parameters.

2255F

Phenome-wide association study (PheWAS) for sex biases in Human Leukocyte Antigen (HLA) associations. J.H. Karnes¹, L. Bastarache², C.M. Shaffer², J.D. Mosley², S. Mallal², D.M. Roden², E.J. Phillips^{2,3}, J.C. Denny². 1) University of Arizona, Tucson, AZ; 2) Vanderbilt University Medical Center, Nashville, TN; 3) Institute of Immunology & Infectious Diseases, Murdoch University, Murdoch, Western Australia, Australia.

Human leukocyte antigen (HLA) genes have more phenotypic associations than any other locus in the human genome. However, recent evidence suggest sex bias in major histocompatibility complex (MHC)-associated shaping of the adaptive immune system, suggesting that reanalysis of HLA association studies by sex is required. We set out to investigate the interaction of sex with HLA associations across a comprehensive set of phenotypes. We performed phenome-wide association studies (PheWAS) for HLA genomic variation by sex and tested HLA*sex interactions using BioVU, a de-identified electronic health records coupled to a DNA biorepository. Using SNP2HLA, we imputed two and four digit HLA alleles from a population of 29,712 European ancestry individuals genotyped on the Illumina HumanExome BeadChip. We tested association of HLA alleles by sex with 1,545 phenotypes derived from EHRs. Significance was achieved if an HLA allele was associated in either males or females at $p < 1 \times 10^{-5}$ and a significant HLA*sex interaction was observed at $p < 0.05$. A total of 84 HLA allele-phenotype associations showed significant HLA*sex interactions. The strongest interaction observed was between *HLA-B*27:05* and 'Other inflammatory spondylopathies' (interaction $p = 3.3 \times 10^{-6}$), which was significant in men (odds ratio 7.97, $p = 4.8 \times 10^{-21}$) but not women (odds ratio 1.58, $p = 0.09$). Phenotypes associated with HLA alleles in men but not women included senile dementia (*DPB1*03:01*), vitamin D deficiency (*B*07*), noninfectious gastroenteritis (*DRB1*04:01*), diabetic retinopathy (*DQB1*02*), viral infection (*B*27*), Graves' disease (*DPB1*17:01*), emphysema (*DRB1*13:02*), and allergic rhinitis (*A*26:01*). Phenotypes associated with HLA alleles in women but not men included psoriasis (*C*06:02*), cardiac arrest (*A*11:01*), chronic liver disease and cirrhosis (*B*14:01*), reticulosarcoma (*DPB1*11:01*), inflammatory bowel disease (*DRB1*01:03*), and ulcerative colitis (*DRB1*01:03*). Several autoimmune phenotypes showed evidence of strong association in both sexes with markedly different odds ratios in males and females, including rheumatoid arthritis (*DQB1*06*), lupus (*DRB1*03:01*), and type 1 diabetes (*DQB1*06:02*). We provide further evidence of a sex bias in MHC-associated influences on the adaptive immune system and further emphasize the need to interrogate sex differences in HLA association studies.

2256W

Co-evolution between *Mycobacterium* and humans influences tuberculosis severity. C.M. Stein^{1,2,3}, M.L. McHenry¹, J. Bartlett⁴, K. Fluegge¹, R.P. Igo, Jr.¹, N.B. Hall¹, H. Mayanja-Kizza^{2,4}, C. Wejse⁵, G. Sirugo⁶, W.H. Boom^{2,3}, S.M. Williams¹. 1) Population & Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH, USA; 2) Medicine, Case Western Reserve University, Cleveland, OH, USA; 3) Uganda-CWRU Research Collaboration; 4) School of Medicine, Makerere Univ., Kampala, Uganda; 5) Department of Infectious Diseases/Center for Global Health, Aarhus Univ., Aarhus, Denmark; 6) Department of Systems Pharmacology and Translational Therapeutics, University of Pennsylvania, Philadelphia, PA, USA.

Susceptibility to tuberculosis (TB) is affected by both host (human) and pathogen (*Mycobacterium tuberculosis*, MTB) genetic variation. For those who develop active TB, there are also likely genetic factors that affect disease severity. Although genetic factors in each of the two species may independently affect TB severity, we have examined the possibility that genetic variants in the human and pathogen have co-evolved such that interactions between MTB lineage and human genetic variation reduce disease severity. We tested this hypothesis using TB cases ascertained through our household contact study in Kampala, Uganda in which we have identified three major MTB lineages, determined using previously characterized MTB markers. Assessing the role of MTB lineage in the context of human genotypes offered a unique opportunity to examine potential host-pathogen interactions in disease severity. First, we tested the association between single nucleotide polymorphisms (SNPs) in 29 candidate genes in the TNF, TLR/NLR, and IFNG/IL12 pathways and disease severity, using the TBscore developed by Wejse. Next, we modeled TB severity as a function of MTB lineage, and interactions between lineage and host genotype. In this case-only study, we analyzed 2 independent cohorts of TB cases with paired human and MTB DNA samples. In the human-only analyses, we observed association between multiple SNPs in *IL12B* and TBscore that replicated in both datasets (most significant SNP rs3212227, $p = 0.007$ and $p = 0.03$, combined $p = 0.0005$), supporting our previous association with TB in a genome-wide association study. With respect to interaction between the human genotypes and MTB lineage, we observed a significant interaction between SNPs in *SLC11A1* and Uganda lineage in both cohorts (rs17235409, $p = 0.003$ and $p = 0.04$, combined $p = 0.0002$). We observed a similar interaction between *STAT1* and Uganda lineage in both cohorts (rs11305, $p = 0.036$ and $p = 0.024$ respectively, combined $p = 0.002$). Importantly, there was no association between TBscore and these genes in the absence of the interaction term. These findings demonstrate that *IL12B* is associated with severity of TB and that the severity of TB disease is modified by MTB lineage and its interaction with at least two genes, providing evidence that host-pathogen co-evolution exists in TB.

2257T

Using *in vivo* eQTL interactions to identify the genetic drivers of variation in the transcriptomic response to sepsis. E.E. Davenport^{1,2}, K.L. Burnham³, C. Goh³, J. Radhakrishnan³, P. Hutton⁴, T.C. Mills³, A. Rautanen³, A.C. Gordon⁵, N. Soranzo⁶, A.S.V. Hill⁶, C.J. Hinds⁷, S. Raychaudhuri^{1,2,8}, J.C. Knight⁸. 1) Division of Genetics and Rheumatology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; 2) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA; 3) Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK; 4) Adult Intensive Care Unit, John Radcliffe Hospital, Oxford, UK; 5) Section of Anaesthetics, Pain Medicine and Intensive Care, Imperial College, London, UK; 6) Wellcome Trust Sanger Institute, Cambridge, UK; 7) William Harvey Research Institute, Barts and The London School of Medicine, Queen Mary University, London, UK; 8) Faculty of Medical and Human Sciences, University of Manchester, Manchester, UK.

Sepsis is defined as organ dysfunction caused by a dysregulated immune response to infection. The clinical diagnosis identifies a heterogeneous population of patients with substantial individual variation in their response, making this a challenging disease to diagnose and treat. We have previously shown that Sepsis Response Signature (SRS) groups are the major predictor of transcriptomic variation in sepsis (Davenport et al., *Lancet Respir Med* 2016) even accounting for different infection sources (Burnham et al., *Am J Respir Crit Care Med* 2017). Specifically the SRS1 group identifies individuals with an immunosuppressed phenotype and a poorer outcome. The contribution of genetic variation to heterogeneity in gene expression can be determined through expression quantitative trait locus (eQTL) mapping. We hypothesized that some eQTL effects are modulated by SRS group resulting in eQTL interactions. Such interactions would indicate specific regulatory variants that are functionally dependent on SRS group, and could point to regulatory drivers of SRS group. There is now increasing interest in eQTL interactions to understand a wide range of complex medical phenotypes. We recruited 468 adult patients with sepsis due to community acquired pneumonia or fecal peritonitis. We profiled gene expression (microarray, Illumina HT12-v4) from 592 peripheral blood leukocyte samples collected within five days of ICU admission and generated genotyping data (Illumina Omni-Express and Core-exome). We first mapped *cis* eQTLs genome-wide using a linear mixed model and identified 2,152 probes with a Bonferroni corrected eQTL $p < 5.2 \times 10^{-8}$. We used the most significant SNP for each of these probes to identify *in vivo* eQTL interactions with SRS group. We observed that 279 eQTLs (>10% of all eQTLs) interact with SRS status (FDR < 0.05). SRS-eQTL interactions include transcription factors such as *IRF5* (FDR = 1.3×10^{-6}) and *EPAS1* (FDR = 7.1×10^{-4}). *IRF5* has a key role in antiviral and inflammatory responses and has an eQTL effect that is dampened in the SRS1 group. *EPAS1* is a hypoxia inducible transcription factor. From network analysis of differential expression between SRS groups, this is a key nodal gene that is both upregulated and has a magnified eQTL effect in the SRS1 group. These eQTL interactions with SRS group provide additional insight into regulatory mechanisms underlying disease and improve our understanding of how genetic variants are driving the individual response to sepsis.

2258F

Genome-wide meta-analysis and functional analysis identified *POGLUT1* as the effector gene driven by rs2293370 in primary biliary cholangitis (PBC) susceptibility locus chromosome 3q13.33 in the Japanese population. Y. Hitomi^{1,9}, K. Ueno^{1,9}, Y. Kawai¹, N. Nishida², K. Kojima^{3,4}, M. Kawashima⁵, Y. Aiba⁶, M. Nagasaki^{3,4,7}, K. Tokunaga¹, M. Nakamura⁸. 1) Department of Human Genetics, Graduate School of Medicine, the University of Tokyo, Tokyo, Japan; 2) The Research Center for Hepatitis and Immunology, National Center for Global Health and Medicine, Ichikawa, Japan; 3) Department of Integrative Genomics, Tohoku Medical Megabank Organization, Tohoku University, Sendai, Japan; 4) Graduate School of Medicine, Tohoku University, Sendai, Japan; 5) Japan Science and Technology Agency (JST), Tokyo, Japan; 6) Clinical Research Center, National Hospital Organization (NHO) Nagasaki Medical Center, Omura, Japan; 7) Graduate School of Information Sciences, Tohoku University, Sendai, Japan; 8) Headquarters of PBC-GWAS study group in Japan, Clinical Research Center, NHO Nagasaki Medical Center, Omura, Japan; 9) These authors contributed equally.

Primary biliary cholangitis (PBC) is a chronic and cholestatic autoimmune liver disease caused by the destruction of intrahepatic small bile ducts. The higher monozygotic/dizygotic (MZ/DZ) ratio and the higher estimated relative sibling risk (λ_s) in PBC patients as compared to unaffected individuals indicates the involvement of strong genetic factors in the development of PBC. Our previous genome-wide association study (GWAS) identified *HLA*, *TNFSF15*, and four other loci as susceptibility loci for PBC in the Japanese population. Here, in order to further elucidate the genetic architecture of PBC in the Japanese population, a GWAS was performed on an additional independent sample set, then a genome-wide meta-analysis with our previous GWAS was performed based on a whole-genome single nucleotide polymorphism (SNP) imputation analysis of a total of 4,045 Japanese individuals (2,060 PBC cases and 1,985 healthy controls). A susceptibility locus on chromosome 3q13.33 (including *ARHGAP31*, *TMEM39A*, *POGLUT1*, *TIMMDC1*, and *CD80*) was previously identified both in the European and Chinese populations and was replicated in the Japanese population (rs57271503: odds ratio [OR] = 0.7241, $P = 3.5 \times 10^{-9}$). Subsequent *in silico* and *in vitro* functional analyses identified rs2293370, previously reported as the top-hit SNP in this locus in the European population, as the primary functional SNP. Moreover, e-QTL analysis in various organs, including pancreas, thyroid and whole blood, indicated that the effector gene of rs2293370 was *Protein O-Glucosyltransferase 1* (*POGLUT1*) ($P < 0.0005$ after Bonferroni multiple comparison correction). This is the first study to demonstrate that *POGLUT1*, and not *CD80* which encodes a well-known co-stimulatory signaling molecule necessary for antigen presentation from HLA class II to T cell receptor (TCR), is the effector gene regulated by the primary functional SNP rs2293370, and that increased expression of *POGLUT1* might be involved in the pathogenesis of PBC.

2259W

The eQTL analysis on blood tissues of a cohort of Korean Crohn's disease patients. B.D. Ye¹, W. Liu², S. Jung³, J. Baek³, J. Moon³, M. Hong³, H.S. Lee³, S.K. Yang¹, J.J. Liu², K. Song³. 1) Department of Gastroenterology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea; 2) Human Genetics, Genome Institute of Singapore, Singapore; 3) Department of Biochemistry and Molecular Biology, University of Ulsan College of Medicine, Seoul, Korea.

Recently, eQTLs has transformed the field of human genetics by providing a comprehensive, easily accessible and interpretable molecular link between genetic variation and organismal phenotypes. Gene expression levels vary among populations presumably due to differences in genotype frequencies, suggesting that population-specific eQTL effects exist. However, non-European ancestries are under-represented in eQTL database. We performed eQTL analysis using RNA-seq of whole blood in a cohort of 101 Korean Crohn's disease (CD) patients. After alignment by STAR, we used RSEM to estimate the transcript abundance, the expected read counts, and fragments per kilobase per million reads (FPKM) for each gene. After removing the low-expressed genes with at least 25% samples not expressed, 10,928 high-expressed genes are obtained. Using 6,633,636 SNPs in genome-wide association study (GWAS) data of 101 samples, eQTL analysis is implemented by matrixQTL, which is computationally efficient for both cis- and trans-eQTLs. The linear model is adopted, covariates including gender, age, age of diagnosis, family history, smoking status, disease phenotype based on Montreal classification. We found 87,086 cis-eQTLs with 74,383 eSNPs on 1,439 high-expressed genes with FDR less than 0.05. The significantly enriched pathways of these 1,439 eGenes include adaptive immune system, interferon signaling, endocytosis, cytokine signaling in immune system, etc. Following the eQTL mapping of 4,727 genome-wide significant SNPs from an independent IBD GWAS cohort, 1,017 SNPs has 1,507 cis-eQTLs on 10 eGenes including *HLA-DQB1*, *HLA-DQA1*, *HLA-DRB1*, *HLA-DRB5*, *HLA-DPB1*, *HLA-C*, *HLA-H*, *MAPK9*, *TUBB2A*, *PLCL1*. We also found some overlap between our eQTLs and the blood eQTL database on a very large cohort. Korean CD eQTL datasets will provide a valuable resource for link between genetic variation and gene expression and regulation in Asians.

2260T

Host genetic variation for susceptibility of ascending *chlamydia* genital tract infection. W. Zhong¹, C. O'Connell², H. Wiesenfeld^{3,6}, S. Hillier^{4,6}, H. Lee^{5,6}, W. Geisler⁷, G. Anyalechi⁸, R. Kirkcaldy⁹, Y. Li^{1,3}, T. Darville², X. Zheng².

1) Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Department of Pediatrics, University of North Carolina at Chapel Hill, Chapel Hill, NC; 3) Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC; 4) Department of Obstetrics, Gynecology, and Reproductive Sciences, University of Pittsburgh, Pittsburgh, PA; 5) Department of Epidemiology, University of Pittsburgh, Pittsburgh, PA; 6) Magee-Womens Research Institute, Pittsburgh, PA; 7) Departments of Medicine and Epidemiology, University of Alabama at Birmingham, Birmingham, AL; 8) US Centers for Disease Control and Prevention, International Research and Programs Branch, Atlanta, GA.

Background and objectives *Chlamydia trachomatis* is the most commonly reported infectious disease in the US. Untreated infection can ascend to the uterus and fallopian tubes resulting in pelvic inflammatory disease (PID), which can lead to infertility. However, the diagnosis of upper genital tract infection is challenging, due to lack of symptoms in most women. The objective of this study is to determine DNA determinants and/or markers for women with ascending *chlamydia* genital tract infection as a target population for vaccine by integrated DNA and mRNA analysis. **Methods** We profiled blood mRNA transcriptome of 212 genotyped women (uninfected n=51, cervical only infected n=82, endometrial infected n=79) from two cohorts, ACE and TRAC, using expression microarray. We mapped expression quantitative trait loci (eQTLs) across the whole cohort, and response eQTLs (reQTLs) with variation of gene expressions in responses to *chlamydia* infection respectively. We tested eQTL effects in ascending infection and infertility in two independent cohorts, a PID Evaluation and Clinical Health (PEACH) longitudinal study, and a *chlamydia* associated tubal factor infertility Case-Control Study by Centers for Disease Control and Prevention (CDC). **Results** We identified 89,368 cis-eQTLs (<1 Mb) in 3,660 expressed genes (eGenes) at a 5% FDR ($P < 7.89 \times 10^{-3}$). Totally 256 cis-eQTLs in 104 genes and 271 cis-reQTLs in 102 genes were nominally associated with infertility in both the PEACH and TIF cohorts ($P < 0.05$). Association of these loci with ascension infection and infertility, suggest plausible roles of these genes in mediating variant effects on *chlamydia* induced ascension and long term sequelae, infertility. Among 256 cis-eQTLs and 271 cis-reQTLs, 71 loci in 18 genes were overlapped. The effects of these 71 loci in uninfected and infected were in the same direction, but of significantly different magnitudes. We also found 6,740 trans-eQTLs (>1 Mb) in 445 eGenes and 5,618 trans-reQTLs in 233 eGenes. **Conclusions** We discovered candidate eQTLs as markers for susceptibility of *chlamydia* ascending infection.

2261F

The genetic architecture of familial vitiligo. *G.H.L. Roberts¹, Y. Jin^{1,2}, S. Santorico^{1,2}, R. Spritz^{1,2}.* 1) Human Medical Genetics and Genomics, University of Colorado, Anschutz Medical Campus, Aurora, CO; 2) Department of Pediatrics, University of Colorado, Anschutz Medical Campus, Aurora, CO; 3) Mathematical and Statistical Sciences, University of Colorado, Denver, CO.

Vitiligo is a common autoimmune disease in which patches of depigmented skin and hair result from destruction of skin melanocytes. Vitiligo is complex, involving multiple genes and environmental triggers. In three GWAS of vitiligo in European-derived Caucasians (EUR), we have identified 50 vitiligo susceptibility loci. Analyzing vitiligo age-of-onset as a secondary phenotype, we found that vitiligo is comprised of two age-of-onset subtypes, early-onset (mean 10.9 years) and late-onset (mean 34.0 years), with distinct underlying genetic architectures. ~20% of vitiligo patients report one or more affected first-degree relatives. We considered that such cases, belonging to 'multiplex families', may likewise constitute a latent vitiligo subtype with different genetic architecture from sporadic cases. We considered three hypothetical genetic architectures in multiplex families: (1) a polygenic architecture, in which common risk variants from the general population drifted to high frequency in the family, conferring higher-than-average risk to close relatives; (2) a mono- or oligogenic architecture, involving segregation of a highly-penetrant variant that is rare in the general population; or (3) a combination of 1 and 2. To address hypothesis 1, we constructed a genetic risk score based on the most-associated variant at each of the 50 GWAS loci, with ORs ranging from 1.15 to 2 and MAFs from 3% to 50%, plus one major histocompatibility complex (MHC) class II high-effect haplotype with OR 7.8 and MAF 3%. The risk score yields a mean difference (Δ) of 1.02 standard deviations comparing all vitiligo cases to controls (AUC=0.77). We compared risk score in one proband from 373 EUR multiplex vitiligo families to 4,069 unrelated sporadic EUR vitiligo cases. Mean risk score is significantly higher in multiplex probands than in singleton cases (Δ =0.16 standard deviations; P =0.004). The difference remains significant after removing the high-effect MHC haplotype from the risk score. Among multiplex probands, higher risk score is associated with increasing number of affected close relatives (P =0.003). Again, the association remains significant after removing MHC class II. This suggests that a high number of affected relatives within families reflects high polygenic risk rather than presence of a major locus. Thus, higher risk among affected probands of vitiligo multiplex families is due, at least in part, to polygenic inheritance of common low-to-moderate effect variants.

2262W

Trans-ethnic analysis of the human leukocyte antigen region for ulcerative colitis reveals common disease signatures. *F. Degenhardt, M. Wendorff, A. Franke, TE HLA working group.* Institute for Clinical Molecular Biology, Kiel, Germany.

Genome-wide association analyses have identified the major histocompatibility complex (MHC), also termed human leukocyte antigen (HLA), as one of the most disease-associated regions in autoimmune and chronic inflammatory diseases, e.g. inflammatory bowel disease (IBD). For IBD, stronger associations of HLA alleles have been observed for ulcerative colitis (UC) than Crohn's disease, the two main subtypes of IBD. Most HLA association analyses have targeted Caucasian populations in the past and systematic trans-ethnic fine mapping studies are lacking because of the lack of cohorts, high costs for wet-lab (e.g. next generation sequencing) based HLA allele typing and the lack of appropriate reference panels for the (cost-effective) imputation of HLA alleles that account for the polymorphism of the HLA in non-Caucasian ethnicities. Using our high-quality, multi-ethnic reference panel, including 1,360 samples of 8 different ethnicities, we imputed HLA alleles into 9 different ulcerative colitis case-control populations, including African American, Iranian, Indian, Chinese, Japanese, Korean, Caucasian, Maltese and Puerto Rican samples with a total of 3,856 cases and 6,875 controls. We further included our previously reported Caucasian data set (19,014 cases 21,677 controls) for comparison purposes in the analyses. We performed allele- and haplotype-based association analyses on alleles with a frequency of > 1% and identified alleles with consistent effect directions across all populations. Overall, classical HLA alleles and haplotypes contributed most to the statistical disease association than single SNPs. Effect sizes were as high as OR=2.8 [95% CI: 2.4-3.3] were observed for HLA-DQB1*06-DRB1*15 haplotypes in African American, East Asian and Caucasian samples while Iranian, Indian and Puerto Rican samples showed distinct disease haplotypes. The previously reported HLA-B*52:01-C*12:02 haplotype was highly associated with UC in the Korean (OR=3.9 [2.5-6.5]), Japanese (OR=2.9 [2.5-3.6]) and Indian (OR=2.1 [1.6-2.9]) samples and are in strong linkage-disequilibrium with the HLA Class II loci, leading to distinct association signatures than in the other ethnicities (roof-top like shape regional association plots). In conclusion, we identify population-specific and cross-population disease associated HLA alleles and is able to assign these signals to specific common haplotypes, showcasing also our previously established multi-ethnic imputation reference panel.

2263T

Whole genome sequencing to identify gene-by-air pollution interactions that influence lung function in minority children. A.C.Y. Mak¹, M.J. White¹, E. Lee², M. Pino-Yanes², K.L. Keys¹, J. Liberto¹, S. Huntsman¹, D. Hu¹, C. Eng¹, S.S. Oh¹, M.A. Seibold³, E.G. Burchard¹, NHLBI TOPMed and NHGRI CCDG Genome Sequencing Project. 1) University of California San Francisco, San Francisco, CA, United States; 2) Research Unit, Hospital Universitario N.S. de Candelaria, Santa Cruz de Tenerife, Spain; 3) National Jewish Health, Denver, Colorado, United States.

Baseline lung function, quantified as forced expiratory volume in the first second of exhalation (FEV₁), is one of the standard diagnostic criterion used by pulmonary clinicians to identify and classify lung diseases. FEV₁ is strongly influenced by genetics, with heritability estimates as high as 55%. Specifically, numerous genetic loci associated with FEV₁ have been identified, and there is also evidence that genetic ancestry plays a role in FEV₁ variation. Several epidemiological studies have identified early-life exposure to air pollution as a significant predictor of baseline lung function. Although there have been numerous genetic and epidemiological studies aimed at identifying the genetic and environmental factors responsible for variation in FEV₁, the majority of the heritability remains undefined. FEV₁ is a complex phenotype that has been shown to be heavily influenced by both environmental and genetic factors. We hypothesize that a portion of the "hidden heritability" for FEV₁ may be explained by gene-by-environment (GxE) interactions. **Methods** We performed whole genome sequencing on 1,133 Puerto Rican children with asthma and available early-life exposure to air pollution data. We then constructed multiplicative interaction models (single nucleotide polymorphism [SNP] x air pollution). We then performed a GWIS (genome-wide interaction study) to test the effect of each interaction model on variation in FEV₁. All regression models were adjusted for age, sex, global ancestry, and the main effects of the specified SNP and air pollution. **Results and Conclusions** We identified several GxE interactions associated with FEV₁ in Puerto Rican children with asthma. Our preliminary results indicate that GxE interactions may play an important role in FEV₁ variation in this population.

2264F

Using ~20,000 public whole genomes to build reference panels for fine-mapping HLA effects in multi-ethnic cohorts. Y. Luo^{1,2}, M. Kana^{1,2}, M. Gutierrez-Arcelus^{1,2}, J.G. Wilson³, S. Kathiresan⁴, J.I. Rotter⁵, S.S. Rich⁶, M.H. Cho⁷, W.S. Choi⁸, B. Han⁹, Y. Okada⁹, A. Metspalu¹⁰, T. Esko^{2,10}, P.J. McLarren^{11,12}, S. Raychaudhuri^{1,2}, NHLBI TOPMed Consortium. 1) Medicine-Brigham and Women's Hospital, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 2) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA; 3) Physiology and Biophysics, University of Mississippi Medical Center; 4) Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA; 5) Institutional for Translational Genomics and Population Sciences, LABiomed at Harbor-UCLA Medical Center, Torrance, CA, USA; 6) Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA; 7) Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; 8) Department of Medicine, Seoul National University College of Medicine, Korea; 9) Department of Statistical Genetics, Osaka University Graduate School of Medicine, Suita, 565-0871, Japan; 10) Estonian Genome Center, University of Tartu, Tartu 51010, Estonia; 11) J.C. Wilt Infectious Diseases Research Centre, National Microbiology Laboratories, Public Health Agency of Canada, Winnipeg, Canada; 12) Department of Medical Microbiology and Infectious Diseases, University of Manitoba, Winnipeg, Canada.

The human leukocyte antigen (HLA) region harbors genes that are crucial to many human diseases. However, it remains a challenge to pinpoint the causal variants for these associations due to the extreme complexity of the region. We constructed the largest multi-ethnic HLA haplotype panel to date to better understand immune related adaptive evolution, and to facilitate fine-mapping studies from genome-wide association studies (GWAS). First, we inferred HLA types at G-group resolution using whole-genome sequences (WGS) from 20,209 individuals of different ancestries (10,699 Europeans, 7,644 African Americans (AA), 1,016 Hispanics and 850 East Asians) using population reference graphs (Dilthey et al. 2016). We evaluated inferred HLA types against sequencing based typing (SBT) among 295 Japanese samples sequenced at 15x coverage. The accuracies for HLA-A, B, C, DQA1, DQB1 and DRB1 were 95.4%, 97.9%, 98.5%, 99.3%, 98.2% and 97.2% respectively. We observed high levels of differentiation in population allele frequencies among inferred HLA types ($P = 8.7e-267$). In many cases these differences are likely to be related to adaptive selection such as enrichment for the B*53:01:01G and C*04:01:01G alleles in AA samples that have been previously associated with malaria protection. We next built a multi-ethnic HLA reference panel based on inferred HLA types and genetic variation in 5,376 multi-ethnic samples. To evaluate the imputation accuracy of the multi-ethnic panel, we compared imputed HLA haplotypes against SBT among 1,067 AA subjects. The average accuracy among the six classical HLA genes was 96.3%, compared to 77.4% when using a European panel of 5,225 samples alone. To illustrate fine-mapping advantages due to increasing ancestral diversity in the reference panel, we meta-analyzed published HIV-1 virus load GWAS in a total of 6,315 European and 2,924 AA subjects. The most significantly associated allele was an amino acid at HLA-B position 97 in both populations. However, conditional analysis identified different secondary associations at an amino acid at HLA-B position 67 and B*08:01:01G for the European and AA respectively. These results highlight the benefits of a multi-ethnic reference panel for the discovery and characterization of HLA-disease associations. In the next phase, we will build an HLA panel using >20,000 WGS. This resource will open an exciting opportunity to understand immune-related genetic architecture across populations of diverse ancestries.

2265W

Validated South Asian HLA imputation and association analysis uncovers a HLA-DPB1 specific effect on hepatitis B vaccine response in infants from Bangladesh. K. Auckland¹, A.J. Mentzer¹, R. Haque², G. Smits³, C. Pomilla⁴, J.A. Noble⁵, K. Jeffery⁶, F. van der Klis⁷, A.V.S. Hill¹, M.S. Sandhu¹, W.A. Petri Jr.¹ 1) Wellcome Centre for Human Genetics, University of Oxford, Oxford, Oxfordshire, United Kingdom; 2) International Centre for Diarrhoeal Disease Research, Dhaka, Bangladesh; 3) National Institute for Public Health and Environment, Bilthoven, The Netherlands; 4) The Sanger Institute, University of Cambridge, Hinxton, Cambridge, United Kingdom; 5) Children's Hospital Oakland Research Institute, Oakland, California, United States of America; 6) Microbiology Department, The John Radcliffe, Oxford, Oxfordshire, United Kingdom; 7) University of Virginia School of Medicine, Charlottesville, Virginia, United States of America.

Background: Vaccine responses are highly variable as a result of host genetic and environmental factors and in some individuals such responses may be absent increasing the risk of vaccine failure. Despite our increasing understanding of the genetic contribution to vaccine response heterogeneity, large-scale GWAS studies of multiple vaccine responses in South Asian populations with high burden of preventable disease are lacking. **Methods:** We have undertaken a pilot-phase genome-wide association study seeking to identify the genetic variants associated with multiple vaccine responses using ~1.7 million variants genotyped in 541 children from Bangladesh enrolled in PROVIDE (Performance of Rotavirus and Oral Polio Vaccines in Developing Countries). Vaccine responses were defined as quantitative antibody levels measured against eight antigens targeting six separate pathogens. Classical HLA types were generated for all individuals with exon-based NGS genotyping to both validate and undertake HLA imputation and association testing. **Results:** Significant associations were observed between allelic, residue and SNP variants in the class II MHC with response to hepatitis B surface antigen ($P=1.3 \times 10^{-20}$), localising only to *HLA-DPB1*, with no *HLA-DRB1* signal observed; and pertussis pertactin ($P=1.2 \times 10^{-8}$), localising to *HLA-DQB1*. Subthreshold associations were observed with response to diphtheria toxin; pertussis filamentous haemagglutinin adhesin and toxin. Little overlap was observed for associated variants between each vaccine response measured. No associations of genome-wide significance were observed with response to tetanus toxin, attenuated measles or the conjugated *Haemophilus influenzae* vaccines. Concordance between classically typed and imputed HLA alleles using SNP2HLA with a customised reference panel demonstrated 4 digit (2-field) concordance estimates across the class II of over 93%. **Discussion:** By undertaking and validating the first HLA-focussed GWAS of multiple vaccine responses in South Asian children, we have identified genetic variants determining vaccine response heterogeneity and have discovered features of HLA associations distinguishing this cohort from other published populations. We are undertaking the replication analysis in a further ~700 children from Bangladesh. Characterising the genetic determinants of vaccine response heterogeneity could offer a cost-effective and feasible method to significantly reduce the incidence of vaccine failure.

2266T

Whole genome sequencing association analysis of red blood cell traits in a multi-ethnic population sample from the Trans-Omics for Precision Medicine (TOPMed) Project. Y. Hu¹, X. Zheng², D. Jain³, C. Laurie², S. Gogarten², P. Auer³, N. Pankratz⁴, L. Polfus⁵, M. Chen^{6,7}, J. O'Connell⁸, J. Lewis⁹, L. Raffield⁹, A. Correa¹⁰, L. Cupples^{6,11}, N. Jenny¹², S. Rich¹³, R. Mathias¹⁴, L. Yanek¹⁴, J. Blangero¹⁵, J. Curran¹⁵, K. Rice², A. Johnson^{6,7}, C. Laurie², A. Reinher¹⁶. 1) Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA; 2) Department of Biostatistics, University of Washington, Seattle, WA; 3) School of Public Health, University of Wisconsin-Milwaukee, Milwaukee, WI; 4) Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN; 5) Center for Genetic Epidemiology, Keck School of Medicine, Department of Preventative Medicine, University of Southern California, Los Angeles, CA; 6) NHLBI's the Framingham Heart Study, Framingham, MA; 7) NHLBI, Population Sciences Branch, Bethesda, MD; 8) School of Medicine, University of Maryland, Baltimore, MD; 9) Department of Genetics, University of North Carolina, Chapel Hill, NC; 10) Department of Medicine, University of Mississippi Medical Center, Jackson, MS; 11) Department of Biostatistics, Boston University, Boston, MA; 12) Department of Pathology and Laboratory Medicine, Larner College of Medicine, University of Vermont, Burlington, VT; 13) Center for Public Health Genomics, University of Virginia, Charlottesville, VA; 14) Department of Medicine, School of Medicine, Johns Hopkins University, Baltimore, MD; 15) Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX; 16) Department of Epidemiology, University of Washington, Seattle, WA.

Red blood cell (RBC) measurements are polygenic traits, and GWAS, exome chip and sequencing analyses have identified hundreds of associated genetic variants in European, Asian, African and Hispanic ancestry populations. However, additional loci remain undiscovered and causal variant(s) at each locus have not been well characterized. We performed WGS-based association analyses for hemoglobin (HGB), hematocrit (HCT), mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), mean corpuscular volume (MCV), RBC count and red cell distribution width (RDW) in multi-ethnic populations from the NHLBI TOPMed Project, consisting of 25,080 European, African, Hispanic and Asian ancestry individuals from nine studies. More than 42,400,000 SNPs ($MAC \geq 10$) were tested for association using inverse-normal transformed residuals in a multi-ethnic mega-analysis, adjusting for age, sex, study, relatedness, population structure and residual heteroscedasticity. We identified six novel loci reaching genome-wide significance ($P < 5E-8$) that remained significant after accounting for all previously reported RBC-associated variants ($P_{\text{conditional}} < 2E-6$). The six novel loci include *MSRA*-rs1484641 ($MAF=0.35$) for HCT, *CREBBP*-rs129965 ($MAF=0.16$) for MCHC, *FOXO6*-rs557606786 ($MAF=0.001$) and *ANKS1A*-rs13207848 for MCV ($MAF=0.14$), and *4q28*-rs192665038 ($MAF=0.003$) and *18q23*-rs558218738 ($MAF=0.001$) for RDW. All novel variants, except for *ANKS1A*-rs13207848, showed disparate allele frequencies across ancestral populations. We sought replication for these novel variants in UK Biobank, Kaiser and WHI-SHARE imputed GWAS datasets, but none of them was replicated ($P > 0.05$). A total of 26 previously reported RBC trait loci were confirmed to be associated with one or more RBC traits at genome-wide significance in our TOPMed analysis, and six of them (*HBA*, *HBB*, *HFE*, *SP1*, *ABCA7* and *RP11-321F6.1*) harbored residual secondary signals after conditional analysis ($P_{\text{conditional}} < 5E-8$). Finally, in the SKAT and burden tests which aggregated rare variants (alternative allele frequency < 0.01) using 5 or 50kb sliding windows, five previously reported GWAS loci were genome-wide significant ($P < 4E-8$ for 5kb window, and $P < 4E-7$ for 50kb window), including the well-established *HBA* and *HBB* loci, *APCDD1L*, *MXD3* and *MOXD1*. In summary, our results suggest that WGS in larger sample sizes are needed for capturing variants or indels with moderate to larger effects on RBC traits that were missed by GWAS.

2267F

Large-scale genome-wide association study identified candidate genes for childhood nephrotic syndrome in Japanese. X. Jia¹, T. Horinouchi², Y. Hitomi¹, Y. Kawai^{1,3,4}, K. Nozu², C. Nagano², T. Yamamura², K. Kojima^{3,4,5}, M. Nagasaki^{3,4,5}, K. Iijima², K. Tokunaga¹. 1) Department of Human Genetics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; 2) Department of Pediatrics, Graduate School of Medicine, Kobe University, Kobe, Japan; 3) Department of Integrative Genomics, Tohoku Medical Megabank Organization, Tohoku University, Sendai, Japan; 4) Graduate School of Medicine, Tohoku University, Sendai, Japan; 5) Graduate School of Information Sciences, Tohoku University, Sendai, Japan.

Idiopathic nephrotic syndrome (INS) is the most common cause of kidney disease in children, but the underlying genetic etiology is not well understood. Our group has conducted the first genome-wide association study (GWAS) for childhood INS in the Japanese population, genome-wide significant association was identified in *HLA-DR/DQ* region and disease-associated *HLA* alleles were further clarified ("Strong Association of *HLA-DR/DQ* Locus with Childhood Steroid-Sensitive Nephrotic Syndrome in the Japanese Population". *Journal of the American Society of Nephrology*). To identify the other candidate loci contributing to the susceptibility of childhood INS, especially the disease-related genes in non-*HLA* regions, we performed an extended GWAS in the Japanese population with large samples. A total of 1,152 patients with childhood-onset INS and 2,807 adult healthy controls were genotyped using Affymetrix 'Japonica Array' in the discovery stage. Whole-genome imputation was conducted using a phased reference panel of 2,049 healthy Japanese individuals (2KJPN panel). Quality control was performed to exclude the individuals with low calling rate (<97%) and the variants with low genotyping rate (<97%), minor allele frequency (MAF) <1% and Hardy-Weinberg equilibrium (HWE) test p-value <1×10⁻⁶. Association analysis was conducted using logistic regression with adjustment of gender and principal components (PCs). In GWAS, the most significant association was detected in *HLA-DR/DQ* region as we reported before (rs754421875, P=5.42×10⁻⁴⁷, odds ratio (OR)=0.32). Another signal on chromosome 19 achieved genome-wide significance (P=9.69×10⁻¹⁸, OR=1.83). The validation and replication of candidate variants will be done to confirm the associations and further elucidate the disease-associated variants. With this large sample size, the power of the current GWAS exceeded 80% to detect common variants (MAF>=5%) with genotypic relative risk above 1.95, making it's possible to detect disease-associated genetic factors with modest associations as well as the associated variants with low allele frequencies. This study will improve our knowledge of the genetic background of childhood INS, providing a better understanding of the pathogenic mechanism.

2268W

Novel transcriptomic analysis of total immunoglobulin E (IgE) in Puerto Rican children. Y. Jiang^{1,2}, T. Wang¹, E. Forno¹, G. Canino³, E. Acosta-Pérez³, W. Chen¹, J.C. Celedón¹. 1) Division of Pulmonary Medicine, Department of Pediatrics, Children's Hospital of Pittsburgh of UPMC, University of Pittsburgh, Pittsburgh, PA; 2) School of Medicine, Tsinghua University, Beijing, China; 3) Behavioral Sciences Research Institute, University of Puerto Rico, San Juan, PR.

Early sensitization with immunoglobulin E (IgE)-mediated allergic inflammation can result in the development of asthma and/or allergic diseases later in life. Although genetic and epigenetic determinants have been associated with serum total IgE and asthma, the underlying mechanisms are still largely unknown. In this study, we performed a novel transcriptomic study of total serum IgE in a cohort of Puerto Rican children, which have the highest prevalence of asthma among all ethnic groups. Whole transcriptome sequencing was performed on white blood cells from 238 children with asthma and 234 controls. A regression framework was applied to identify genes associated with total IgE. Genes that were significantly associated with total IgE were combined to build a statistical prediction model. A total of 83 genes were significantly associated with total IgE (false discovery rate P < 0.1), of which genes *IL5RA* and *IL17RB* are known to affect allergic inflammation. Most of our findings were validated in microarray data of whole blood samples from another Puerto Rican cohort (n=84). Integrated with DNA genotyping data, the expression quantitative trait locus (eQTL) studies identified significant genotype-expression pairs, including several that have been reported, such as *rs11136344/PARP10*, *rs745307/CD40*. Furthermore, pathway analysis revealed 9 significantly enriched pathways (P < 0.01), including Th1 and Th2 activation pathways. A prediction model achieved satisfactory performance (AUC=0.80). Our findings advance the knowledge of allergy pathogenesis and offer a potential diagnostic tool for allergic diseases in a clinical setting.

2269T

Large-scale PheWAS of the MHC region in ~170,000 Japanese using NGS-based classical and non-classical HLA imputation. Y. Okada^{1,2}, J. Hirata¹, K. Hosomichi², S. Sakaue¹, M. Kanai¹, H. Nakaoka⁴, K. Ishigaki², K. Suzuki¹, M. Akiyama², T. Kishikawa¹, K. Ogawa¹, T. Masuda¹, K. Yamamoto¹, M. Hirata⁵, K. Matsuda⁵, I. Inoue⁴, M. Kubo², Y. Kamatani². 1) Osaka University Graduate School of Medicine, Osaka, Japan; 2) RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 3) Kanazawa University, Kanazawa, Japan; 4) National Institute of Genetics, Shizuoka, Japan; 5) The University of Tokyo, Tokyo, Japan.

The major histocompatibility complex (MHC) region confers genetic risk on a variety of human phenotypes, while detailed fine-mapping has yet to be elucidated. Here, we conducted NGS-based typing of the 33 human leukocyte antigen (HLA) genes of the Japanese population ($n = 1,120$), providing high resolution allele catalogue up to 6-digit HLA alleles and linkage disequilibrium (LD) structure of both classical and non-classical HLA genes. Together with population-specific deep whole-genome sequencing (WGS) data of Japanese (on average $25\times$, $n = 1,276$), we conducted NGS-based HLA, SNV, and Indel imputation of the large-scale genome-wide association (GWAS) genotype data of Japanese from BioBank Japan Project, a nation-wide hospital-based cohort of Japan ($n = 166,190$). A phenome-wide association study (PheWAS) assessing 106 clinical phenotypes including human complex diseases, as well as anthropometric, biochemical, and hematological quantitative traits, identified abundant significant genotype-phenotype associations across 52 phenotypes. Fine-mapping within the MHC region revealed multiple association patterns conferring independent risks from the classical HLA genes including those represented as non-classical HLA gene variants, non-HLA variants within MHC, and long-range haplotype spanning the entire MHC region. We also identified novel MHC associations of the traits (pollinosis, hyperlipidemia, myocardial infarction, stable angina, type 2 diabetes, liver cancer, liver cirrhosis, and nephrotic syndrome). MHC region-wide heritability estimates and genetic correlation network analysis elucidated polygenic architecture shared across the phenotypes, which was distinct between classical HLA variants and other variants in MHC. This study should contribute to our understanding of genetic and phenotypic landscape of MHC.

2270F

Specific interaction between host immune-related genetic factors and lineage in *Mycobacterium tuberculosis* identified from pathogen lineage based genome-wide association study in tuberculosis. Y. Omae¹, L. Toyo-oka¹, S. Mahasirimongkol², H. Yanai¹, N. Smittipat³, P. Palittapongarnpim³, S. Nedsuwan², S. Wattanapokayakit², N. Satproedprai², P. Sawanpanyalert², W. Inunchot², E. Pasomsub², N. Wichukchinda², T. Mushiroda⁴, M. Kubo⁴, K. Tokunaga¹. 1) The University of Tokyo, Bunkyo-ku, Tokyo, Japan; 2) Ministry of Public Health, Nonthaburi, Thailand; 3) National Center for Genetic Engineering and Biotechnology, Pathum Thani, Thailand; 4) RIKEN Center for Integrative Medical Sciences, Yokohama, Japan.

Tuberculosis (TB) is one of the three major global infectious diseases caused by pathogenic bacterium, *Mycobacterium tuberculosis* (MTB). Although twin studies suggested the contribution of host genetic factors to TB onset, previous genome-wide association studies (GWASs) for TB have faced difficulty in identifying the genetic factors concordant among independent GWASs. MTB is known to have the heterogeneity in its genome structure and seven major global lineages have been reported worldwide. We recently conducted a stratified GWAS for TB in Thailand based on the lineage information of MTB from 686 TB patients and assessed possible interaction between pathogen genomic heterogeneity and host genetic factors. The pathogen-lineage based GWAS revealed a genome-wide significant association of one SNP on chromosome 1p13, which was specifically associated with non-Beijing lineage infected cases with age at onset older than 45 years of age ($p=2.54E-08$, odds ratio (OR)=1.74 [95%CI=1.43-2.12]). The SNP was located near the *CD53* gene, which encodes a leukocyte surface glycoprotein and known as a modulator of cytokine production (Y. Omae, et al., Journal of Human Genetics 2017). In this study, we further applied whole genome SNP imputation for our dataset and revealed that SNPs on chromosome 6p21 showed the strongest association with Beijing lineage infected cases ($p=2.14E-07$, OR=2.00 [95%CI=1.53-2.61]). The 6p21 locus included class II human leukocyte antigen (*HLA*) genes, whose associations with TB have been reported from the previous GWASs in European and Han Chinese. HLA allele imputation based on the surrounding SNP genotypes revealed the strongest association of *HLA-DRB1*09:01* allele with Beijing lineage infected cases ($p=5.47E-06$, OR=1.81 [95%CI=1.40-2.35]). Beijing lineage isolates were reported to induce lower levels of the inflammatory cytokines in monocyte-derived macrophages and dendritic cells than non-Beijing lineage isolates *in vitro*, thus the phenotypic differences in MTB can affect the risk of host genetic factors and acquired immunity via class II *HLA* might be an important step to fight against specific lineage of this pathogen. Our findings demonstrate the importance of analyzing the interaction between host and pathogen genomes in TB.

2271W

Published and novel human associations of blood cell traits in the Healthy Nevada Project cohort. K. Schlauch, R. Read, J. Metcalf, S. Rybar-ski, S. Sweet, J. Grzymalski. Desert Research Institute, Reno, NV.

Background: In 2017, Renown Health and DRI began the Healthy Nevada Project, consisting of an extensive Electronic Medical Record database and genotypic data of more than 10,000 individuals in Northern Nevada based on the Illumina Human OmniExpress-24 BeadChip. **Methods:** In this study, a genome-wide association study (GWAS) was performed on 4,610 Healthy Nevada Project participants to assess the genotypic association of Mean Platelet Volume (MPV), Mean Corpuscular Volume (MCV), and Platelet Counts (PC) against 602,000 genotypes from the Illumina Human OmniExpress-24 BeadChip. After performing standard quality control measures (verifying 95% SNP call rates, 95% individual call rates, Hardy-Weinberg Equilibrium with $p < 1 \times 10^{-6}$), 498,709 quality-controlled genotypes with minor allele frequency $> 1\%$ and 4,532 patients were tested for association using PLINK v. 1.9. Linear regression was performed on each SNP to measure the genetic association between genotype and quantitative MPV, MCV and PC measures, using age and gender as covariates, and the standard log-additive genetic model, following previous similar studies. **Results:** This study identified 215 statistically significant SNPs putatively associated ($p < 5 \times 10^{-8}$) with at least one of the blood cell traits. SNPs associated with MPV lie in eight genomic regions; five of these regions were reported previously in at least one other MPV study. Specifically, SNPs rs342293, rs342296 and rs342275 in region 7q22.3 represent our highest association with MPV ($p < 2 \times 10^{-13}$) and were linked to MPV, PC, and platelet formation in five previous GWAS. All three SNPs had positive effect sizes. Additionally, we found rs385893 in region 9p24.1 to be statistically significantly associated to PC in our cohort ($p = 2.67 \times 10^{-10}$, and beta coefficient = -.09); this SNP was linked to PC and MPV in at least three earlier GWAS. We also identified ten SNPs in a region within the *MYB* gene in 6q23.3 with MCV p -values $< 1 \times 10^{-10}$ and positive effect sizes; these SNPs were previously identified as associated to MCV in at least three other GWAS. Five SNPs in the 22q12.3 region were also identified here to be associated with MCV, as in several previous GWAS. Our study also links several yet unidentified regions with high statistical significance with one or more blood cell traits awaiting further investigation. **Conclusion:** This study shows previously identified and putative novel associations in blood cell traits in the Healthy Nevada Project cohort.

2272T

Genome-wide association study of HIV-1 subtype C in Botswana population. A.K. Shevchenko¹, S.V. Malov¹, D.V. Zhernakova^{1,2}, N. Cherkasov¹, A. Svitin¹, W. Xie^{3,4}, V. Novitsky^{3,4}, M. Essex^{3,4}, S.J. O'Brien¹. 1) Dobzhansky Center for Genome Bioinformatics, SPSU, St. Petersburg, St. Petersburg, Russian Federation; 2) University of Groningen, University Medical Center Groningen, Department of Genetics, Groningen, the Netherlands; 3) Harvard T.H. Chan School of Public Health AIDS Initiative, Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA; 4) Botswana Harvard AIDS Institute, Gaborone, Botswana.

HIV-infection is a deadly chronic disease spreading across all continents. More than 36 000 000 people currently live with HIV and more than 1 000 000 die due to AIDS (UNAIDS, 2015). Despite efforts and resources directed to the fight against the HIV epidemics for more than three decades, still currently there exists no effective vaccine or cure. According to UNAIDS, Botswana falls into top 3 countries with the highest HIV prevalence in the world with adult aged 15 to 49 HIV prevalence rate of 22.2% (UNAIDS, 2015). Moreover, HIV-1 subtype C (widespread in Botswana), which accounts for over half of HIV-1 infections worldwide, is understudied, because most studies focus on other subtypes. Genome-wide association studies (GWAS) allow to identify common variants associated with the trait in question. In order to efficiently search for the genetic associations we have previously developed GWATCH web platform (Svitin et al, 2014). The broad goal of the Botswana GWAS project is to identify genetic determinants of susceptibility and resistance to infection by HIV-1 subtype C among people severely affected by HIV/AIDS in Botswana. By conducting GWAS analysis on HIV-1C case/control dataset consisting of 762 Tswana people (combined from two partly overlapping datasets of 809 microarray and 362 WGS samples), we found several gene regions slightly below significance level. One of these loci was successfully replicated in the longitudinal study on CD4 and viral load (VL) trajectories in a total of 556 treatment-naïve HIV-1C infected individuals in Botswana, which was the first GWAS for HIV-1C acquisition and control of viral replication and the first GWAS in this population (Xie et al, 2016). Despite different types of clinical data in the compared cohorts (HIV infection & AIDS progression), this is an interesting and noteworthy replication. Currently we are working on the replication of our findings in other available datasets. 1. Svitin A. et al. (2014). GWATCH: a web platform for automated gene association discovery analysis. GigaScience. 2. Xie W. et al. 2016. Genome-Wide Analyses Reveal Gene Influence on HIV Disease progression and HIV-1C Acquisition in Southern Africa. AIDS Research and Human Retroviruses.

2273F

Genetic basis of hypoxia-induced excessive erythrocytosis: Role of *SENP1* gene regulation. P. Azad¹, A. Akbari², G. Chen³, D. Zhou¹, O. Poulsen¹, Y. Hsiao¹, E. Bouhassira⁴, V. Bafna⁵, P. Mali⁶, G. Haddad^{1,5,6}.

1) Department of Pediatrics, University of California San Diego, La Jolla, CA; 2) Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA; 3) Department of Bioengineering, University of California San Diego, La Jolla, CA; 4) Department of Medicine, Albert Einstein College of Medicine, Bronx, NY; 5) Department of Neurosciences, University of California, San Diego, La Jolla, CA; 6) Rady Children's Hospital, San Diego, CA.

Since polycythemia is a predominant trait in some high altitude dwellers (Chronic Mountain Sickness, CMS, or Monge's disease) but not others living at the same altitude in the Andes, we took advantage of this human "experiment in nature" and studied both populations (with CMS and without, non-CMS). In order to understand the molecular basis for polycythemia of high altitude, we generated a disease in-the dish-model by re-programming fibroblasts from CMS and non-CMS subjects. In addition, we have done whole genome sequencing of subjects from this region (CMS and non-CMS) to study whether there was a genetic adaptation to high altitude. Using both iPS cells and genomic data, we demonstrated that *SENP1* (a candidate gene) plays an important role in the polycythemia response to hypoxia in CMS. By altering *SENP1*, we converted the non-CMS into CMS cells and vice versa in our in-vitro studies. In this current study, using bioinformatics tools (PreCIOS), we prioritized the potential causal SNP(s) (out of the 95 differential SNPs) for up-regulating *SENP1* in CMS or preventing this up-regulation in non-CMS cells under hypoxia. We found 7 SNPs that could be critically linked to the upstream regulation in *SENP1* expression. In addition, we identified the transcriptional regulators that bind to the genomic region containing the candidate SNPs in the *SENP1* gene region. Indeed, when we compared the transcriptional response of CMS to that of non-CMS cells in hypoxia, we found that CMS cells up-regulated *SENP1*, suggesting that this up-regulation could underlie the excessive erythrocytosis present in the CMS population. Additionally, we are starting to evaluate the contribution of each candidate SNP (or their combination) in the hypoxia-induced up-regulation of *SENP1*, such as switching (A to G) the allele in CMS subjects for one shortlisted SNPs (rs11613781) in non-CMS, as we have done recently using CRISPR/Cas9. We conclude that the increased erythropoietic hypersensitivity to hypoxia in CMS subjects is genetic in nature and depends on an increase in *SENP1* expression under hypoxia. By combining the iPS technology with this unique Andean population that has adapted (or mal-adapted) to chronic hypoxia over thousands of years, we have built an in-vitro model to probe at the exact genetic mechanisms underlying this adaptation to high altitude.

2274W

Identification of three novel loci associated with systemic lupus erythematosus (SLE) and evaluation of SLE genetics in drug repositioning. Y.F. Wang¹, Y. Zhang², T.Y. Wang³, Z. Zhu³, D.L. Morris⁴, J.J. Shen¹, H. Zhang¹, H.F. Pan⁵, J. Yang¹, S. Yang³, D.D. Ye⁵, T.J. Vyse⁴, Y. Cui⁶, X. Zhang³, Y. Sheng³, Y.L. Lau¹, W. Yang¹.

1) The University of Hong Kong, Hong Kong, China; 2) Guangzhou Women and Children's Medical Center, Guangzhou, China; 3) Institute/department of Dermatology, no.1 Hospital, Anhui Medical University, Hefei, China; 4) Division of Genetics and Molecular Medicine, King's College London, London, UK; 5) Department of Epidemiology and Biostatistics, School of public Health, Anhui Medical University, Hefei, China; 6) Department of Dermatology, China-Japan Friendship Hospital, Beijing, China.

Systemic lupus erythematosus (SLE) is prototype autoimmune disease with a diverse spectrum of clinical symptoms, ranging from skin rash to end-organ damage. Through trans-ethnic genome-wide association studies (GWAS), we recently identified ten novel loci associated with SLE. In this study, we aimed to replicate loci with suggestive association signals in an additional 6,728 SLE cases and 8,631 healthy controls of Chinese ancestry. We identified three novel loci that passed the genome-wide significance, including *ST3AGL4* (rs13238909, $P_{\text{meta}} = 4.40\text{E-}08$), *MFHAS1* (rs2428, $P_{\text{meta}} = 1.17\text{E-}08$), *CSNK2A2* (rs2731783, $P_{\text{meta}} = 1.08\text{E-}09$), and confirmed the association of *CD226* locus with SLE (rs763361, $P_{\text{meta}} = 2.45\text{E-}08$). We further performed fine mapping analyses at the newly identified loci and found two putative causal variants in *CSNK2A2* locus. The putative causal variants reside in an enhancer and are associated with expression of *CSNK2A2* in B-lymphocytes, suggesting a potential mechanism of the association. To explore biological insight underlying the SLE susceptibility loci, we prioritized putative risk genes by integrating the GWAS findings with functional annotations. We identified 78 putative SLE risk genes and demonstrated that they are more likely to encode proteins interacting with molecular targets of approved SLE drugs. Our study also suggested that several drugs approved for other diseases might have effects on SLE patients. In summary, this study identified three novel loci associated with SLE and demonstrated the role of SLE GWAS findings in drug repositioning.

2275T

Endometrial microRNA networks associated with pelvic inflammatory disease progression. L. Dong¹, C. O'Connell¹, J. Xu², H. Wiesenfeld^{4,5}, S. Hillier^{4,5}, T. Darville², X. Zheng². 1) Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Department of Pediatrics, University of North Carolina at Chapel Hill, Chapel Hill, NC; 3) College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, China; 4) Department of Obstetrics, Gynecology, and Reproductive Sciences, University of Pittsburgh, Pittsburgh, PA; 5) Magee-Womens Research Institute, Pittsburgh, PA.

Background and objectives: *Chlamydia* (CT) and *gonorrhea* (GC) are the most common sex transmitted infection. Untreated infection can ascend from cervix to uterus and fallopian tubes resulting in pelvic inflammatory disease (PID). Our previous study revealed distinct blood mRNA transcriptome profiles associated with PID. MicroRNAs (miRNAs) modulate processing of up to 60% of human mRNA transcripts. The objective of this study is to determine human endometrial tissue miRNA networks associated with PID progression, key drivers of these networks and their target mRNAs. **Methods and results:** Endometrial miRNA from 57 women from two cohorts was profiled by seq based whole transcriptome assay. We compared profiles of women with clinical PID and endometrial GC and/or CT infection and inflammation (n=10), defined as cases, to those of asymptomatic women with cervical only infection and no inflammation (n=12), defined as controls, to determine disease-associated immune responses. We also compared the responses of these 10 cases to responses of 19 women who had pelvic pain, but no infection. Cases possessed a profile that distinguished them from controls, as well as women with pelvic pain without infection. The responses of asymptomatic women without inflammation but with cervical only infection were indistinguishable from asymptomatic, uninfected women. Weighted gene co-expression network analysis identified 4 miRNA modules (clusters of highly correlated miRNAs) associated with PID. Three out of four disease associated modules were validated in 5 women with subclinical PID who had asymptomatic endometrial infection and inflammation. The disease modules abated one month after curation of infection. These findings indicate the disease associated modules were driven by endometrial infection induced inflammation instead of abdomen pain. Experimentally validated and high-confidence predicted mRNA targets of disease modules were identified by Ingenuity Pathways Analysis, and validated by endometrial mRNA seq using 7 cases and 6 uninfected controls. The target mRNAs were enriched in T cell signaling, myeloid cell genes and cell growth and death pathways. The key miRNAs regulating these pathways were further revealed by miRNA-miRNA functional synergistic network analysis. **Conclusions:** Endometrial miRNA profiles reveal miRNA networks associated with PID progression that are driven by endometrial infection induced inflammation.

2276F

In-depth analysis of differential DNA methylation reveals novel disease genes and pathways in SLE. W. Yang, M. Guo, T.Y. Wang, Y.F. Wang, J.J. Shen, Y.L. Lau. University of Hong Kong, Hong Kong, China.

Systemic lupus erythematosus (SLE) is a prototype autoimmune disease with unclear pathogenesis. DNA methylation is an important regulatory mechanism on gene expression, which provides a key angle to understand disease mechanisms. Using public available methylation data for SLE, we performed integrative analyses of differentially methylated regions (DMR) and methylation sites with differential variability (DVS). We identified the most DMRs in T cells, while majority of DVSs were found in B cells, featuring hyper variability in enhancers. We observed a prominent T cell receptor (TCR) signaling cluster with consistent hypermethylation and a B cell receptor (BCR) cluster with high variability in SLE samples. Genes involved in innate immunity were often found hypomethylated, while adaptive immunity genes were featured with hypermethylation. Significant co-localization of DMRs and DVSs with SLE susceptibility loci was also observed. Using LASSO, a machine learning method, we selected a panel of CpGs that accurately distinguished SLE patients from healthy individuals, and correlated with disease activities. These differentially methylated genes may further help us understand the pathways involved in the disease, and to develop biomarkers for its diagnosis and treatment.

2277W

Differential gene expression analysis in CD4+ T cells from the Barbados asthma genetics study. M.P. Boorgula¹, S. Chavan¹, M.A. Taub², M. Daya¹, A. Shetty¹, T.M. Brunetti¹, N.M. Rafaels¹, M. Campbell¹, P. Maul¹, T. Maul¹, A.R. Greenidge³, D. Walcott⁴, A. Franklin¹, R.C. Landis⁵, H. Watson⁶, T.H. Beaty⁶, I. Ruczinski⁶, R.A. Mathias³, K.C. Barnes¹. 1) Department of Medicine, University of Colorado, Aurora, CO; 2) Department of Biostatistics, Bloomberg School of Public Health, JHU, Baltimore, MD, USA; 3) Department of Epidemiology, Bloomberg School of Public Health, JHU, Baltimore, MD, USA; 4) University of West Indies at Cavehill, Barbados, WI; 5) Edmund Cohen Laboratory for Vascular Research, The George Alleyne Chronic Disease Research Centre, The University of the West Indies at Cave Hill, Barbados W.I; 6) Faculty of Medical Sciences, The University of the West Indies, Queen Elizabeth Hospital, Bridgetown, St. Michael, Barbados.

RATIONALE: CD4+ T helper cells are known to play a role in pathogenesis of asthma and are therefore good target tissues for elucidating disease mechanisms. In this study, we aimed to characterize differential gene and transcript level expression of CD4+ T cells in asthmatics and non-asthmatics from the Barbados asthma genetics study (BAGS). **METHODS:** Paired-end RNA-Seq on RNA from CD4+ T-cells isolated from 84 samples was performed on the Illumina HiSeq4000, using the TruSeq Total RNA RiboZero kit for library preparation. 10 samples were excluded from further analysis due to high percentages of ribosomal bases (greater than 20%). Low quality RNA-Seq read removal and adapter trimming were performed using BBDUK. The trimmed reads for all samples were mapped and assembled using the HISAT2 graph aligner and STRINGTIEv1.3.0 with the UCSC hg38 reference genome. Assembled transcripts from all samples were merged using STRINGTIEv1.3.0 to create a uniform set of transcripts and to re-estimate the transcript abundance. Gene/transcript expression tables were created with the Ballgown R package. Transcripts whose FPKM (fragments per kilobase of transcript per million mapped reads) measures had a cross-sample variance less than 1 were removed to filter out low abundance genes. Ballgown's "stattest" function was used to perform differential gene and transcript expression using FPKM measurements, between 43 atopic asthmatic cases and 19 non-atopic non-asthmatic controls. **RESULTS:** Among the 43 atopic asthmatics and 19 non-asthmatic controls included in the analysis, there were 8,416 genes and 24,060 transcripts that had transcripts with variance greater than 1 across all samples. Differential expression analysis identified 17 genes with a P-value less than 0.001. *ANAPC15* and *EPC1* were the most significant genes with a P-value of 6.3×10^{-5} and 1.1×10^{-4} , respectively. **CONCLUSIONS:** Differentially expressed genes like *ANAPC15* and *EPC1* might play a role in asthma pathogenesis. Future work will include incorporating additional subjects and using gene set enrichment analysis to identify gene regulatory networks associated with dysregulation of CD4+ T cells in asthma. Future analyses will also include identifying expression quantitative trait loci (eQTL) specific to cases and identify any novel eQTLs in cases compared to non-atopic controls.

2278T

Leukocyte gene expression signatures of asthma severity. J.A. Resztrak¹, A.K. Farrell², J.E. Mair-Meijers¹, A. Alazizi¹, S. Zilioli², R. Pique-Regi¹, R.B. Slatcher², F. Luca¹. 1) Center for Molecular Medicine and Genetics, Wayne State University, Detroit, MI; 2) Department of Psychology, Wayne State University, Detroit, MI.

Allergic asthma is the most common chronic disease in children in the industrialized world, and it remains the most common reason for hospitalization other than infections. There are striking disparities in asthma prevalence in the US population: African American children are 1.6 times overrepresented within asthmatics and three times more likely to be hospitalized and die as a consequence of asthma compared to European American children. GWA studies found several hundred SNPs associated with asthma (most of them located outside the coding regions), but the factors influencing asthma severity and response to therapy remain largely unknown. We followed a cohort of almost two hundred predominantly African American asthmatic children living in Detroit Metropolitan Area annually for a period of two years. We collected genotypic and extensive phenotypic information on over one hundred asthma-related variables, including asthma symptoms, medication use, pulmonary function, and others. Here we present the analyses of leukocyte RNA-seq in 165 subjects. Expectedly, blood cell composition has a major effect on gene expression, after correcting for technical and biological confounders. For example, we find 15808 genes (DEGs; FDR = 10%) whose expression is associated with lymphocyte proportion. Additionally, we performed eQTL mapping and identified 4325 eGenes (FDR = 10%), over a thousand of which have not previously been detected in GTEx blood data. Our eGenes are enriched for genes associated with GWAS catalog traits ($OR \approx 1.14$, $pval = 0.0004$), and include 30 genes previously associated with asthma. This dataset will be useful to identify the genetic and environmental exposures that modify the molecular pathways and contribute to disease severity.

2279F

Single-cell RNA-sequencing of bulk peripheral blood mononuclear cells from idiopathic multicentric Castleman disease (iMCD) reveals leukocyte heterogeneity underlying disease pathology. M. Gonzalez¹, D. Shilling², F. Mafrá¹, J. Smiler¹, C. Kao¹, J. Garifallou¹, R. Pellegrino¹, D. Fajgenbaum², H. Hakonarson^{1,3}. 1) Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Division of Translational Medicine and Human Genetics, Hospital of the University of Pennsylvania, Philadelphia, PA; 3) Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

Castleman disease (CD) describes a group of heterogeneous lymphoproliferative disorders that share a distinct spectrum of lymph node histopathological features, including atrophic or hyperplastic germinal centers, hypervascularization, polyclonal lymphoproliferation, and/or polyclonal plasmacytosis. Hallmarks of idiopathic multicentric CD (iMCD) include multiple regions of enlarged lymph nodes, cytokine-driven multiple organ system impairment, and systemic inflammatory symptoms, which ultimately lead to a ~35% 5-year mortality rate. Despite approximately 1,500 cases of iMCD diagnosed per year in the US, very little is known about the cell types, signaling pathways and molecules that drive the disease. Therefore, we performed single-cell RNA-sequencing (scRNA-seq) to characterize 39,457 bulk peripheral blood mononuclear cells (PBMCs) isolated from an iMCD patient at distinct time points during his disease course; PBMCs were collected while the patient was in remission (no abnormal laboratory values or clinical features) and three days prior to admission to the intensive care unit for an acute flare episode, characterized by constitutional symptoms, ascites, thrombocytopenia, hypoalbuminemia, and kidney failure. The cellranger pipeline (10x Genomics, v.2.1.0) was used for initial analysis and aggregation of samples. 19,322 cells (17,327 mean reads/cell, 823 median genes/cell) were recovered from bulk PBMCs in flare and automated clustering identified 18 distinct cellular populations. 20,135 cells (16,283 mean reads/cell, 799 median genes/cell) were recovered from bulk PBMCs collected during remission identifying 21 distinct cellular populations. Further, single-cell transcriptomes from both time points were aggregated and normalized to allow comparison between immune cell populations in flare and remission. These comparisons indicate multiple population specific alterations and changes in gene expression during flares that may explain disease specific pathology. This collection of single-cell transcriptomes and distinct classification of cellular subsets provide a novel resource for understanding gene expression and cell population changes in iMCD flare.

2280W

Pleiotropic effects for Parkin and LRRK2 in leprosy type-1 reaction and Parkinson's disease. V.M. Fava^{1,2,14}, Y.Z. Xu^{1,2}, G. Lettre^{3,4}, N.V. Thuc⁵, M. Orlova^{1,2}, V.H. Thai⁶, G. Cambré⁷, S. Tao^{7,8}, R. Lahiri⁹, L. Adams⁹, A. Cobat^{10,11}, A. Alcaïs^{10,11}, L. Abel^{10,11,12}, E. Schurr^{1,2,13,14}. 1) The Research Institute of the McGill University Health Centre. Program in Infectious Diseases and Immunity in Global Health, Montreal, Canada; 2) McGill International TB Centre, Montreal, Canada; 3) Montreal Heart Institute, Montreal, Canada; 4) Department of Medicine, Faculty of Medicine, Université de Montréal, Montreal, Canada; 5) Hospital for Dermato-Venerology, Ho Chi Minh City, Vietnam; 6) Pontificia Universidade Católica do Paraná, Graduate Program in Health Sciences, Curitiba, Parana, Brazil; 7) Division of Experimental Medicine, Faculty of Medicine, McGill University, Montreal, Canada; 8) The Translational Research in Respiratory Diseases Program, The Research Institute of the McGill University Health Centre, Montreal, Canada; 9) Health Resources and Services Administration (HRSA), National Hansen's Disease Program (NHDP), Baton Rouge, United States of America; 10) Laboratory of Human Genetics of Infectious Diseases, Necker Branch, Institut National de la Santé et de la Recherche Médicale 1163, Paris, France; 11) Imagine Institute, Paris Descartes-Sorbonne Paris Cité University, Paris, France; 12) St. Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, Rockefeller University, New York, United States of America; 13) Department of Human Genetics, Faculty of Medicine, McGill University, Montreal, Canada; 14) Drs. Fava and Schurr contributed equally to this project.

Background Type-1 reactions (T1Rs) are pathological inflammatory episodes and main contributors to nerve damage in leprosy. Previous studies revealed a significant overlap in the genetic control of T1R with Crohn's disease. Here, we searched for rare coding variants as risk factors for T1R and evaluated to what extent such rare amino acid changes were shared with Parkinson's disease (PD), a neurodegenerative disease where inflammation is a key component. **Methods** Seven known T1R associated genes were targeted for deep resequencing in 474 leprosy-patients from Vietnam. Of these, 237 were T1R-affected cases that were compared to 237 T1R-free matched controls. The gene-wise enrichment of nonsynonymous variants was tested with both kernel based (SKAT) and burden methods. The impact of amino acid changes on protein structure was assessed with PyMol. Wild-type RAW cells and RAW cells edited with CRISPR/Cas9 to introduce T1R and PD associated mutations were challenged with BCG or *Mycobacterium leprae* for up to 6 hours. Upon infection, we estimate the impact of selected mutations in the ROS production and in apoptosis. **Results** Of the seven genes tested two showed statistical evidence of association with T1R. For the Parkin coding gene *PRKN* (formerly *PARK2*) seven rare variants were enriched in T1R-affected cases ($p_{\text{SKAT-O}} = 7.4 \times 10^{-6}$). T1R-risk amino acids substitutions in Parkin were enriched in Parkin residuals observed mutated in patients suffering from PD ($p = 1.1 \times 10^{-3}$). The second gene-wise association with T1R was found for the *LRRK2* where nonsynonymous variants were enriched in T1R-free controls ($p_{\text{SKAT-O}} = 1.6 \times 10^{-4}$). This gene-wise association was driven almost entirely by the gain of function variant R1628P ($p = 0.004$; OR = 0.29) that is a known PD risk factor in Asian populations. We observed that genome-edited RAW 1628 P/P mutants challenged with BCG and *M. leprae* had increased ROS production compared to wildtype cells. RAW 1628 P/P also strongly reduced apoptosis and necrosis increasing cell survival upon BCG challenge ($p < 0.001$). **Conclusion** Amino acid mutations in *LRRK2* and *PRKN* are critically involved in the pathogenesis of T1R. The mutations displayed a strong pleiotropic effect with Parkinson's disease. Hence, neuroinflammation in PD and peripheral nerve damage in T1R share overlapping mechanisms of pathogenicity.

2281T

Chromatin accessibility landscapes of large and small airway cells annotate multiple COPD susceptibility GWAS regions. C.J. Benway, F. Du, M.H. Cho, E.K. Silverman, X. Zhou. Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA.

Genome-wide association studies (GWAS) have identified hundreds of genome-wide significant loci for respiratory disease, including Chronic Obstructive Pulmonary (COPD), asthma, pulmonary fibrosis, lung function and COPD-related phenotypes. However, identification of causal variants and functional annotation in the appropriate cell type at these loci remain a major challenge. To gain a better understanding of COPD and other respiratory diseases, we sought to functionally annotate GWAS loci using chromatin accessibility profiling of small airway and bronchial epithelial cells from healthy human lungs. We performed Omni-ATAC-seq in small airway epithelial cells (SAEC) from two biological replicates and bronchial epithelial cells (NHBE) in one biological replicate as well as the 16HBE human bronchial epithelial cell line. We confirmed that Omni-ATAC-seq outperforms Fast-ATAC-seq in 16HBE cells with regard to mtDNA contamination, signal-to-noise, library complexity, and enrichment in annotated genomic regions such as transcription start sites and CTCF sites. We tested for enrichment of 22 COPD susceptibility GWAS loci in the lung epithelial ATAC-seq peaks. Sets of SNPs within GWAS loci were generated either by a heuristic or Bayesian fine-mapping approach. For the heuristic approach, LD proxy SNPs ($LD\ r^2 > 0.8$) were determined for each lead SNP based on 1000 Genomes data. For the bayesian fine-mapping approach, candidate causal variants for each locus were determined using the Probabilistic Identification of Causal SNPs (PICS) algorithm. To aid in the functional annotation of COPD GWAS loci, we compared chromatin accessibility peaks and signals of these different cell types. Both primary lung subtypes were highly correlated with one another, but the 16HBE bronchial epithelial cell line was only moderately correlated with its primary counterpart. We identified 9,112 differentially accessible peaks ($FDR < 0.05$) between large and small airway epithelial cells. Further, we profiled genome-wide footprinting of 533 transcription factors using the Protein Interaction Quantification (PIQ) computational method. Our work confirms the performance of Omni-ATAC-seq, and suggests cell lines substantially differ from primary cells in chromatin accessibility. These factors should be considered when mapping respiratory disease loci. Further work identifying peak enrichment in fine-mapping regions is ongoing.

2282F

Phenotypic clustering reveals distinct subtypes of polycystic ovary syndrome with novel genetic associations. M. Dapas¹, F.T.J. Lin¹, R. Sisk¹, R.S. Legro², M. Urbanek^{1,3,4}, M.G. Hayes^{1,4,5}, A. Dunai⁶. 1) Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL; 2) Department of Obstetrics and Gynecology, Penn State College of Medicine, Hershey, PA; 3) Center for Genetic Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL; 4) Center for Reproductive Science, Northwestern University Feinberg School of Medicine, Chicago, IL; 5) Department of Anthropology, Northwestern University, Evanston, IL; 6) Division of Endocrinology, Diabetes & Bone Disease, Icahn School of Medicine at Mount Sinai, New York, NY.

Understanding the genetic architecture of complex diseases is a central challenge in human genetics. Often defined according to arbitrary diagnostic criteria, complex diseases can represent the phenotypic convergence of numerous genetic etiologies. Identifying and addressing genetic heterogeneity in complex diseases could increase power to detect causal variants and improve treatment efficacy. Polycystic ovary syndrome (PCOS) is a complex disease diagnosed by the presence of at least two of its three cardinal reproductive features: hyperandrogenism, chronic anovulation, and polycystic ovarian morphology. Accordingly, there are four different affected phenotypes with a net prevalence of 15% in women of reproductive age. PCOS is also frequently associated with insulin resistance, obesity, and dysglycemia. This phenotypic variation may reflect an underlying genetic heterogeneity, which would limit findings from genome-wide association studies (GWAS). We performed this study test the hypothesis that there are subtypes of PCOS with distinct genetic architectures. Using data from 932 cases from our previously published GWAS (Hayes et al., 2015), we performed unsupervised hierarchical clustering on eight adjusted quantitative traits: testosterone, dehydroepiandrosterone sulfate (DHEAS), insulin, glucose, BMI, sex hormone binding globulin (SHBG), luteinizing hormone (LH), and follicle-stimulating hormone (FSH). The clustering revealed two distinct phenotypic subtypes: 1) a "reproductive" group (23.4%) characterized by higher LH and SHBG levels with relatively low BMI and insulin levels and 2) a "metabolic" group (37.4%) characterized by higher BMI and glucose and insulin levels with relatively low SHBG and LH levels. We then repeated the GWAS limiting the cases to either the reproductive or metabolic subtype. We identified alleles in five novel loci that were associated with the reproductive PCOS subtype at genome-wide significance (*PRDM2/KAZN1*, $P=3.3 \times 10^{-10}$; *IQCA1/ACKR3*, $P=1.4 \times 10^{-9}$; *BMPR1B/UNC5C*, $P=4.6 \times 10^{-9}$; *CDH10*, $P=1.2 \times 10^{-8}$; *ANXA6*, $P=1.2 \times 10^{-8}$) and one locus that was significantly associated with the metabolic subtype (*KCNH7/FIGN*, $P=1.0 \times 10^{-8}$). These findings support the hypothesis that there are reproductive and metabolic subtypes of PCOS with distinct genetic architectures and demonstrate how precise phenotypic delineation can be more powerful than increases in sample size for genetic association studies.

2283W

Characterization of the genomic and gene expression of *SEMA3D* in multifactorial Hirschsprung disease. . Gunadi¹, NYP. Budi¹, K. Iskandar^{1,2}, A. Makhmudi¹. 1) Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia; 2) UGM Academic Hospital.

Background: Recently, the *semaphorin 3D (SEMA3D)* gene has been implicated in intestinal aganglionosis. We aimed to characterize the genomic and gene expression of *SEMA3D* in the development of Hirschsprung disease (HSCR) in Indonesia. **Methods:** We performed Sanger sequencing on 54 HSCR patients to find pathogenic variants in *SEMA3D*. Next, we determined *SEMA3D* expression in the colon of 18 HSCR patients and 13 non-HSCR controls by quantitative real-time polymerase chain reaction (RT-PCR). **Results:** Sanger sequencing showed no rare variants in the *SEMA3D*, but one common variant in exon 17 (rs7800072). The risk allele frequency (C) at rs7800072 among HSCR patients was 22.5%. Quantitative RT-PCR showed that the *SEMA3D* expression was strongly up-regulated (5.5-fold) in the ganglionic colon of HSCR patients compared to control colon (10.8 ± 2.1 vs. 13.3 ± 3.9 ; $p=0.034$). **Conclusion:** *SEMA3D* gene expression, but not *SEMA3D* rare variant, might contribute to the development of HSCR in Indonesia. In addition, this study is the first report of *SEMA3D* investigation in Asian ancestry.

2284T

Whole genome sequence analysis of pulmonary function and COPD in >16,000 multi-ethnic participants of the NHLBI trans-omics for precision medicine (TOPMed) program identifies new associated loci. A.W. Manichaikul¹, X. Zhao², D. Qiao³, C. Yang⁴, J.N. Nguyen¹, P. Sakornsakolpat¹, D. Prokopenko⁴, E.C. Oelsner⁵, P. Balte⁵, B. Yu⁶, L. Lange⁷, S.J. London⁸, J. Dupuis⁹, G. O'Connor¹⁰, R. Reed¹¹, B. Cade¹², S. Gharib¹³, M. Daya¹, B. Psaty¹⁴, S. Redline¹², K.C. Barnes¹, S.S. Rich¹, G.R. Abecasis², E.K. Silverman¹, R.G. Barr⁵, M.H. Cho⁴, TOPMed PFT, COPD and Lung Working Groups. 1) Center for Public Health Genomics, University of Virginia, Charlottesville, VA; 2) Department of Biostatistics, University of Michigan; 3) Department of Medicine, Brigham and Women's Hospital; 4) Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School; 5) Department of Medicine, Columbia University; 6) Department of Epidemiology, Human Genetics & Environmental Sciences, UTHealth School of Public Health; 7) Division of Biomedical Informatics and Personalized Medicine, Department of Medicine, University of Colorado School of Medicine Anschutz Medical Campus; 8) Epidemiology Branch, National Institute of Environmental Health Sciences; 9) Department of Biostatistics, Boston University School of Public Health; 10) Department of Medicine, Boston University School of Medicine; 11) Department of Medicine, University of Maryland; 12) Division of Sleep and Circadian Disorders, Department of Medicine, Brigham and Women's Hospital; 13) Division of Pulmonary, Critical Care and Sleep Medicine, University of Washington; 14) Departments of Epidemiology and Medicine, University of Washington.

Chronic obstructive pulmonary disease (COPD), the third leading cause of death in the United States, is diagnosed by a decrease in lung function, namely forced expiratory volume in one second (FEV1) and its ratio to forced vital capacity (FEV1/FVC). Genome-wide association studies (GWAS) of lung function and COPD have already identified over 100 loci associated with one or more of these measures and conditions. We performed whole genome sequence (WGS) analysis of lung function traits (FEV1, FEV1/FVC and FVC) and COPD in the NHLBI Trans-Omics for Precision Medicine (TOPMed) Program to examine evidence for rare and/or common trait-associated variants that were not identified in previous GWAS. WGS had deep (~30X) coverage with joint-sample variant calling and comprehensive variant level quality control in >50,000 TOPMed samples (freeze 5b). Lung function measures and COPD were analyzed in a subset of 8,332 subjects from four population-based and two family-based studies, as well as 8,471 individuals from the COPD-ascertained COPDGene study. These samples included 4,232 moderate-to-severe COPD cases out of which 1,644 had severe COPD. We conducted pooled analyses combining White (n=11,667), African Americans (n=3,912), Hispanic (n=711) and Asian (n=515), in addition to race/ethnic stratified analyses, with covariate adjustment for age, sex, height, weight, current and former smoking, pack-years of smoking, PCs of ancestry and sequencing center. After removing variants with Heterozygosity Count < 30, single variant quantitative trait analysis of population- and family-based samples identified novel associations with FVC in Whites near *GALR1* (minor allele frequency [MAF]=0.39; $P=3 \times 10^{-8}$) and with FEV1 in African Americans near *LGALS8* (MAF=0.04; $P=1.5 \times 10^{-8}$). In combined analysis across population-based and COPD-ascertained samples, we identified novel loci for FEV1/FVC near *ZNF462* (MAF=0.39; $P=3.2 \times 10^{-8}$) in pooled analysis of all race/ethnic groups, and for FEV1 near *GALNT18* (MAF=0.30; $P=4.3 \times 10^{-8}$) and FVC near *CMIP* (MAF=0.13; $P=1.3 \times 10^{-8}$) in stratified analysis of African Americans. We further identified a novel association for moderate-to-severe COPD in pooled analysis of all race/ethnic groups near *GRK7* (MAF=0.006; $P=3.4 \times 10^{-8}$). Replication of these findings will be needed. Our study thus identified multiple novel signals for lung function in African Americans and multi-ethnic samples, demonstrating the benefits of dense coverage across the genome by WGS. .

2285F

Clonal expansion and diversification of cancer-associated mutations in endometriosis and normal endometrium. H. Nakaoka¹, K. Suda², K. Yoshihara², T. Enomoto², I. Inoue¹. 1) National Institute of Genetics, Mishima, Japan; 2) Department of Obstetrics and Gynecology, Niigata University Graduate School of Medical and Dental Sciences, Niigata, Japan.

Endometriosis is characterized by ectopic endometrial-like epithelium and stroma. A more complete molecular characterization of the pathogenesis of endometriosis is to develop better treatments. Whole-exome and target-gene sequencing were performed for 107 ovarian endometriotic epithelium and 82 normal uterine endometrial epithelium samples isolated from frozen sections by laser-microdissection. In both normal and endometriotic epithelium samples, numerous somatic mutations were identified within genes frequently mutated in endometriosis-associated cancers. Analysis of mutant allele frequency (MAF) in endometriotic epithelium revealed a unique mutational landscape, wherein *KRAS* was frequently mutated, with a higher MAF accompanied by arm-level copy number alterations, leading to clonal expansion of cells carrying these changes. MAF distributions in normal endometrium were lower than in endometriotic epithelium. To show that each endometrial gland was clonal in origin, we prepared and sequenced single endometrial glands and found that each gland carried a clonal state of distinct mutations, proving how heterogeneous and mosaic-like the cellular and genomic architecture of endometrial epithelium is. The identified somatic mutations can be the pathogenic footprints showing a lineage from normal uterine epithelium to endometriosis to subtypes of ovarian cancers, corroborating with previous epidemiological evidence. Remarkable increases in MAF of *KRAS* mutations in endometriotic epithelium suggests that retrograde flow of endometrial cells already harboring *KRAS* mutations have selective advantages at ectopic sites, leading to the development of endometriosis. Our genomic study supports Sampson's century-old retrograde hypothesis of the origin of endometriosis.

2286W

Primary sclerosing cholangitis, a common co-morbidity in inflammatory bowel disease. J. Koskela^{1,2,3}, M.D. Voskuil⁴, M. Rivas⁵, B. Avila^{2,6}, C. Stevens^{2,6}, S. Ripatti⁷, A.N. Ananthakrishnan⁷, R. Xavier⁸, H. Soko⁶, R.K. Weersma⁴, M. Färkkilä³, M.J. Daly^{2,6}. 1) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Finland, Helsinki, Uusimaa, Finland; 2) Broad Institute of MIT and Harvard, Cambridge, MA, USA; 3) Helsinki University and Helsinki University Hospital, Clinic of Gastroenterology Helsinki, Finland; 4) Department of Gastroenterology and Hepatology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands; 5) Department of Biomedical Data Science, Stanford University; 6) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; 7) Massachusetts General Hospital and Harvard medical school, Boston, MA, USA; 8) Department of Gastroenterology, Saint Antoine Hospital, APHP, Paris, France.

Primary sclerosing cholangitis (PSC) is a rare and progressive inflammatory disease of the bile ducts with no effective therapy currently available. While the prevalence of PSC has been reported only at 15.3-32/100,000, PSC is the most common cause of liver transplantations in Nordic countries. Additionally, PSC patients are at high risk of cholangiocarcinoma (up to 1000x) compared to population control and substantial risk of colorectal cancer. PSC is commonly associated with inflammatory bowel disease (IBD) as ~70% of PSC-patients present with PSC-IBD. When co-occurring with IBD, PSC is most often (80% of cases) associated with ulcerative colitis. So far it is unclear whether PSC-IBD is an independent entity or whether it only manifests as an IBD subtype, however up to 8% of patients with IBD have diagnostic findings in MRI. A recent (and largest) genome wide association study using imputed array data of 4,796 PSC cases estimated the genetic correlation of IBD and PSC to 29%, while confirming most previously reported variants and identifying 4 new loci, as up to 19 loci have now been associated with PSC. We are currently analyzing the largest whole exome sequencing (WES) study on PSC, aiming at meta-analyzing > 1000 patients of European and Finnish ancestry. So far, we have finished the analysis of 511 Finnish samples and 8913 population controls. These PSC patients already represent more than 10% of total estimated Finnish PSC population. In addition to DNA, deep phenotyping and health status information, sequential endoscopic retrograde cholangiographic examinations (ERC), biliary cytology, bile samples have been collected from the Finnish patients. Out of the 19 loci described so far, we identified nominally significant ($p < 0.05$) associations in 8 genes reported earlier and in 12 loci within a 500kb window of the previously implicated gene ($p < 0.01$). After excluding significant associations from a large WES study on IBD (unpublished), we confirm a strong MHC component and two suggestive loci specific to PSC in *3p21.31* with protective intronic variant on *RNF123* (OR=0.7, case MAF 20%, $p=3.22e-5$) and a risk-carrying missense variant on *QRICH1* (OR=1.6, case MAF 12%, $p=9.19e-6$), both regions being described associated with IBD previously. Interestingly, *RNF123* (probability of LoF intolerance = 0.97) is a neighbor gene to *MST1*, a previously described PSC risk locus. The Finnish results are available at: <http://35.233.15.189/pheno/PSC.txt>.

2287T

Multi-omic analysis of discordant and concordant sib-pairs with inflammatory bowel disease. A.B. Stienke¹, S.R. Brant^{2,3}, C.L. Simpson¹. 1) Department of Genetics, Genomics, and Informatics, University of Tennessee Health and Science Center, Memphis, TN; 2) Department of Medicine, Department of Genetics and the Human Genetics Institute of New Jersey, Rutgers, The State University of New Jersey, New Brunswick and Piscataway, NJ; 3) Meyerhoff Inflammatory Bowel Disease Center, Department of Medicine, School of Medicine and Department of Epidemiology, School of Public Health, Johns Hopkins University, Baltimore, MD.

Background: Inflammatory bowel disease (IBD) is an immune-mediated chronic intestinal disorder that is typically divided into two distinct types; ulcerative colitis (UC) and Crohn's disease (CD). Over 240 IBD susceptibility loci have been identified however, much of the etiology remains unexplained and even within families the disease can have a heterogeneous clinical presentation. Here we present a study attempting to discern the cause of that heterogeneity in discordant and concordant sib-pairs by comparing whole exome sequencing, epigenetic differences using DNA methylation analysis and gene expression using RNAseq. **Methods:** A total of 96 discordant and concordant sib-pairs were included in the study. Genomic DNA and RNA were extracted from peripheral blood mononuclear cells that were isolated using cell sorting from whole blood. DNA methylation values were measured using the Illumina EPIC array. Whole exome sequencing was performed using the Illumina TruSeq Exome kit, and RNAseq was performed using the Illumina RNA Kit v2. **Results:** Current results have identified over 5000 variants that are being filtered and prioritized for further analysis and cross comparison to additional datasets, as well as employing integrative approaches. Additional results will be presented. **Conclusion:** Combining data from multiple technologies is an important next step for interpreting the results of genome-wide association studies (GWAS) and there is currently no agreement on the best approach for integrating these data. Given the known genetic and clinical heterogeneity in IBD, using discordant and concordant sib-pairs attempts to leverage the expected sharing between the sibs as supporting information to inform future studies.

2288F

Genetic association between human leukocyte antigen (HLA)-DR, -DQ, -DP and gallstone disease in Han Chinese. H. Yang¹, P. Yang^{2,3,4,5}, C. Tsai^{2,4}, C. Tsai^{2,4}, T. Liu^{2,4}, K. Ho^{2,4}, J. Lee^{2,3,4}, M. Li^{2,4}, D. Chu^{2,4}, F. Lee^{2,4}, H. La^{2,4}, S. Shyr⁶, S. Lee⁶, C. Chung⁶, Y. Wu⁶, P. Lin¹, Y. Lien¹, C. Lin¹, M. Lin¹, S. Shih^{3,4,7}, Y. Lee^{1,8,9,10}. 1) Department of Medical Research, MacKay Memorial Hospital, New Taipei City, Taiwan; 2) Department of General Surgery, MacKay Memorial Hospital, Taipei, Taiwan; 3) Department of Medicine, MacKay Medical College, New Taipei City, Taiwan; 4) Department of Nursing, MacKay Junior College of Medicine, Nursing, and Management, Taipei, Taiwan; 5) Department of Pharmacology and Graduate Institute of Medical Sciences, Taipei Medical University, Taipei, Taiwan; 6) Department of Nursing, MacKay Memorial Hospital, Taipei, Taiwan; 7) Division of Gastroenterology, Department of Internal Medicine, MacKay Memorial Hospital, Taipei, Taiwan; 8) Department of Pediatric Endocrinology, MacKay Children's Hospital, Taipei, Taiwan; 9) Institute of Biomedical Sciences, MacKay Medical College, New Taipei City, Taiwan; 10) Department of Pediatrics, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan.

Background and Goal: Biliary calculi in gallbladder induce latent and persist inflammatory responses and related multiple symptoms called gallstone disease (GSD). We know the pathological and environmental risk factors of GSD. However, the immune-regulation in GSD are still elusive. Human leukocyte antigens (HLA) are correlated with many gastrointestinal disorders and expressed in gallbladder epithelial cells. In this study, we investigated whether the HLA will link to GSD susceptibility. **Methods:** Genomic DNA was isolated from 577 GSD patients and 371 gallstone-free controls. HLA-DRB1, -DQB1, -DPB1 sequence variants were typed using Sequence-Based Typing and identified by IMGT database alignment. Differences in the frequencies of these alleles between GSD patients and control subjects were evaluated using the chi-squared test with Yates' correction and Fisher's exact test where appropriate. The odds ratio (OR) and 95% confidence interval (95% CI) for each allele were calculated. Haplotype and Hardy-Weinberg Equilibrium test were evaluated by Pypop 0.7.0. **Results:** We found the associations between GSD and HLA compared to controls. HLA-DRB1*12:02 (OR=1.36, $P=0.04$), -DRB1*12:01 (OR=1.68, $P=0.02$), -DRB1*16:02 (OR=0.60, $P=0.01$). Serotype of DRB1*02 (OR=1.49, $P=0.002$); DRB1*16 (OR=0.59, $P=0.009$). HLA-DQB1*03:01 (OR=1.30, $P=0.01$). Haplotype HLA-DRB1*12:01-DQB1*03:01 (OR=1.70, $P=0.02$). DRB1*16:02-DQB1*05:02 (OR=0.61, $P=0.01$). DRB1*12:01-DPB1*05:01 (OR=2.29, $P=0.01$). DRB1*16:02-DPB1*05:01 (OR=0.51, $P=0.006$). DRB1*12:01-DQB1*03:01-DPB1*05:01 (OR=2.22, $P=0.02$). DRB1*16:02-DQB1*05:02-DPB1*05:01 (OR=0.57, $P=0.02$). **Conclusions:** Our finding is the first report in specific HLA correlate to GSD. Our data facilitated an evaluation of the effects of individual HLA alleles consisting of disease-prone/disease-resistant.

2289W

Association between the *SLCO1B* locus and alcohol-associated chronic pancreatitis. T. Nagpal¹, P. Greer¹, C. Shelton^{1,2}, D. Yadav¹, D.C. Whitcomb^{1,2}, NAPS2 Consortium. 1) Gastroenterology, Hepatology and Nutrition, University of Pittsburgh, Pittsburgh, PA; 2) Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA.

Background: Chronic pancreatitis (CP) is an acquired, complex, progressive, fibro-inflammatory disorder of the pancreas characterized by variable atrophy, fibrosis, pain, pancreatic exocrine and endocrine insufficiency and dysplasia. Multiple pathogenic genetic variants were originally identified by linkage (*CFTR*, *PRSS1*) and candidate gene studies (*SPINK1*, *CTRC*, *CPA1*, *CASR*). We conducted the first 2-stage GWAS using the North American Pancreatitis Study 2 (NAPS2) consortium cohort (Whitcomb, *Nature Genetics*. 2012) with 676 cases and 4,507 controls, followed by 910 cases and 4,170 controls to identify a protective *PRSS1-PRSS2* locus and *CLDN2* risk locus. Additional loci with suggestive association were identified, including loci at 12p12.1–12p12.2 ($P < 10^{-5}$). **Aim:** To expand the NAPS2 cohort of pancreatitis patients and identify potential pancreatitis risk alleles on chromosome 12. **Methods:** Well-phenotyped patients with recurrent acute pancreatitis (RAP) and CP were recruited into the NAPS2 cohort. 25 common SNPs that spanned the region of interest across 12p12.1–12p12.2 were genotyped. The allele frequencies in alcohol-CP/RAP ($n = 433$) vs non-alcohol-CP/RAP ($n = 762$) cases were compared using Fisher's exact test. Pancreatic tissue samples that underwent RNA-Seq were examined for expression of candidate genes. **Results:** 8 out of the 25 variants are in linkage disequilibrium ($LD > 0.8$) and form a haplotype block with lead SNPs rs11045681 and rs11045699 ($LD = 1$). Both SNPs are deleterious missense mutations in *SLCO1B7* (SIFT scores of 0 and 0.01) and negatively associated with alcohol-related pancreatitis ($P = 6.3 \times 10^{-6}$ and $OR = 0.61$ for rs11045681; $P = 3.1 \times 10^{-6}$ and $OR = 0.60$ for rs11045699). These genes produce organic anion transporter proteins and are mainly expressed in the human liver. Specifically, *SLCO1B1* and *SLCO1B3* uptake complex bile acids, hormones, etc. from the blood into the liver. RNA-Seq identified expression of *SLCO1B3* transcript in pancreatic tissue from patient with RAP. **Conclusion:** The *SLCO1B3-SLCO1B1-SLCO1B7* haplotype was underrepresented in alcoholic pancreatitis patients and may be protective against alcohol-related pancreatitis. We hypothesize that expression of these transporters in the regenerating pancreas after an acute pancreatitis attack may increase toxicity in pancreatic cells. Lack of ectopic expression, as expected from this haplotype, results in protection from the exposure to alcohol and tobacco in alcoholic CP.

2290T

Discovery of novel genetic loci for kidney traits using whole genome sequencing: The trans-omics for precision medicine (TOPMed) project. B.M. Lin¹, L. Raffield², J. Brody³, J.A. Perry⁴, D.Y. Lin¹, N. Franceschini⁵, TOPMed Kidney Working Group. 1) Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC; 3) Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC; 4) Cardiovascular Health Research Unit, University of Washington, Seattle, WA; 5) Department of Medicine, University of Maryland, Baltimore, MD.

Chronic kidney disease (CKD) is a common but frequently unrecognized medical condition that is associated with high morbidity and mortality. Presence of CKD is defined by low estimated glomerular filtration rate (eGFR), a heritable measure of kidney function. To identify genetic loci for eGFR, we conducted whole genome sequencing analysis of 23,732 participants of 10 multi-ethnic (European, African, East Asian, and Hispanic) cohort studies within the NHLBI TOPMed Project. An extensive, centralized quality control was performed in the sequencing data. Calibrated serum creatinine was used to estimate eGFR, which was inverse normalized within study and ethnic groups and rescaled to recover original trait variances. We applied linear mixed models adjusted for age, sex, study, and ethnicity; the variance components were modeled by the genetic relationship matrix estimated from the whole genome sequencing data. In single-variant tests with minor allele counts > 10 , there was no evidence of inflation ($\lambda = 1.009$). Four loci were significantly associated with eGFR at the genome-wide significance threshold of $P = 5.0 \times 10^{-8}$, including three novel loci (*NBP7*, rs184708696, MAF 0.04%, $P = 2.95 \times 10^{-8}$; *BOLL*, rs375059924, MAF 0.07%, $P = 4.57 \times 10^{-10}$; and *FAM46A*, rs143018620, MAF 0.09%, $P = 2.13 \times 10^{-9}$). At the *FAM46A* locus, the signal extended for 94.2 kb upstream of the gene and was comprised of several highly correlated low frequency variants present in individuals of East Asian ancestry ($n = 2,208$). Compared to individuals with the rs143018620 reference allele, each additional copy of the minor allele increased eGFR by 8.9 ml/min/1.73m² (standard error = 1.5 ml/min/1.73m²). Several signals that almost reached genome-wide significance were located in known eGFR loci, including *GATM*, *GCKR*, *BCAS3*, and *PDILT/UMOD*. Gene-based burden tests of loss of function variants with MAF $< 1\%$ did not reach genome-wide significance. In conclusion, analyses of deep sequenced TOPMed whole genome sequencing in multi-ethnic studies have identified three novel genome-wide significant loci for eGFR, including a rare, long-range haplotype specific to East Asians that confers a protective effect for kidney function.

2291F

Genetic analysis of 400 patients with aHUS refines understanding and implicates a new gene in this disease. *F. Bu^{1,2}, Y. Zhang¹, K. Wang³, N. Ghiringhelli Borsa¹, M. Jones¹, A. Taylor¹, E. Takanami¹, N. Meyer¹, K. Frees¹, C. Thomas¹, C. Nester^{1,4}, R. Smith^{1,4}.* 1) Molecular Otolaryngology and Renal Research Laboratories, University of Iowa, Iowa City, IA, USA; 2) Medical Genetics Center, Southwest Hospital, Chongqing, China; 3) College of Public Health, University of Iowa, Iowa City, IA, USA; 4) Divisions of Nephrology, Department of Internal Medicine and Pediatrics, Carver College of Medicine, University of Iowa, Iowa City, IA, USA.

Atypical hemolytic uremic syndrome (aHUS) is a life-threatening thrombotic microangiopathy (TMA) that primarily affects the kidney although other organ involvement can be seen. Dysregulation of the alternative pathway of complement at the endothelial cell surface triggered by a variety of events has been implicated in disease pathogenesis. Genetic variation in complement genes and autoantibody against complement proteins are known to be predisposing factors, however the interpretation of variant effect in known aHUS genes can be ambiguous. In addition, in half of patients genetic/autoimmune causality is unclear. In this study, we sequenced and analyzed 93 TMA-related genes in 516 aHUS patients. 400 patients passed controls for sequencing quality, population stratification and relatedness. 63,345 non-Finnish European subjects from the Genome Aggregation Database and 600 in-house healthy subjects were used as two control groups. We applied the Fisher's exact, modified Poisson exact and SKAT-O tests to show that aHUS patients are enriched for novel/ultra-rare variants in the *CFH*, *C3*, *CD46*, *CFI* and *DGKE* genes. We show that variants accumulate in featured domains of *CFH*, *CD46* and *C3*, and using a sliding-window analysis, propose a minor allele frequency of 0.1% as the threshold above which a variant in known aHUS genes should not be considered pathogenic unless significant association is reported. Based on this threshold and a literature review, we reclassify 20 variants previously implicated in aHUS as non-pathogenic. In addition, we show that aHUS patients are not enriched for multiple causal variants. Finally, we implicate *VTN*, which encodes an inhibitor of the terminal complement pathway, as an aHUS-associated gene.

2292W

Genome-wide association meta-analysis highlights the role of cholesterol and bile acid homeostasis in gallstone disease. *E. Ferkingstad¹, A. Oddsson¹, S. Gretarsdottir¹, S. Benonisdottir¹, G. Thorleifsson¹, A.M. Deaton¹, O.A. Stefansson¹, G.L. Norddahl¹, F. Zink¹, G.A. Arnadottir¹, B. Gunnarsson¹, G.H. Halldorsson¹, A. Helgadottir^{1,2}, B.O. Jensson¹, R.P. Kristjansson¹, G. Sveinbjornsson¹, D.A. Sverrisson¹, G. Masson¹, H. Holm¹, I. Jonsdottir^{1,2,3}, S. Olafsson⁴, T. Steingrimsdottir^{2,5}, T. Rafnar¹, E.S. Bjornsson^{2,4}, U. Thorsteinsdottir^{1,2}, D.F. Gudbjartsson^{1,6}, P. Sulem¹, K. Stefansson^{1,2}.* 1) deCODE genetics/Amgen, Inc., Reykjavik, Iceland; 2) Faculty of Medicine, University of Iceland, Reykjavik, Iceland; 3) Department of Immunology, Landspítali University Hospital, Reykjavik, Iceland; 4) Department of Internal Medicine, Landspítali University Hospital, Reykjavik, Iceland; 5) Department of Obstetrics and Gynecology, Landspítali University Hospital, Reykjavik, Iceland; 6) School of Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland.

Gallstones are responsible for one of the common diseases in the Western world. Due to the threat of complications, such as acute pancreatitis, it is commonly treated with cholecystectomy. We performed a meta-analysis of two genome-wide association studies of gallstone disease from Iceland and the UK, totaling 27,174 cases and 736,838 controls, uncovering 21 novel variants at 20 loci that associate with the disease. Two distinct low frequency missense variants in a gene involved in bile acid transport associate with an increased risk of gallstone disease. Based on these associations and reported functional studies, we conclude that lower bile acid transport is accompanied by greater risk of gallstone disease. Two other low frequency missense variants in other genes also associate with an increased gallstone disease risk. Additionally, we describe 17 novel common variants that associate with gallstone disease. Taken together with replication of previously reported variants, this highlights the role of cholesterol metabolism and the enterohepatic circulation of bile acids in gallstone disease.

2293T

Phenome-wide association study of polycystic ovary syndrome using polygenic risk prediction.

Y.Y. Joo¹, K. Actkins², J.A. Pacheco¹, T.L. Walunas³, J. Harley⁴, H. Hakonarson⁴, F.D. Mentch⁴, K. Ho⁵, S. Pendergrass⁵, A.O. Basile⁶, M.D. Ritchie⁷, S.J. Hebbinger⁸, D.R. Crosslin⁹, A.S. Gordon⁹, G.P. Jarvik⁹, I.B. Stanaway⁹, D.R. Velez Edwards², R.J. Carroll², J.C. Denny², A.E. Dunai¹⁰, M. Urbanek¹, M.E. Smith¹, R.L. Chisholm¹, A.N. Kho¹, L. Davis², M.G. Hayes¹. 1) Northwestern University Feinberg School of Medicine, Chicago, IL; 2) Vanderbilt University Medical Center, Nashville, TN; 3) Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio; 4) Children's Hospital of Philadelphia, Philadelphia, PA; 5) Geisinger, Danville, PA; 6) Pennsylvania State University, State College, PA; 7) University of Pennsylvania, Philadelphia, PA; 8) Marshfield Clinic Research Institute, Marshfield, WI; 9) University of Washington Medical Center, Seattle, WA; 10) Icahn School of Medicine at Mount Sinai, New York, NY.

Polycystic ovary syndrome (PCOS) is the most common endocrine disorder affecting 5-15% of reproductive age women worldwide. It is a leading risk factor for infertility and type 2 diabetes in females, but its genetic etiology is highly complex hindering diagnosis and treatment in clinical practice. To date, Genome-Wide Association Studies (GWAS) of PCOS have identified 30 genome-wide significant variants spanning 20 unique loci, although these account for a limited proportion of the estimated heritability. Leveraging integrated phenotype and genotype data of 83,717 multi-ethnic individuals from the eMERGE (electronic MEDical Records and GENomics) network, we constructed a polygenic risk score (PRS) for PCOS by summing the allele dosage of the susceptibility variants in each individual weighted by the reported effect size. The same phenome-wide association study (PheWAS) was performed in an independent biobank sample (BioVU, n=18,370) to replicate our finding. We also performed comparative PheWAS using individual risk variants as well as the PRS to identify unique risk factors with high pleiotropy and/or comorbidity with PCOS. We identified several clinical phenotypes associated with the PRS for PCOS approaching phenome-wide significance ($P < 2.5 \times 10^{-5}$) that may be either symptomatic or cryptogenic, including neoplasms and reproductive diseases such as 'Malignant neoplasm of renal pelvis' ($P = 1.44 \times 10^{-4}$, OR=0.88), 'Polycystic ovaries' (2.04×10^{-4} , OR=1.08), 'Viral hepatitis C' (5.48×10^{-4} , OR=0.96), and 'Ovarian dysfunction' (1.01×10^{-3} , OR=1.06). The phenotype-PRS association of 'Polycystic ovaries' ($P = 0.047$, OR=1.21) was confirmed and remained significant in our replication sample. In the single variant PheWAS, two phenotypes reached phenome-wide significance: 'Neuralgia, neuritis, and radiculitis NOS' (rs2272046, *HMG2*, $P = 9.75 \times 10^{-5}$, OR=1.36) and 'Type 1 diabetes with renal manifestations' (rs705702, *SUOX*, $P = 2.30 \times 10^{-5}$, OR=1.27). Several of these associations were previously unknown, thereby providing new insight into the etiology of PCOS. Our findings demonstrate that individual genome-phenome data available from electronic health records can provide opportunities to stratify patients' genetic risk of PCOS and discover the potential shared biology between PCOS and other diseases.

2294F

Genetic variation in the *SIM1* locus is associated with erectile dysfunction.

E. Jorgenson¹, N. Matharu^{2,3}, M.R. Palmer⁴, J. Yin¹, J. Shan¹, T.J. Hoffmann^{2,5}, K.K. Thai¹, X. Zhou^{2,3}, J.M. Hotaling⁶, G. Jarvik⁴, N. Ahituv^{2,3}, H. Wessells⁷, S.K. Van Den Eeden¹. 1) DOR, Kaiser Permanente, Oakland, CA; 2) Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco, California, 94158, USA; 3) Institute for Human Genetics, University of California San Francisco, San Francisco, California, 94158, USA; 4) Division of Medical Genetics, University of Washington School of Medicine, Seattle, Washington, 98195, USA; 5) Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, California, 94158, USA; 6) Department of Surgery (Urology), University of Utah School of Medicine, Salt Lake City, Utah, 84132, USA; 7) Department of Urology, University of Washington School of Medicine, Seattle, Washington, 98195, USA.

Erectile dysfunction affects millions of men worldwide. Twin studies support the role of genetic risk factors underlying erectile dysfunction, but no specific genetic variants have been identified for erectile dysfunction. We conducted the first large-scale genome-wide association study (GWAS) of erectile dysfunction in 36,649 men in the multiethnic Kaiser Permanente Northern California GERA cohort. We also undertook replication analyses in 222,358 men from the U.K. Biobank. In the discovery cohort, we identified a single locus on chromosome 6 near the single-minded family bHLH transcription factor 1 (*SIM1*) gene that was significantly associated with the risk of erectile dysfunction (rs17185536-T, Odds Ratio = 1.26, $p = 3.4 \times 10^{-26}$). The association replicated in the U.K. Biobank sample (Odds Ratio = 1.25, $p = 6.8 \times 10^{-14}$), and the effect is independent of known erectile dysfunction risk factors, including body mass index (BMI). The risk locus resides on the same topologically associating domain as *SIM1*, interacts with the *SIM1* promoter, and the rs17185536-T risk allele and two other associated alleles in the locus showed differential enhancer activity in luciferase assays. *SIM1* is part of the leptin-melanocortin system, which has an established role in body weight homeostasis and sexual function. Because the variants associated with erectile dysfunction are not associated with differences in BMI, our findings suggest a mechanism that is specific to sexual function.

2295W

Genome-wide association study of lipedema phenotype in the UK Biobank. Y.C. Klimentidis, K.L. Herbst, Z. Chen. University of Arizona, Tucson, AZ.

Lipedema is a common disorder characterized by excessive deposition of subcutaneous adipose tissue (SAT) in the legs and hips, occurring mainly in women after puberty. It often results in pain and impaired mobility. Although it may be relatively common, the etiology of lipedema is poorly understood, and no specific genes have yet been linked to it. We used data on approximately 250,000 women from the UK Biobank to identify those with and without a lipedema phenotype. Measures of segmental body composition from bioelectric impedance analysis, hip and waist circumferences, along with information on pain in legs were used to define cases and controls. Genome-wide association analyses were performed to identify genetic risk factors for lipedema, exploring a range of different case/control definitions and prevalence estimates. We identified several loci, including *CCDC92*, *COBLL1*, and *LYPLAL1* as being consistently associated with the lipedema phenotype. Using a gene-based analysis, we also identified the *HMGCS1* gene. Follow-up analyses suggest an enrichment of genes expressed in SAT. Our findings point to specific genetic risk factors for lipedema, some of which have previously been identified as sex-differentiated loci for waist-to-hip ratio, and provide a starting point for better understanding the pathophysiology of lipedema.

2296T

Identification of genetic loci related to circulating reproductive hormone levels by GWAS in Japanese men. Y. Sato¹, A. Tajima², S. Kogusuri¹, A. Fujii¹, T. Sato², I. Imoto³, T. Iwamoto⁴. 1) Graduate School of Biomedical Sciences, Tokushima University, Tokushima, Japan; 2) Graduate School of Advanced Preventive Medical Sciences, Kanazawa University; 3) Aichi Cancer Center Research Institute; 4) Reproduction Center, Sanno Hospital.

BACKGROUND: Genetic contributions to population-level variations in circulating reproductive hormone levels are expected; however, the genetic determinants of reproductive hormone levels remain poorly understood. In this study, to identify loci related to plasma inhibin B, luteinizing hormone (LH), follicle-stimulating hormone (FSH), testosterone, and sex hormone-binding globulin (SHBG) levels, we conducted two-stage genome-wide association study (GWAS) in Japanese men. Although previous genome-wide association studies (GWASs) of several reproductive hormones such as testosterone, dihydrotestosterone, SHBG, LH and FSH have been conducted in several populations, this is the first GWAS in individuals of Japanese nationality. **METHODS:** In the discovery stage, 811 men were genotyped using the Illumina Human Core V.1.0 DNA Analysis kit (Illumina, San Diego, CA, USA) and then imputation was performed using BEAGLE V.4.1. Association analysis between each single-nucleotide polymorphism (SNP) and hormone levels was performed using multiple linear regression analysis with age and body mass index as covariates. Subsequently, we performed a replication study using an additional 722 Japanese men. The results of the discovery and replication stages were combined in a meta-analysis. **RESULTS:** For all hormones, GWAS failed to reveal any SNPs showing genome-wide significance ($P < 5 \times 10^{-8}$) in the discovery stage. On setting a suggestive significance threshold (P of $< 5 \times 10^{-6}$) in the discovery stage, a total of 12 loci were found to be associated with inhibin B (five loci), LH (two loci), testosterone (four loci), and SHBG (one locus). We selected only the most significant SNP in each locus for replication genotyping. Only one SNP on chromosome 12 showed a significant association with inhibin B levels (Bonferroni-corrected $P < 0.05$) in the replication stage. When we combined the discovery and replication results in a meta-analysis, this SNP was nominally associated with plasma inhibin B levels ($P = 1.2 \times 10^{-7}$). **CONCLUSION:** In this GWAS, we identified one SNP on chromosome 12 that was nominally associated with plasma inhibin B levels in Japanese men, though we detected no loci related to the other four reproductive hormones.

2297F

Large scale GWAS meta-analyses unravel loci associated with albuminuria. A. Teumer^{1,2} on behalf of the CKDGen Consortium. 1) Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany; 2) DZHK (German Center for Cardiovascular Research), partner site Greifswald, Greifswald, Germany.

Elevated concentrations of albumin in the urine (albuminuria) are associated with increased risk of kidney disease progression, end-stage renal disease, and cardiovascular morbidity and mortality. Albuminuria is a hallmark of diabetic kidney disease (DKD), the leading cause of end-stage renal disease for which effective treatment is limited. Previous genome-wide association studies (GWAS) have identified very few loci at level of genome-wide significance (p -value $<5E-8$). To gain additional insights into the pathophysiological mechanisms underlying albuminuria, we conducted a meta-analysis of 54 GWAS of the urinary albumin-to-creatinine ratio (UACR), after inverse normal transformation of the age- and sex-adjusted residuals of log-transformed UACR, in 564,257 individuals of different ancestries, overall and stratified by diabetes. High-quality 1000 Genomes or HRC-imputed genotypes were available at up to 8 million genetic variants. In the overall sample, we identified 58 genomic 1-Mb regions containing at least one SNP associated with UACR at level of genome-wide significance, explaining 0.7% of the trait variance. Restricting the analysis to the 547,361 European ancestry (EA) individuals resulted in 61 genome-wide significant loci, whereas 56 overlapped with the trans-ethnic meta-analysis. Among the 47,816 EA individuals with diabetes, four loci were genome-wide significant, whereas two of them were specific for the diabetes sample. Fine-mapping of the EA associations revealed 63 independently associated SNPs. In the EA analysis, 18 regions contained credible sets of ≤ 10 SNPs with a cumulated 99% probability for including the causal variant underlying the GWAS signal (median region size 14 kb, IQR 5-43 kb). For several loci, colocalization of GWAS signals with gene expression in different tissues from the GTEx Project and from additional kidney resources highlighted potentially causal genes. This GWAS meta-analysis using a substantially increased sample size and state-of-the-art imputation panels revealed more than 60 loci associated with albuminuria, most of them novel. These results highlight novel pathways influencing albuminuria, and provide a list of candidate targets to improve treatment of kidney disease.

2298W

Whole exome sequencing analysis in biliary atresia: A follow-up study.

W.Y. Lam¹, J.S. Hsu², C.S. Tang¹, M.T. So¹, P.H.Y. Chung¹, P.K.H. Tam¹, M.M. Garcia-Barcelo¹. 1) Department of Surgery, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong; 2) Department of Psychiatry, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong.

Biliary Atresia (BA) is a severe and complex neonatal disorder characterized by aggressive fibrosclerosis and inflammatory obliteration of the extrahepatic biliary tract. The Southeast Asian population has the highest prevalence (1/5000 live births). The Kasai portoenterostomy is often performed to improve bile drainage, but the majority of the BA patients ultimately require liver transplant, making the disease the most common cause of pediatric liver transplant. Most BA cases are sporadic, and the etiology of BA remains unclear. In cholangiocytes, vascular endothelial growth factor A (VEGF-A) regulates cell proliferation during development, and mutations in VEGF were suggested to be associated with BA. Previously, we identified *ADD3* as a susceptibility factor for BA through genome-wide association analysis and showed that reduced *ADD3* expression levels correlated with the risk *ADD3* allele. In this follow-up study, we performed whole exome sequencing on 92 BA trios from the Southeast Asian population to identify potential variants, genes and biological pathways associated with BA. Our preliminary analysis found a de novo frameshift insertion in the *FLT1* gene, which encodes VEGF receptor 1 to regulate VEGFA signaling. *FLT1* is highly intolerant to loss-of-function mutation ($pLI=1.0$), and the insertion is extremely rare, not found in the ExAC or gnomAD databases. Further analysis results will be presented for additional insights on the pathogenesis of BA, which will guide our functional studies using in-house human organoid system. This project is supported by Li Ka Shing donation account - enhanced new staff start-up packages to Patrick H.Y. Chung.

2299T

Large scale Mendelian randomization scan reveal novel pathway linking education attainment and central adiposity to gastroesophageal reflux disease. *JS. Ong, J. An, M. Law, D. Whiteman, P. Gharahkhani, S. MacGregor.* QIMR Berghofer Medical Research Institute, 300 Herston Road, Brisbane, Australia.

Gastroesophageal reflux disease (GERD) is a common chronic disorder where affected individuals experience frequent acid reflux from the stomach into the esophagus. Previous work has shown GERD is related to complications in the upper digestive system and is strongly associated to risk of Barretts's esophagus and the rapidly fatal cancer esophageal adenocarcinoma. Apart from obesity, little is known about what actually causes GERD. We recently identified new loci implicated in this disease through a GWAS (72,334 cases and 263,352 controls) using data from both the UK Biobank and the Australian QSkin cohort. To determine which modifiable risk factors alter GERD risk, we performed a large scale Mendelian randomization (MR) analysis. MR is an instrumental variable technique based on using genetic variants as confounding-free instruments for exposures of interest. We evaluated the evidence for causality for a comprehensive set of possible risk factors for GERD. Odds ratios on GERD were estimated based on per SD change in the risk factor evaluated. We tested a series of anthropometric traits, fasting glucose, cholesterol levels, systolic blood pressure, vitamin D, coffee intake, alcohol consumption and education attainment; all were not associated with GERD except for traits related to obesity, height and education. We showed that BMI is causally associated with GERD, and for the first time show that education may be causally associated (1 SD lower years of schooling confers OR=1.09 for GERD, $p=2e-4$). We also found a suggestive association between white blood cell count (OR=1.006, $p=0.003$) and GERD, although this was not significant following multiple testing correction. Using instruments for anthropometric traits that are not associated with BMI, we also showed that decreased height (OR=1.02, $p=8e-15$) and more interestingly, higher BMI-adjusted waist-hip ratio (OR=1.08, $p=1e-10$) were associated with higher susceptibility to GERD. Our findings identify a novel causal link between adiposity (and specifically central adiposity) and GERD development. It is unclear whether the observed association between education attainment and GERD is mediated through BMI. At the meeting we will further present our sex-stratified findings and a phenotype-based stratification MR technique to dissect other potential mechanisms (stratified by obesity pathway) linking BMI and GERD.

2300F

Gut microbiome and irritable bowel syndrome with or without anxiety and depression symptoms. *K.H Bahk¹, H.A Park^{1,2}, J.H Sung¹.* 1) Genome Epidemiology, Department of Epidemiology, Graduate School of Public Health, Seoul National University, Seoul, Korea; 2) Department of Emergency Medicine, Hallym University Dongtan Sacred Heart Hospital, Gyeonggi, Korea.

BACKGROUND: Irritable bowel syndrome (IBS) is a common functional gastrointestinal disorder, but the risk factors of IBS are largely elusive. Recent evidence has suggested that in the development of IBS, similar to inflammatory bowel diseases, "gut-brain" axis and microbial dysbiosis might matter. We attempted to identify microbial characteristics for those having IBS according to comorbid psychological symptoms. **METHODS:** We asked IBS symptoms (Rome III), state-trait anxiety inventory, depressive symptoms (CES-D), and also analyzed stool microbial samples (16S rRNA gene analysis) for 672 individual from the Korean Healthy Twin Study. Using the Illumina's MiSeq platform, we obtained pyrosequencing data and filtered it through Chimera removal (QIIME) and sequence quality check. We measured operational taxa units (OTUs) using QIIME based on Uclust open reference, and aligned sequence using 13_8 revision of the Greengenes database. We estimated the α -diversity using four indices (Observed, Chao1, Shannon, and InvSimpson) and β -diversity using UniFrac distances and Bray-Curtis dissimilarity. The multivariate analysis, MaAsLin (Multivariate Association with Linear Models), was used to association test between microbial taxa and "gut-brain" axis phenotype. **RESULTS AND INTERPRETATIONS:** We compared α -diversity between IBS with psychological symptoms ($n=56$), IBS only ($n=25$), psychological symptoms only ($n=420$), and control group ($n=171$). The α -diversity of all IBS patients was slightly lower than those IBS-free subjects without psychological symptoms. IBS with anxiety and the IBS with depressive symptoms showed significantly lower α -diversity than IBS without psychological symptoms. The distance matrices, created by β -diversity (Unweighted/weighted UniFrac distances, Bray-Curtis dissimilarity) to compare MZ twin pairs, DZ twins and non-twin siblings, family and unrelated group, indicated that there is no significant difference between the MZ and DZSib, but between the other groups. When we compared OTU levels, the Firmicutes phylum, including the *Catenibacterium* genus, *Clostridium* genus, and *Ruminococcaceae* family was significantly less populated for IBS group. Our study supports that microbial profile is associated with gut-brain axis which might have consequences on the development of IBS.

2301W

Mitochondrial genetic variation is not associated with increased risk of developing uterine fibroids in African American or European American women. B.S. Mautz^{1,2,3}, J.N. Hellwege^{1,2,3}, T.L. Edwards^{1,2,3,4}, D.R. Velez Edwards^{2,3,4,5}. 1) Department of Medicine, Vanderbilt University Medical Center, Nashville, TN; 2) Institute for Medicine and Public Health, Vanderbilt University Medical Center, Nashville, TN, United States; 3) Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, United States; 4) Vanderbilt Epidemiology Center, Vanderbilt University Medical Center, Nashville, TN, United States; 5) Department of Obstetrics and Gynecology, Vanderbilt University, Nashville, TN, United States.

Uterine fibroids affect up to 70% of women by menopause and lead to large healthcare and lost work costs. African American (AA) women have up to a threefold increased risk of developing fibroids, as well as larger and multiple fibroids relative to European American (EA) women. Several autosomal SNP associations have been reported. However, no study has investigated the impact of mitochondrial genetic variation on fibroids despite differences in haplogroups in those with European and African ancestry which might contribute to the racial disparities. We used data from Vanderbilt University's BioVU, a biorepository of de-identified electronic medical records (EMR) linked to DNA, to investigate potential mitochondrial SNP and haplogroups associated with the risk of fibroids. Mitochondrial SNPs were extracted from Affymetrix Axiom Biobank arrays in EAs (N = 2273) and AAs (N = 1383). Fibroid status was assessed in the EMR through validated algorithms, including pelvic imaging and an intact uterus to define controls. Mitochondrial haplogroup assignment was performed using Haplogrep2 software. Fibroid characteristics (number, volume, and largest dimension) were extracted from imaging reports for a subset of AA cases. Single SNP association identified 4 suggestive variants in AAs (rs2854122, OR = 0.69 [95% CI: 0.508, 0.938], p = 0.018; rs2854128 OR = 0.439 [0.206, 0.936], p = 0.033; rs41337244 OR = 0.51 [0.271, 0.959], p = 0.037; rs2853496, OR = 0.868 [0.766, 0.983], p = 0.026). One SNP (rs2854122) is in a non-coding region, while the other 3 define Haplogroup H, or are in *CY-TOCHROME B* and *ND4* genes, respectively. The three top SNPs in the EAs (Affx-52321495, OR = 1.36 [1.01, 1.84], p = 0.04; Affx-52321475, OR = 1.61 [1.01, 2.59], p = 0.047; Affx-52321472, OR = 1.95 [1.01, 3.76], p = 0.048) are all in the *ND5* gene and unnamed. However, all SNPs are not statistically significant after accounting for multiple tests. The majority (77.4%) of EAs were assigned to haplogroup H or U, and 5 and 8 different haplogroups were represented among EA and AA participants, respectively. We found no relationship of haplogroup with risk of developing fibroids in either cohort, and mitochondrial haplogroup was unrelated to fibroid number or size in AA women. Our study did not show evidence for mitochondrial variants being associated with fibroid risk or specific fibroid characteristics.

2302T

Overlap of genetic risk between computed tomography interstitial lung abnormalities and idiopathic pulmonary fibrosis. B.D. Hobbs^{1,2}, R.K. Putman², T. Araki^{3,4}, G. Gudmundsson^{5,6}, V. Gudnason^{6,7}, G. Eiriksdottir⁷, N.R. Zil-hao⁷, J. Dupuis^{8,9}, H. Xu⁸, G.T. O'Connor^{9,10}, A. Manichaikul^{11,12}, J.N. Nguyen¹³, A. Podolanczuk¹³, D. Lederer^{13,14}, E.J. Ampleford¹⁵, V. Ortega¹⁵, S.P. Peters¹⁵, J.D. Newell Jr.^{16,17}, T.E. Fingerlin^{18,19}, D.A. Schwartz^{18,20,21}, G.R. Washko^{2,4}, I.O. Rosas², E.K. Silverman^{1,2}, H. Hatabu^{3,4}, M.H. Cho^{1,2}, G.M. Hunninghake^{2,3}, COPDGene Investigators, ECLIPSE Investigators, SPIROMICS Research Group. 1) Chan-ning Division of Network Medicine, Brigham and Women's Hospital, Boston, MA, USA; 2) Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, MA, USA; 3) Department of Radiology, Brigham and Women's Hospital, Boston, MA, USA; 4) Center for Pulmonary Functional Imaging, Brigham and Women's Hospital, Boston, MA, USA; 5) Dept of Respiratory Medicine, Landspítali University Hospital, University of Iceland, Reykjavik, Iceland; 6) Faculty of Medicine, University of Iceland, Reykjavik, Iceland; 7) Icelandic Heart Association, Kopavogur, Iceland; 8) Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA; 9) National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, MA, USA; 10) Pulmonary Center, Department of Medicine, Boston University, Boston, MA, USA; 11) Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA; 12) Department of Public Health Sciences, University of Virginia, Charlottesville, VA, USA; 13) Department of Medicine, College of Physicians and Surgeons, Columbia University, New York, NY, USA; 14) Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY, USA; 15) Center for Genomics and Personalized Medicine, Wake Forest School of Medicine, Winston-Salem, NC, USA; 16) Division of Cardiovascular and Pulmonary Imaging, Department of Radiology, University of Iowa Carver College of Medicine, Iowa City, IA, USA; 17) Department of Radiology, University of Washington, Seattle, WA, USA; 18) Center for Genes, Environment and Health, National Jewish Health, Denver, CO, USA; 19) Department of Biostatistics and Informatics, University of Colorado Denver, Aurora, CO, USA; 20) Department of Medicine, School of Medicine, University of Colorado Denver, Aurora, CO, USA; 21) Department of Immunology, School of Medicine, University of Colorado Denver, Aurora, CO, USA.

Rationale: Individuals with interstitial lung abnormalities (ILA) share a clinical syndrome similar to that observed in idiopathic pulmonary fibrosis (IPF) including physiologic decrements, radiologic progression, accelerated lung function decline, and an increased risk of death. ILA are associated with the most common and highest genetic risk locus in IPF, the *MUC5B* promoter polymorphism rs35705950. However, additional variants associated with ILA and the extent of overlap of the genetic risk between ILA and IPF are not recognized. **Methods:** ILA and the subpleural-predominant ILA subtype were visually assessed on chest computed tomography (CT) scans in European- and African-ancestry participants from the AGES, COPDGene, Framingham Heart, ECLIPSE, MESA, and SPIROMICS studies. Genome-wide variants were imputed utilizing the Haplotype Resource Consortium panel. We performed a genome-wide association study (GWAS) of ILA in each cohort using logistic regression adjusting for age, sex, pack-years smoking, and principal components of genetic ancestry. Cohort-level results were combined using a fixed-effects meta-analysis. Sixteen distinct ($R^2 < 0.8$) IPF GWAS SNPs from the NHGRI-EBI GWAS catalog were assessed for association with ILA and subpleural ILA. **Results:** Genome-wide SNP genotyping data were available in 1,868 ILA cases and 12,871 controls. The *MUC5B* promoter variant rs35705950 was significantly associated with both ILA ($p = 2.1 \times 10^{-11}$) and subpleural ILA ($p = 1.1 \times 10^{-10}$). We discovered a novel genome-wide significant association with ILA near *IPO11* (rs6886640, $p = 5.8 \times 10^{-9}$) and with subpleural ILA near *HTR1E* (rs7744971, $p = 2.7 \times 10^{-6}$). In an assessment of ILA association with prior 16 IPF GWAS SNPs, we identified 13 of 16 directionally consistent associations. Of these 13 SNPs, 5 SNPs (near *MUC5B*, *DSP*, *DPP9*, *TOLLIP*, and *FAM13A*) were significantly associated ($p < 0.05/16$) with ILA and 4 SNPs were nominally significant ($p < 0.05$). **Conclusions:** In a GWAS of ILA and subpleural ILA in six studies, we confirmed the previously described association with a *MUC5B* promoter variant and discovered additional significant associations near *IPO11* and *HTR1E*. Additional investigation is needed to replicate novel findings; however, our study demonstrates that IPF GWAS SNPs are enriched for ILA associations, suggesting a common genetic pathway between ILA and IPF, and advances the possibility of ILA as a precursor lesion for IPF.

2303F

Evaluation of polygenic risk scores for chronic obstructive pulmonary disease in a biobank linked to electronic health records. V.L. Martucci^{1,2}, B. Richmond³, T. Blackwell⁴, L.K. Davis^{1,4,5}, M.C. Aldrich^{1,2,6,7}. 1) Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN; 2) Department of Thoracic Surgery, Vanderbilt University Medical Center, Nashville, TN; 3) Division of Allergy, Pulmonary, and Critical Care Medicine, Vanderbilt University Medical Center, Nashville, TN; 4) Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN; 5) Department of Psychiatry and Behavioral Sciences, Vanderbilt University Medical Center, Nashville, TN; 6) Division of Epidemiology, Vanderbilt University Medical Center, Nashville, TN; 7) Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN.

Chronic obstructive pulmonary disease (COPD) is the third leading cause of mortality in the U.S. and worldwide. COPD is characterized by persistent respiratory symptoms and reduced lung function, causing significant morbidity. Pulmonary function testing (PFT) is used to determine disease status, with the ratio of forced expiratory volume in one second (FEV₁) to forced vital capacity (FVC) less than 0.7 considered diagnostic for COPD. Electronic health records (EHR) provide large-scale, routine healthcare data for COPD research, and linking EHR with genetic information provides unique opportunities for discovery of genetic risk factors. We investigated genetic predictors of COPD and lung function within a spirometry-defined clinical cohort at Vanderbilt University Medical Center. We identified individuals over 45 years of age who received PFTs between August 2014-September 2017 (N = 14,058). We selected participants from BioVU, the Vanderbilt de-identified biobank linked to EHR. Individuals were previously genotyped on the Illumina MEGA-Ex, Omni and Quad arrays. Over 6 million SNPs and 1,065 individuals remained after standard QC and imputation to 1000 Genomes. We defined COPD cases as individuals with a pre- or post-bronchodilator FEV₁/FVC < 0.7 (N = 343 cases, 722 controls). SNPs associated with COPD or lung function were identified from the NHGRI-EBI GWAS catalog to construct weighted polygenic risk scores (PRS) using PRSice software. Logistic or linear regression models examined associations between PRS and COPD or lung function, respectively, adjusting for sex and age at last clinic visit. The odds ratio for a 1 standard deviation change in a PRS with 91 SNPs was 2.60 [95% confidence interval (CI): 0.83 to 8.21]. A PRS with 64 SNPs was inversely associated with FEV₁/FVC (β = -0.46, 95% CI: -1.24 to 0.33), although not statistically significant. Inverse associations were observed with FEV₁ (β = -0.04, 95% CI: -0.09 to -0.01) and FVC (β = -0.05, 95% CI: -0.10 to 0.001), although only the association with FEV₁ was statistically significant ($p \leq 0.05$). These preliminary results demonstrate the potential for PRS to elucidate polygenic contributions to COPD. We will present extended analyses using PRS applied to 200,000 adults in BioVU using a natural language processing algorithm to identify COPD patients.

2304W

Genome-wide polygenic risk scores for chronic kidney disease in the UK biobank. *H.T. Vy¹, L. Chan², K. Chaudhary¹, S. Coca², R. Do^{1,2}, G.N. Nadkarni².* 1) Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine, New York, NY; 2) Division of Nephrology, Department of Medicine, Icahn School of Medicine, New York, NY; 3) Department of Genetics and Genomic Sciences, Icahn School of Medicine, New York, NY.

Early detection and treatment of chronic kidney disease (CKD) can potentially prevent disease onset and delay disease progression. In addition to traditional methods of detection which are based on measurements of laboratory values indicative of kidney function, genetic testing has emerged as a potential tool for the identification of high-risk individuals. Recent genome-wide association (GWA) studies have tested for associations between millions of single nucleotide variants (SNVs) and CKD, and have successfully identified a number of genetic loci for CKD. However, the predictive ability of the aggregate effect of these SNVs on CKD risk at a genome-wide level has not yet been explored. In this study, we generated per-individual genome-wide additive polygenic risk scores using publicly available CKD GWA association statistics and tested these scores for association with CKD in 8,432 cases and 400,529 controls of self-identified white British ancestry in the UK Biobank (UKBB) data set. We calculated different polygenic scores from four different sets of SNVs: 1) 33 loci significantly associated with estimated glomerular filtration rate (eGFR-crea); 2) 53 loci significantly associated with CKD risk; 3) >38,000 independent SNVs using a linkage disequilibrium (LD) pruning approach of $r^2 < 0.80$ and $P < 0.05$; and 1,797,807 SNVs with re-assigned coefficients adjusted for the effects of LD by LDpred. Among the four polygenic scores, the LDpred-adjusted polygenic score demonstrated the most robust association with CKD risk with an area-under-curve of 0.703. Using this score, individuals at the top 5% of the distribution had a more than 2-fold risk for CKD compared to those at the bottom 20%. Similar associations were also observed for other ancestries (African, East Asian) within UKBB. In conclusion, we demonstrate the utility of polygenic predictors for CKD risk through the identification of polygenic high-risk individuals for CKD across different ancestral populations.

2305T

RNA-seq from human pancreatic tissue with chronic pancreatitis provides insight into etiopathogenesis and classification. *C.A. Shelton^{1,2}, D.C. Whitcomb^{1,2}.* 1) Division of Gastroenterology, University of Pittsburgh, Pittsburgh, PA; 2) Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA.

Pancreatitis is a complex, genetically heterogeneous acute and chronic inflammatory disorder of the pancreas. There is wide variability in disease progression, pain and loss of function, even among patients with apparently identical susceptibility factors. Human tissue has been largely unavailable for study, and the majority of knowledge regarding molecular pathways of disease progression is from mouse models. However, few therapies proven to be effective in mouse models have had a positive impact on human disease, and there is no effective treatment for chronic pancreatitis (CP). Leftover tissue from *total pancreatectomy with islet autotransplantation* (TPIAT)—a surgical option for some patients with CP—represents an untapped resource to study biological drivers of CP in humans. This tissue is collected at the optimal time to evaluate underlying pathogenesis — *after* a pathogenic, progressive natural history is predicted, but *before* insulin-producing islets are lost. RNA was extracted from TPIAT RNALater samples (1 *PRSS1* CP; 2 idiopathic CP; 1 alcoholic recurrent acute pancreatitis (RAP)) and histologically normal tissue from pancreatic cancer cases, matched for age, sex and race. RNA integrity was consistent among samples, and paired-end total RNA-sequencing was used for optimal coverage. On principal component analysis, controls clustered and clear separation was seen between TPIAT tissue by stage (recurrent acute vs. chronic) and etiology. With a false discovery rate < 0.05 , nearly 500 genes were found to be differentially expressed by at least 2-fold. The most significantly upregulated genes were immunoglobulins, indicating a humoral immune response. Additionally, stellate cell activation was found to be the most significantly upregulated pathway in cases — an expected response to pancreatic injury. Differences between cases by stage and etiology also revealed new and expected pathways indicative of etiopathogenesis. For example, the most significant pathways upregulated in the alcoholic RAP tissue compared to idiopathic and genetic chronic pancreatitis were oxidative phosphorylation and mitochondrial dysfunction—established pathologic responses to alcohol. Although this is a small sample size, these results replicate findings from mouse models, lending biological plausibility. Our findings suggest that RNA-Seq may be useful for further classification of pancreatitis subtypes and reveal targets for early diagnosis and targeted treatment.

2306F

Genetic studies of body proportion (sitting height ratio) shed light on the recent evolution of human skeletal growth. *E.R. Bartell^{1,2,3}, K.M. Tsuo^{1,2,3}, K. Lin⁴, J.B. Cole^{1,2,3}, S. Vedantam^{1,2,3}, R.S. Fine^{1,2,3}, R.M. Salem^{1,2,3}, Y. Chan^{1,2,3}, R. Walters¹, J.N. Hirschhorn^{1,2,3}*. 1) Department of Genetics, Harvard Medical School, Boston, MA; 2) Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, MA; 3) Broad Institute of the MIT and Harvard, Cambridge, MA; 4) Nuffield Department of Population Health, Oxford, UK.

Genome-wide association studies (GWAS) have identified hundreds of common genetic variants associated with human height; however, less is known about the genetics of body proportion. A greater understanding of the genetics of Sitting Height Ratio (SHR, sitting height divided by standing height), a commonly-used measure of skeletal proportion, would expand our understanding of skeletal growth. Taken together with the genetics of human height, genetic variants can be classified into those affecting predominantly long bones (legs), spine/head, or overall height. Objective: To expand our understanding of skeletal growth using genetic studies of SHR. We use data from the UK Biobank to expand our genome-wide association study of SHR (Chan et al. 2015) from ~25,000 to >400,000 individuals. In this abstract, we use association results from BOLT-LMM, in ~500,000 individuals. Height-increasing alleles that decrease SHR were deemed to act primarily on the legs/long bones, and height-increasing alleles that increase SHR were deemed to act primarily on the head/spine. For analyses of selection, we used methods from our earlier work (Turchin et al., 2012), and expanded study populations to include those genotyped in the China Kadoorie Biobank (CKB), constituting ~100,000 further individuals. We identified >600 genome-wide significant SHR-, Height-, or both-associated variants ($p < 5 \times 10^{-8}$). We used the SHR association data to classify height variants according to their upper or lower body effect. We next re-examined a signal of selection we had previously demonstrated for height-associated variants (the difference in height between northern and southern Europeans is partly explained by differential selection on height-associated variants). In preliminary analyses, the signal of selection appears stronger in height-associated variants that act on long bones than in variants that act on spine/head length; if confirmed in the larger sample, this would indicate that recent evolutionary pressures differentially altered limb length within Europe. Results also show more differentiation in allele frequencies at loci that affect torso length in CKB populations as compared to leg length; this may argue for different selection pressures acting on the two different study populations. We have identified many novel loci associated with SHR that offer new insights into the genetic, biological and evolutionary basis of body proportion and skeletal growth.

2307W

Translating osteoarthritis GWAS signals into therapeutic targets. *L.C. McCarthy¹, I. Tachmazidou¹, K. Hatzikotoulas², L. Southam^{2,3}, J. Esparza-Gordillo⁴, J. Hoffman¹, J. Larkin⁴, P. Fisher¹, T. Johnson¹, M. Koprulu^{2,5}, E. Zengin^{6,7}, J. Steinberg^{2,8}, J.M. Wilkinson⁹, S. Bhatnagar⁹, M.N. Nelson⁴, R.A. Scott¹, E. Zeggini², arcOGEN Consortium*. 1) GlaxoSmithKline, Stevenage, United Kingdom; 2) Wellcome Trust Sanger Institute, Wellcome Genome Campus, UK; 3) Wellcome Centre for Human Genetics, University of Oxford, UK; 4) GlaxoSmithKline, King of Prussia, US; 5) University of Cambridge, UK; 6) University of Sheffield, UK; 7) Dromokaiteio Psychiatric Hospital, Greece; 8) Cancer Council NSW, Australia; 9) McGill University, Canada.

Osteoarthritis (OA) is the leading cause of disability globally. Pain management and joint replacement are the therapeutic options currently available to OA patients, and there remains an enormous unmet medical need for effective disease modifying drugs. We conducted the largest genome-wide association study for OA to date (70,588 cases and 263,702 controls) using UK Biobank data, and meta-analysed the data with arcOGEN consortium OA data, to reveal 43 genome-wide signals, 32 of which are novel. Our focus is the translation of these genetic signals into the identification of targets for new drugs which impede OA joint degeneration and may promote OA joint healing. The following evidence elements contributed to refining the list of genes likely to drive OA association signals: fine mapping, variant effect prediction (VEP), colocalization with eQTL signals, gene mutations driving rare human skeletal/joint phenotypes including early onset OA, differential protein and/or RNA expression and/or methylation in OA vs. normal joint tissue, gene mutations driving relevant skeletal/joint phenotypes in animal models. Following integration of these evidence strands, 15 loci had only 1 gene with supportive evidence for driving the OA signal. There were 21 loci containing ≥ 2 genes with OA related evidence. 7 loci had no genes with plausible evidence for driving the OA signal. Five (*TGFB1*, *IL11*, *MST1R*, *DPEP1*, *PTH1LH*) of the potential effector genes, based on human genomics and animal model evidence, have therapeutics approved or in clinical trials for a range of indications, and mechanisms of action consistent with potential efficacy for OA. Of the 137 genes in OA loci with any evidence supporting their role in OA susceptibility, 100 are predicted to be potentially tractable to small molecule modulation, and 70 are predicted to be tractable to biological therapeutic intervention. Where molecular tools exist, or can be created, to modulate the products of these genes, we will test their activity in models of OA. We describe approaches to translating OA GWAS signals to potential new drug targets.

2308T

Integration analysis of multi-omics data for osteoporosis biomarker discovery. C. Qiu, F.T. Yu, K.J. Su, H.W. Deng, H. Shen. Tulane University School of Public Health and Tropical Medicine, New Orleans, LA.

Osteoporosis is an increasingly serious public health problem, which is characterized by low bone mineral density (BMD). To date, the functional roles of the most reported biomarker that are associated with BMD remain unclear. To increase the understanding of the biological mechanisms of BMD variation, it is necessary to leverage multiple high dimensional datasets to capture the complexity of biological networks. Recent advent of high throughput technologies has created an opportunity for integrating multi-omics data to build a comprehensive and dynamic model of the molecular changes in osteoporosis. In this study, we generated the multi-omics datasets with three data types (mRNA, methylation, and metabolomics) from 120 subjects with extremely high (n=61) and low (n=59) hip BMDs. By focusing on potential interesting biomarkers (differentially expressed genes, differential methylated CpG sites, and differential metabolites) generated from single layer analysis, we applied a data integration analysis for biomarker discovery using latent components (DIABLO) to identify the key omics biomarkers for BMD variation. We identified a multi-omics signature of 22 and 15 mRNA biomarkers, 36 and 11 metabolomics biomarkers, 11 and 5 methylation biomarkers on the first two components, respectively. The DIABLO model displayed a better discrimination of BMD variation with the methylation and metabolomics than with mRNA data. Correlation analysis showed a strong correlation between mRNA and metabolomics, and 17 mRNA biomarkers, 23 metabolomics biomarkers, and 6 methylation biomarkers in the first component were highly negatively correlated with the first component. A relevance network analysis highlighted two clusters, both including features from the three types of omics data. In addition, we performed Gene Ontology analysis on the largest cluster consisting of 32 omics biomarkers. Interestingly, several biomarkers (e.g., *ICA1L*, *PCAT1*) have been associated with BMD variation through previous genome-wide association studies. In conclusion, we reconstructed the biological network in multi-omics level and identified several novel biomarkers that may mediate variation in risk of osteoporosis. Our results highlighted the advantages of using integration analysis to add value to the traditional single layer analysis.

2309F

Cellular crosstalk between retinoic acid and pesticides: A model for understanding osteoporosis through genomic and non-genomic pathways. H.S. Sandhu^{1,2}, G. Singh^{1,4}, S. Puri^{2,3}, A.J.S. Bhanwer¹. 1) Department of Human Genetics, Guru Nanak Dev University, Amritsar, India; 2) Centre for Stem Cell and Tissue Engineering, Panjab University, Chandigarh, India; 3) Biotechnology Branch, UIET, Panjab University, Chandigarh, India; 4) Department of Anthropology, Panjab University, Chandigarh, India.

The altered bone quality and increased risk of fractures featuring the debilitating bone loss disease osteoporosis involves a crosstalk between various environmental, nutritional and genetic factors. Environmental intoxicants such as pesticides adversely affect the skeletal homeostasis by binding to receptors on bone marrow mesenchymal stem cells (MSCs) and exacerbating the adipogenic differentiation of these inherent stem cells. To understand this mechanism, a dual approach was employed comprising of (1) a murine cell-based test system for continuous, label-free monitoring of the effect of retinoic acid, a model compound for osteoporosis, and an organophosphate pesticide chlorpyrifos (CPF) on adipogenic differentiation, and (2) a human case-control sample set. The *in vitro* model of C3H10T1/2 MSCs generated by combined treatment with 2 μ M RA+ 50 μ M CPF showed vastly increased adipogenic differentiation as compared to vehicle control and independent treatments. These results were characterized by microscopic and biochemical studies. Expression analysis of major adipogenic genes by RT-PCR suggested an upregulation of *adipsin* (*ADN*) and *perilipin-1* (*PLIN1*). Screening for various pathway inhibitors revealed GSK3 β to be the key player, with its inhibitor LiCl resulting in complete blockade of lipid accumulation and downregulation of these adipogenic markers. For validating the role of these genes, peripheral blood was collected from 102 male/female subjects from North India having direct exposure to pesticides and bone quantitative ultrasound t-score ≤ -1 , and compared with 167 age/gender matched healthy controls with t-score > -1 and without any direct pesticide exposure. CPF residues were quantified in a representative sample set using HPLC and showed a higher mean concentration in cases (198.231 ng/ml) as compared to controls (114.45 ng/ml). Analysis of the genetic variants of *ADN* and *PLIN1* by PCR-based methods revealed that rs1683563 and rs2304796 respectively, conferred 1.7-2.3 fold risk towards the susceptibility of bone loss upon pesticide exposure. However, only redundant gene-gene interactions were prevalent among these variants. The present study highlights that *ADN* and *PLIN1* might be important susceptibility loci for an increased risk of osteoporosis upon pesticide exposure, however, ethnicity differences need to be considered. Hence, impairment in metabolic pathways resulting in obesity and related phenotypes also exerts its influence on bone health.

2310W

A variant in *MEGF6* gene predisposes to osteoporosis. C.C. Teerlink¹, J. Stevens¹, D.C. Hughes^{1,2}, C.P. Brunke^{1,3}, M.J. Juryne¹, D. Grunwald⁴, L.A. Cannon-Albright^{1,6,7}. 1) Genetic Epidemiology, Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City, UT; 2) Rehabilitation Services, University of Utah, Salt Lake City, UT; 3) Geriatric Medicine, Intermountain Healthcare, Salt Lake City, UT; 4) Orthopedic Surgery, University of Utah, Salt Lake City, UT; 5) Human Genetics, University of Utah, Salt Lake City, UT; 6) George E. Wahlen Department of Veterans Affairs Medical Center, Salt Lake City, UT; 7) Huntsman Cancer Institute, Salt Lake City, UT.

Osteoporosis is a common skeletal disorder characterized by deterioration of bone tissue in later life. The set of genetic factors contributing to osteoporosis is not completely specified. We performed a relatively small exome-sequencing study on previously sampled osteoporosis pedigree members designed to identify rare, highly penetrant variants that segregate with disease following a dominant mode of inheritance. The osteoporosis pedigree resource consists of 1,871 individuals in 276 multi-generational pedigrees with 4–46 sampled individuals per pedigree. All study participants received bone mineral density (BMD) scans via Hologic QDR 4500A fan-beam dual X-ray absorptiometry bone densitometer of the hip and spine. BMD scores were converted to a normalized age- and sex-adjusted score (T-score). We performed exome sequencing on affected-relative-pairs, approximately cousins, from 10 high-risk osteoporosis pedigrees with an extreme phenotype defined as the lowest 5% of the distribution of T-scores across all subjects to identify rare coding variants that may contribute to osteoporosis in these families. Variants were filtered on population frequency (minor allele < 0.005) and concordance between sequenced relative pairs, resulting in 319 candidate variants. We further reduced the set of candidate variants by requiring that the gene of origin had at least one previously reported association to osteoporosis identified via PubMed search, resulting in 12 candidates. The 12 remaining candidates were tested for segregation via Taqman assay in the additional, previously sampled, affected relatives of the index carriers, which revealed a rare variant in the *MEGF6* gene (rs755467862 at chr1:3,440,693 (hg19)) with strong evidence for segregation (6 affected relatives carrying the variant, RVsharing probability = 1.2E⁻⁴). The *MEGF6* gene resides in the 1p36.3 cytogenetic band which has two previous reports of evidence for linkage to osteoporosis. We are currently generating loss-of-function alleles in the 2 zebrafish *MEGF6* orthologues, *megf6a* and *megf6b*, to determine the biological role for this gene and variant in osteoporosis predisposition.

2311T

Genetic variants in ER stress signaling molecules *IRE1* and *XBP1* are associated with human bone and tooth phenotypes. Y. Zhou¹, H. Ouyang^{2,3}, D. McNeil⁴, R. Weyant⁵, R. Crout⁶, R. Slayton⁶, T. Burns⁷, S. Levy⁸, M. Marazita¹, J. Shaffer⁹. 1) Department of Oral Biology, University of Pittsburgh, Pittsburgh, PA, 15219; 2) Texas A&M University, Dallas, Texas 75246; 3) University of Texas Southwestern Medical Center; 4) West Virginia University, Morgantown, West Virginia; 5) School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA; 6) School of Dentistry, University of Washington, Seattle, WA, USA; 7) College of Public Health, University of Iowa, Iowa City, IA; 8) University of Iowa College of Dentistry, Iowa City, IA; 9) Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA, 15261.

Introduction Our work in mice suggests that the endoplasmic reticulum stress *IRE1/XBP1*s signaling axis is important for mineralized tissues, with roles in bone formation and tooth eruption. Specifically, mice with the *IRE1*-coding gene *ERN1* knocked out in *Osterix*-positive cells showed delayed tooth eruption, skeletal and dental morphological differences, and low bone mineral density (BMD). Whether genetic variants in these genes influence bone and dental phenotypes in humans remains unknown. Therefore, we employed an *in silico* candidate gene approach testing association of variants in *ERN1* and *XBP1* with BMD, tooth eruption, and dental caries. **Methods** Participants were white, non-Hispanic children from the Center for Oral Health Research in Appalachia (N=506; ages 0-6), Iowa Fluoride Study (N=410; ages 3-6) and Iowa Head Start Study (N=266; ages 0-6). These cohorts were genotyped together using an Illumina SNP chip. Common SNPs in the candidate genes plus 500kB up- and downstream were tested for genetic association with timing of primary tooth eruption, dental caries, and DXA-derived spine and hip BMD and BMC while adjusting for ancestry and pertinent covariates. A total of 1363 and 1395 SNPs were tested for *ERN1* and *XBP1*, respectively. The locus-wise threshold for significance given the effective number of independent tests was P<2E-4. **Results** We observed significant associations for both *ERN1* and *XBP1* with bone phenotypes. For the *XBP1* region, we detected the association of rs5752758 with spine BMC (P=2E-6). This SNP also showed suggestive evidence of association for all other bone phenotypes (hip BMD P=7E-4, hip BMC P=9E-4, spine BMD P=7E-4). For the *ERN1* region, SNP rs71376922 (P=2.5E-5) was associated with hip BMD. This SNP is in strong LD with a variant in known promoter and enhancer regions in osteoblasts. A different SNP in the 5' regulatory region of *ERN1*, rs75160892, was associated with dental caries (P=1E-4); this SNP is in LD with several SNPs showing enhancer chromatin marks in multiple cell types and one SNP that is a known eQTL for *ERN1*. **Conclusion** Our findings suggest that common variants at the *ERN1* and *XBP1* loci are associated with both bone and dental phenotypes in humans. These data, along with the mouse model showing deregulated bone and tooth development, highlight a role of the *IRE1/XBP1*s axis in regulating these biological processes. **Funding** U01-DE018903, R03-DE024264, R01-DE014899, R01-DE009551, R01-DE12101, R01-CA182418.

2312F

Comprehensive functional annotation of susceptibility SNPs associated with osteoporosis. Y. Guo, Y.J. Zhang, H.M. Niu, S.S. Dong, T.L. Yang. Key Laboratory of Biomedical Information Engineering of Ministry of Education, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, China.

We aimed to provide a comprehensive functional annotation of osteoporosis-associated SNPs identified by previous genome-wide association studies (GWASs) and genetic association studies and offer guides for post-GWAS functional experiments. In this study, a total of 10,891 susceptibility SNPs (487 sentinel SNPs and 10,839 proxy SNPs in LD) were included. The target genes of all functional SNPs are enriched in osteoporosis-related pathways, including basal cell carcinoma, Wnt signaling pathway, and Hedgehog signaling pathway. Over 90% of the susceptibility SNPs are located in non-coding regions and it is difficult to understand their functional roles. Therefore, we annotated these SNPs via various functional databases. We identified 937 functional SNPs, including 1 nonsense SNPs, 17 missense SNPs with protein damaging effects, 47 UTR SNPs which might affect microRNA binding, 20 promoter and 851 enhancer SNPs which might affect transcription factor (TF) binding. The enhancer SNPs and promoter SNPs are enriched in 19 TFs, and 9 of these 19 TFs have been confirmed in mouse knockout models for association with osteoporosis. To integrate the GWAS and eQTL data, we used SMR approach and discovered a total of 111 instances of gene-trait-associated changes in expression at a false-discovery rate < 0.05. These genes were primarily identified in eQTL derived from pathophysiological relevant tissues, like whole blood and muscle skeletal. 19 of 111 (17.12%) trait-associated genes have been shown to be associated with osteoporosis-associated phenotypes in mouse knockout model. 21/111 trait-associated gene were differentially expressed during osteoblast differentiation. There are 5 trait-associated genes (*ARHGAP1*, *CBFB*, *CTNNB1*, *FGFRL1*, and *SOC2*) were annotated in both mouse knockout models and osteoblast differentiation. Our work is not only useful for understanding the etiology of osteoporosis and the biology of non-coding sequences in the genome, but also guides for post-GWAS functional experiments.

2313W

miRNAs targeting the mechanotransduction pathway associated with osteon number in aging baboons. E.E. Quillen^{1,2}, L.A. Cox^{1,2}, D.P. Nicoletta³, T.L. Bredbenner⁴. 1) Department of Internal Medicine - Molecular Medicine, Wake Forest School of Medicine, Winston-Salem, NC; 2) Center for Precision Medicine, Wake Forest School of Medicine, Winston-Salem, NC; 3) Materials Engineering Department, Southwest Research Institute, San Antonio, TX; 4) Department of Mechanical and Aerospace Engineering, University of Colorado Colorado Springs, Colorado Springs, CO.

The key cells regulating bone and cartilage remodeling – osteocytes, osteoblasts, osteoclasts and chondrocytes – are acutely responsive to external mechanical signals through a process known as mechanotransduction. The goal of this study is to determine if miRNAs targeting the mechanotransduction pathway are associated with increased bone fragility using a baboon model. No work was done on living animals and no animals were euthanized specifically for this study. RNA was extracted from white blood cells collected at necropsy and miRNA libraries were sequenced on the Illumina HiSeq2500. Predicted miRNAs were identified based on miRbase. miRNAs from 20 miRNA families predicted to target genes identified as part of the generalized mechanotransduction pathway were identified using TargetScan version 7.2. Skeletal fragility is a major risk factor for fracture but is incompletely indexed by the most common clinical indicator, bone mineral density. Histological analysis of the cortical midshaft and femoral neck of each baboon was used to characterize a panel of traits including osteon number and size as well as the area of Haversian canals. Linear models of the influence of miRNAs on these measures of bone structure were fit while controlling for age, sex, and body size of the baboons. Associations with 9 miRNAs passed Bonferroni correction, with the three largest effect sizes in isoforms of miR-9-5 on osteon number. This miRNA has previously been associated with reductions in skeletal cell proliferation, differentiation, and adhesion *in vitro*. Regulation of this miRNA is a possible mechanism of action for the osteoporosis drug strontium ranelate. This finding suggests This work was funded by NIH/NIAMS R01 AR064244 and R01 AR060341.

2314T

A genome-wide SNP x smoking interaction study identifies novel genetic risk loci for aggressive periodontitis in smokers and gives evidence that *ST8SIA1* is regulated by tobacco smoke. A.S. Schaefer¹, S. Freitag-Wolf¹, M. Munz¹, R. Wiehe¹, A. Franke², W. Lieb³, H. Dommisch¹, *European Periodontitis Consortium*. 1) Charité - University Medicine Berlin, Research Center ImmunoSciences, Dept. of Periodontology and Synoptic Dentistry, Berlin, Germany; 2) Institute of Medical Informatics and Statistics, University Clinic Schleswig-Holstein, Kiel, Germany; 3) Institute of Clinical Molecular Biology, Christian-Albrechts-University Kiel, Germany; 4) Institute of Epidemiology, Christian-Albrechts-University Kiel, Germany.

Periodontitis (PD) is an inflammatory disease of the oral cavity that leads to the resorption of the alveolar bone. The early-onset and severe phenotype aggressive periodontitis (AgP) has a heritability of 50%. Smoking is a well-established environmental risk factor for PD but there is considerable variation in age of disease onset and bone loss among smokers. Differences in smoking intensity are not solely responsible for this variation. It is also influenced by genotype differences between patients, which respond differently to exposure to the same environmental factor, indicating that to some extent, the sensitivity to smoking rather than AgP itself is inherited. Here, we investigated whether the relative risk of AgP in smokers is modified by genetic variants. For this purpose, we used imputed genotype data from a recently performed genome-wide association study. We regarded the early-onset phenotype AgP as particularly suited for this investigation, because the proposed gene x environment interaction (GxE) is revealed within years of exposure rather than within decades. We adopted a 2-tiered approach, including the verification in control individuals of the gene-smoking independence assumption implicit to the case-only design, followed by a case-only analysis to identify gene-smoking interactions that are relevant for the disease etiology of PD (stage I). Subsequently, we performed a case-control analysis, to identify variants that show additional gene-smoking interaction beyond the specific disease phenotype of PD (stage II). To proof the transcriptional response of identified SNP associations at protein coding genes to cigarette smoke, we stimulated primary gingival fibroblasts with cigarette smoke extract *in vitro*. Genome-wide analysis of GxE was performed with 896 AgP cases and 7,104 controls. In the case-only design, 13 loci corresponded to the pre-assigned significance cut-off for suggestive association with $P < 5 \times 10^{-5}$, supported by ≥ 2 flanking SNPs with $P < 10^{-5}$. In the case-control design, performed to assess GxE beyond the specific disease phenotype of AgP, we found 5 additional loci. In the smoke stimulated cells, *ST8SIA1* showed significant upregulation ($p < 10^{-7}$, five biological replicates). We conclude that *ST8SIA1* is a risk gene for AgP and the effects of the associated variants are exposed in conjunction with smoking in genetically predisposed individuals.

2315F

CRISPR/Cas9 mediated edits in osteoblast cell lines to characterize a novel osteoporosis gene, *DAAM2*. L. Laurent¹, J.A. Morris^{1,2}, A. Luco³, L. Nguyen-Yamamoto³, A. Pramatarova⁴, D. Goltzman⁵, J.B. Richards^{1,2,6,7}. 1) Lady Davis Institute, Jewish General Hospital, McGill University, Montreal, Quebec, Canada; 2) Department of Human Genetics, McGill University, Montreal, Quebec, Canada; 3) Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada; 4) McGill University and Genome Quebec Innovation Centre, Montreal, Quebec, Canada; 5) Division of Endocrinology, Department of Medicine, McGill University, Montreal, Quebec, Canada; 6) Departments of Medicine and Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Quebec, Canada; 7) Department of Twin Research and Genetic Epidemiology, King's College London, London, UK.

Introduction: Osteoporosis is a common and complex disease for which novel therapeutic options are a health priority. Through large-scale GWAS followed by integrative analyses with statistical fine-mapping and functional genomics, we have identified a list of target genes with strong evidence for being novel osteoporosis genes. We are developing a genome editing validation pipeline to validate target genes and have selected one top candidate gene as a proof-of-principle study, *DAAM2*. Using CRISPR/Cas9 and the human osteoblast cell line, SaOS-2, we studied the effect of knocking down *DAAM2* in this osteoblast-like cell with the ability to mineralize. **Methods:** Three guide RNAs (gRNA) targeting exon 2 of *DAAM2* were designed and cloned in the PX458 plasmid expressing green fluorescent protein (pSpCas9(BB)-2A-GFP). SaOS-2 cells were transfected with the constructed plasmids or an intact control plasmid and GFP-sorted in a single-cell model. Clones edited successfully with double-stranded breaks in *DAAM2* were identified by immunofluorescence and Western blot as well as MiSeq sequencing. Bone phenotypes such as mineralization, alkaline phosphatase (ALP) activity and collagen type I (*COL1A1* and *COL1A2*) expression were assessed. **Results:** Successful edits were induced using two of three gRNAs, and decreases of 82.5% and 66.5% of *DAAM2* protein level was detected for the two edited clones compared to control cells. We observed a 9-fold increase of mineralization in control cells treated with osteogenic factors, whereas induced mineralization was not observed in clones with successful *DAAM2* edits. No difference of ALP activity was reported in the three strains. Finally, an increase of *COL1A2* mRNA levels was observed for only one of the edited clones compared to control cells. **Conclusions:** In summary, by successfully applying CRISPR/Cas9 editing to decrease *DAAM2* protein expression in a human osteoblast cell line, we have confirmed that *DAAM2* is crucial for bone biology. Further studies will improve our understanding of the *DAAM2* molecular pathway and are currently underway. Finally, this study is a proof of concept, which will validate our target gene approach to discover new therapeutic targets for osteoporosis.

2316W

Large scale meta-analysis of genome-wide association studies for height in multiple ancestries. S. Vedantam^{1,2}, A. Locke³, E. Marouli⁴, S. Berndt⁵, L. Yengo⁶, A.R. Wood⁷, T. Ferreira⁸, S. Graham⁹ on behalf of the Genetic Investigation of Anthropometric Traits (GIANT) Consortium. 1) Endocrinology, Boston Children's Hospital, Boston MA 02115, MA; 2) Broad Institute of MIT and Harvard, Cambridge, MA, USA; 3) McDonnell Genome Institute, Division of Genomics and Bioinformatics, and Department of Medicine, Washington University School of Medicine, St. Louis, MO, USA; 4) Heart Centre, William Harvey Research Institute, Barts & the London Medical School, Queen Mary University of London, Charterhouse Square, London, EC1M 6BQ; 5) National Cancer Institute, Bethesda, MD; 6) Program in Complex Trait Genomics, Institute for Molecular Bioscience, University of Queensland, St Lucia, Brisbane, Australia; 7) Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Exeter, EX2 5DW, UK; 8) Big Data Institute, Li Ka Shing Center for Health Information and Discovery, Oxford University, Oxford, UK; 9) Department of Internal Medicine, Division of Cardiovascular Medicine, University of Michigan, Ann Arbor, MI.

Human adult height is a highly heritable, classic polygenic trait. Previous genome wide association study (GWAS) meta-analyses and exome chip studies by the GIANT consortium and others have identified >900 independent common variant association signals in >400 near-independent genetic loci that explain >25% of the heritability in height. To identify additional associations with height, we performed a GWAS on 456,426 European-ancestry participants from UKBiobank (UKB) using BOLT-LMM and combined these results with available GWAS summary statistics from GIANT (N~250,000; Wood et al. 2015). We identified 3,290 genome-wide significant associations in 712 loci, explaining >30% of heritability. In parallel, we are performing a meta-analysis of >250 studies from 6 different ancestries (not including UKB, total N = 906,589). All have genotype data imputed from the 1000Genomes Phase3 reference panel; many European-ancestry samples were also imputed to the Haplotype Reference Consortium or population-specific reference panels. Linear mixed effects models (implemented in Rvtest) and the first four principal components were used to adjust for relatedness and population structure. Association testing used inverse normally transformed age-adjusted residuals, stratified by sex and (if relevant) disease status. A preliminary meta-analysis using raremetal in 140 of these studies with European ancestry samples (total N = 618,362), imputed to 1000Genomes Phase3, identifies 1186 independent loci (defined as 1Mb apart from each other), including novel signals. Of the 1000 lead variants where UKB results were available, 918 strongly replicated in UKB ($p < 5 \times 10^{-8}$, which is $p < 0.05$ corrected for 1000 tests). Encouragingly, despite the large sample size, little evidence of association was observed at a highly stratified variant known to be associated with lactase persistence (rs4988235, p -value = 0.002). We will be adding data from cohorts from multiple ancestries and from the UKBiobank, providing a multiethnic discovery meta-analysis sample size of >1.5 million; large additional replication cohorts will be available. Secondary analyses will include fine mapping, aggregate testing of low frequency variants, and pathway analyses to delineate likely causal genes and biological mechanisms. Through our collaborative efforts, we have assembled a unique, large resource for understanding the genetic architecture of height and, more generally, polygenic traits.

2317T

Expression quantitative trait locus data from human osteoclasts suggests a role for the *C7orf73* and *DCSTAMP* genes in Paget's disease of bone. S.G. Wilson^{1,2,3}, B.H. Mullin^{1,2}, K. Zhu^{1,2}, S.J. Brown², S. Mullin^{1,2}, J. Tickner¹, N.J. Pavlos¹, F. Dudbridge⁴, J. Xu¹, J.P. Walsh^{2,5}. 1) School of Biomedical Sciences, The University of Western Australia, Crawley, WA, Australia; 2) Department of Endocrinology & Diabetes, Sir Charles Gairdner Hospital, Nedlands, WA, Australia; 3) Department of Twin Research & Genetic Epidemiology, King's College London, London, UK; 4) Department of Health Sciences, University of Leicester, Leicester, UK; 5) Medical School, University of Western Australia, Crawley, WA, Australia.

Paget's disease of bone (PDB) is a focal disorder of bone remodeling with clinical features of bone pain, deformity, neurological complications and risk of malignant transformation. The disease process is initiated by increases in osteoclast mediated bone resorption with a subsequent compensatory increase in new bone formation resulting in a mosaic of woven and lamellar bone at affected sites that disrupts bone architecture and structure. While the primary feature of PDB at the cellular level is increased bone resorption, the osteoclasts in affected bone are larger in size than normal osteoclasts, have more nuclei and display characteristic intra-nuclear inclusion bodies. Genetic factors have been shown to play an important role in the pathogenesis of PDB, and genome-wide association studies (GWAS) have identified 7 genetic loci as associated with PDB at the genome-wide level. Expression quantitative trait locus (eQTL) studies using disease-relevant cell types have increasingly been integrated with the results from GWAS to identify genes through which the observed GWAS associations are potentially mediated. We recently generated a unique osteoclast-specific eQTL resource using cells differentiated *in vitro* from 158 participants for study of the genetics of bone disease. Considering the major role osteoclasts have in PDB, we used this resource to investigate potential genetic regulatory effects for the 7 genome-wide significant loci for PDB. All expressed genes located within 500 kb of each GWAS variant were included in the analysis, which was performed using the FastQTL software. After correction for multiple testing, we observed statistically significant associations for rs4294134 with expression of the gene *C7orf73*, and rs2458413 with expression of the genes *DPYS* and *DCSTAMP*. The eQTL associations observed for rs4294134 with *C7orf73*, and rs2458413 with *DCSTAMP* were further supported by eQTL data from other tissue types. The product of the *C7orf73* gene has not been extensively studied, however the *DCSTAMP* gene has an established role in osteoclast differentiation and the associations seen between rs2458413 and PDB are likely mediated through regulatory effects on this gene. This study further highlights the value of eQTL data in determining which genes are relevant to GWAS loci.

2318F

RNA sequencing of osteocalcin-positive blood cells from adolescent idiopathic scoliosis patients reveals abnormal expression of cytoskeletal genes. E.A. Terhune¹, C.I. Wetthey¹, R.M. Baschal¹, M.T. Cuevas¹, K.D. Robinson¹, K.A. Payne¹, K.L. Jones², N.H. Miller^{1,3}, E.E. Baschal¹. 1) Department of Orthopedics, University of Colorado Anschutz Medical Campus, Aurora, CO; 2) Department of Pediatrics, University of Colorado Anschutz Medical Campus, Aurora, CO; 3) Musculoskeletal Research Center, Children's Hospital Colorado, Aurora, CO.

Background: Adolescent Idiopathic Scoliosis (AIS) is the most common pediatric spinal deformity, which affects ~3% of healthy children. The hereditary nature of this disorder is well accepted, but the overall genetic etiology of AIS is not well understood. Recent studies have identified osteocalcin-expressing blood cells originating from the bone marrow compartment, which represent a unique opportunity to study bone biology without requiring the collection of patient bone samples. **Methods:** In this study, we isolated osteocalcin-positive peripheral blood mononuclear cells (PBMCs) from the blood of four female AIS patients and four controls via fluorescence-activated cell sorting (FACS). We isolated RNA from these cells and completed RNA sequencing to identify the transcriptome profiles of osteocalcin-positive PBMCs in AIS and control individuals. Cufflinks software was used to calculate FPKM (normalized transcript levels) and p-values for AIS vs. control gene expression. Principal component analysis revealed that the samples were clustered based on both scoliosis status and cell sorting quality. We decided to analyze only the samples with high quality cell sorting, leaving us with two AIS and two control samples. We then used the Cytoscape plugin BiNGO to identify pathways and GO terms that are over-represented in the genes that have significantly different expression levels in AIS vs. control individuals. **Results:** We found that 34 transcripts were differentially expressed between AIS and control cells (either >2 or <-2 FPKM fold change and $p < 0.05$, or $p < 0.01$). These transcripts included those from microtubule genes, which were mildly upregulated in AIS (*KIF2A*, 1.23 fold change, $p = 0.0043$; *CALM1*, 1.2 fold change, $p = 0.0085$) and actin genes, which were downregulated in AIS (*CAPG*, -2.03 fold change, $p = 0.031$; *FGD3*, -2.13 fold change, $p = 0.00047$; *VCL*, -2.19 fold change, $p = 0.023$; and *RAB10*, -2.15 fold change, $p = 0.0075$). BiNGO analysis revealed that these 34 transcripts were enriched in GO terms related to actin-based cell projections, other cytoskeletal proteins, and glycoproteins. **Conclusions:** This is the first study to analyze circulating osteocalcin-positive cells in AIS, and we suggest that abnormal expression of cytoskeletal genes in bone-forming precursor cells may contribute to AIS etiology.

2319W

Investigating the GWAS-implicated loci for rheumatoid arthritis in Pakistani population. M. Aslam^{1,2}, P. John², F. Kang-Hsien¹, A. Bhatti², E. Feingold¹, F. Demirci¹, M. I. Kamboh¹. 1) University of Pittsburgh, Pittsburgh, PA; 2) National University of Sciences and Technology, Islamabad, Pakistan.

Rheumatoid arthritis (RA) is a complex, multifactorial, autoimmune disorder that is characterized by chronic inflammation of the joints. A complex interplay between genetic and environmental factors initiates the abnormal response of the immune system and leads to disease manifestation. The heritability estimate for RA is approximately 60%. RA affects ~1% of the general population worldwide while its prevalence is ~0.5% in Pakistan. Genome-wide association studies (GWAS) have identified more than 50 RA genetic loci in European populations. Given the anticipated overlap of RA-relevant genes and pathways across different ethnic groups, we sought to replicate 58 GWAS-implicated RA-related SNPs reported in Europeans in our Pakistani case-control sample. We genotyped 58 SNPs in 1959 unrelated Pakistani subjects comprising 1,222 RA cases and 737 controls. All cases met the ACR 1987 classification criteria for RA. Genotyping was performed using iPLEX or TaqMan methods. A total of 42 SNPs were included in the final association analysis after excluding those that failed assay design/run or post-run QC analysis. Ten SNPs (*PVT1*/rs1516971, *PADI4*/rs2240336, *CEP57*/rs4409785, *STAT4*/rs13426947, *HLA-B*IMICA*/rs2596565, *C5orf13*/rs26232, *CCL21*/rs951005, *GATA3*/rs2275806, *CD5*/rs595158 and *HLA-DRB1*/rs660895) were replicated in our Pakistani sample at false discovery rate (FDR) <0.20 with *P* values ranging from 4.73E-06 to 4.02E-02. Our results indicate that several RA susceptibility loci are shared between Pakistani and European populations supporting the role of common genes/pathways.

2320T

Associations between gene expression and clinical disease activity during and after pregnancy among women with rheumatoid arthritis: A pilot study.

D. Jawaheer^{1,2,3}, *D. Goin*^{1,4}, *M. Smed*⁵, *N. Jewell*⁶, *L. Pachter*⁶, *J.L. Nelson*^{7,8}, *J. Olsen*⁹, *M. Hetland*^{9,10}, *V. Zoffmann*⁹, *H. Kjærgaard*⁹. 1) UCSF Benioff Children's Hospital Oakland Research Institute, Oakland, CA; 2) University of California San Francisco, San Francisco, CA; 3) Aarhus University, Aarhus, Denmark; 4) University of California Berkeley, Berkeley, CA; 5) Juliane Marie Center, Copenhagen, Denmark; 6) California Institute of Technology, Pasadena, CA; 7) Fred Hutchinson Cancer Research Center, Seattle, WA; 8) University of Washington, Seattle, WA; 9) DANBIO, Glostrup, Denmark; 10) University of Copenhagen, Copenhagen, Denmark.

Objective: To identify genes whose expression patterns during the pregnancy and post-partum periods are associated with clinical disease activity among women with rheumatoid arthritis (RA). **Methods:** We have established a prospective RA pregnancy cohort, with clinical data and blood samples collected before, during and after pregnancy. Women who improved during pregnancy were selected for analysis (n=8). RA disease activity (DAS28CRP4) was assessed at each time point. Global gene expression profiles were generated using RNA-sequencing (RNA-seq). Raw reads were pseudo-aligned and quantified using kallisto. Random effects regression models were used to test for association between gene expression and RA disease activity (a) during pregnancy (P1), and (b) in the post-partum period (P2). Significance was assessed using a FDR-adjusted threshold of $q < 0.05$ and a fold-change (FC) of at least 2. **Results:** During pregnancy, 444 genes had expression patterns significantly associated with disease activity, of which 20 showed at least 2-fold change in expression. While these 20 genes were not significantly enriched in specific pathways, several had immune-related functions and included the GNLY, FKBP8, CTSE, IFIT3 and IGKV3-20 genes. GNLY encodes a protein that is a component of cytotoxic granules in T cells; FKBP8 is involved in immune-regulation, lymphocyte signaling, and regulation of apoptosis; CTSE is involved in antigen processing; IFIT3 and IGKV3-20 are involved in immune-regulation. In the post-partum period, 2,030 genes had expression patterns significantly associated with disease activity, with 64 of these showing at least 2-fold change in expression from the third trimester. This set of 64 genes was significantly enriched in several immune-related pathways including inflammatory response, defense response, innate immune response and regulation of inflammatory response. Interestingly, there was no overlap between the 20 genes and the 64 genes with expression patterns associated with RA disease activity during and after pregnancy, respectively. **Conclusion:** In this pilot RA pregnancy cohort study with longitudinal RNA-seq data, several candidate genes were identified as significantly associated with disease activity during the pregnancy and post-partum periods. However, the findings also suggest that different mechanisms may be involved in the pregnancy-induced improvement and post-partum flare.

2321F

GWAS suggests an association between lipid metabolism and avascular necrosis of the hip.

*M.M. Lee*¹, *M. Shivakumar*², *S.A. Lietman*², *T.R. Bowen*², *Y. Zhang*¹. 1) Genomic Medicine Institute, Geisinger, Danville, PA; 2) Orthopedic Department, Geisinger, Danville, PA; 3) Biomedical & Translational Informatics Institute, Geisinger, Danville, PA.

Background: Avascular Necrosis (AVN) is the death of bone tissue due to loss of blood to the bone. Trauma, alcoholism and steroids are risk factors for AVN. Familial and twin studies also suggested the existence of a genetic predisposition. In this study, we conducted a GWAS in patients with AVN of the hip from the Geisinger MyCode® Community Health Initiative to search for possible genetic causes of AVN of the femoral head. **Method:** The AVN cohort was identified from electronic health record (EHR) using ICD codes. Cases were patients with either ICD-9 code of 733.42 or ICD-10 code of M87.05. Controls did not have any of these codes. The controls were randomly selected and were matched for age and BMI. Genotype data was generated using Illumina OmniExpressExome. SNPs with marker call rate <99% or minor allele frequency <1% were removed. Samples with missing genotypes > 5% were removed from further analysis. One of the pairs of individuals with first- or second-degree relatedness were also removed from the analysis. A total of 197 cases and 1778 controls were included in the final association tests. Logistic models were adjusted for age, sex, BMI, lipids levels and the first 6 genetic principle components. **Results:** SNPs with suggestive association ($p < 1 \times 10^{-5}$) were identified in the *LMF1* loci, which encodes for lipase mutation factor 1. The most significant SNP is rs3751666 ($P = 4.056 \times 10^{-7}$, OR = 0.5398, 95% CI [0.4253, 0.685]). *LMF1* has been shown to be associated with hypertriglyceridemia. In our analysis, *LMF1* remained associated with AVN after adjusting for LDL, HDL and TG, potentially indicating its direct role on AVN. **Conclusion:** Our results implicates a role of lipid metabolism in the pathogenesis of AVN of the hip. Future studies with a larger sample size are needed to replicate our findings.

2322W

Using co-expression network analysis to inform genome-wide association studies for bone mineral density. O.L. Sabik^{1,2}, G.M. Calabrese¹, C.L. Ackert-Bicknell³, C.R. Farber^{2,4}. 1) Center for Public Health Genomics, School of Medicine, University of Virginia, Charlottesville, VA 22908; 2) Department of Biochemistry and Molecular Genetics, School of Medicine, University of Virginia, Charlottesville, VA 22908; 3) Center for Musculoskeletal Research, University of Rochester Medical Center, University of Rochester, Rochester, NY 14624; 4) Department of Public Health Sciences, University of Virginia, Charlottesville, VA 22908.

Osteoporosis is a disease characterized by reduced bone mineral density (BMD) and increased bone fragility. About half of the variation in BMD across the human population can be explained by genetic variation, and genome-wide association studies (GWASs) have identified over 200 regions of the human genome containing genetic variants influencing BMD. Despite the wealth of genetic signals identified, the genes and mechanisms through which these loci impact bone remain unknown. Here, we hypothesized that we could identify the causal genes underlying GWAS loci for BMD by analyzing GWAS data in the context of a co-expression network. Our approach was based on the idea that causal genes likely influence BMD through the same biological process (e.g. osteoblast-mediated bone formation) and genes involved in specific biological processes are often co-expressed. We applied weighted gene co-expression network analysis (WGCNA), to genome-wide gene expression profiles from mouse osteoblasts and identified 65 modules of co-expressed genes. Of these, one module was significantly enriched for genes located within BMD GWAS loci (odds ratio (OR) = 3.4, $P = 4 \times 10^{-9}$) and its expression was significantly correlated with *in vitro* levels of mineralization ($r = 0.49$; FDR = 0.012). This module was also enriched for skeletal development genes (OR = 8.58; $P < 2.2 \times 10^{-16}$), and genes that result in a bone phenotype when knocked-out (OR = 4.06; $P = 2.14 \times 10^{-9}$). Using independent gene expression profiles measured throughout osteoblast differentiation, we observed that genes in this module were expressed in two distinct patterns, either early or late in osteoblast differentiation, and the enrichments listed above were increased in the late differentiation group. Additionally, we observed that the mean connectivity of GWAS-implicated genes was significantly higher than non-GWAS genes, leading us to prioritize genes based on their degree of connectivity within the module. Of the 30 most highly connected genes in the module, 13 had not been previously linked to bone mass and three, *B4GALNT3*, *CADM1*, and *SLC8A3*, were located within GWAS loci for BMD. For all three genes, their expression in humans was influenced by expression quantitative trait loci (eQTL) that colocalized with the BMD associations and mouse knockouts demonstrated altered BMD. These data have identified three genes underlying BMD GWAS loci and provide a resource for understanding the complex network of genes directing bone mass.

2323T

HLA-C*06:02 genotype is a predictive biomarker of biologic treatment response in psoriasis. N. Dand on behalf of the PSORT consortium. King's College London, London, United Kingdom.

Biologic therapies can be highly effective for the treatment of immune mediated inflammatory diseases, but some patients respond poorly to certain biologics. Predictive biomarkers to guide drug selection in clinical practice could significantly improve patient outcomes and reduce wasted healthcare expenditure. In psoriasis, *HLA-C*06:02* has been established as the primary genetic susceptibility allele. Patients with and without this allele can exhibit clinical differences and there is some evidence associating *HLA-C*06:02* status with response to ustekinumab (a biologic inhibiting IL-12/23 signalling). Using a large national psoriasis registry in the UK, we present the first pharmacogenetics study of comparative response to the two most commonly prescribed biologics for psoriasis: ustekinumab and adalimumab (anti-TNF α). Specifically we sought to test whether *HLA-C*06:02* is an effective predictive biomarker for treatment selection. The study includes longitudinal treatment and response observations and detailed clinical data on an unparalleled scale. HLA alleles were imputed from Illumina HumanOmniExpressExome genotype data for 1,326 patients for whom PASI90 response status (90% reduction in psoriasis area and severity index relative to baseline) was observed after 3, 6 or 12 months of treatment. We developed regression models of PASI90 response, examining the interaction between imputed *HLA-C*06:02* dosage and drug type (adalimumab or ustekinumab) while accounting for potentially confounding clinical variables. A statistically significant interaction was observed at all time-points (most strongly at six months: $P = 3.76 \times 10^{-5}$). Non-carriers of the *HLA-C*06:02* allele were significantly more likely to respond to adalimumab than ustekinumab (odds ratio at six months 2.95, $P = 5.85 \times 10^{-7}$), and this difference was greater when selecting patients with psoriatic arthritis, a common comorbidity (OR = 5.98, $P = 6.89 \times 10^{-5}$). No significant difference in response was seen in *HLA-C*06:02* carriers. Furthermore, results from HLA-wide analysis were consistent with *HLA-C*06:02* itself being the mechanistically relevant allele. We conclude that reference to *HLA-C*06:02* status could offer substantial clinical benefit when selecting treatments for severe psoriasis. Our study design provides a model that could facilitate more effective deployment of these expensive biologic therapies across immune-mediated disease more broadly.

2324F

Genetic susceptibility to autoimmune diseases: The vitiligo model. L. T. Martins¹, H.R. Frigeri², C.C.S. Castro³, M.T. Mira^{1,3}. 1) Graduate Program in Health Sciences, Pontifícia Universidade Católica do Paraná, Curitiba, Brazil; 2) School of Life Science, Pontifícia Universidade Católica do Paraná, Curitiba, Brazil; 3) School of Medicine, Pontifícia Universidade Católica do Paraná, Curitiba, Brazil.

Vitiligo is an acquired, multifactorial pigmentary disease characterized by progressive loss of melanocytes. The reasons leading to melanocyte damage and its subsequent disappearance of the skin remain largely unknown and several hypotheses have been investigated; among them, the most explored by researchers assumes an autoimmune basis for the disease. Supporting this hypothesis is the frequent co-occurrence of autoimmune diseases with vitiligo. In addition, various genetic *loci* associated with vitiligo harbor key immune response genes. Symmetrically, genetic studies involving autoimmune pathologies have identified variants also associated with vitiligo. Our hypothesis is that vitiligo can be used as a model for the study of the genetic component controlling autoimmunity. To test this hypothesis, we investigated for association between vitiligo and genes *CYP27B1*, *REL*, *TNFAIP3* and *IL2/IL21*, all previously related to autoimmune diseases associated with vitiligo. The study was performed using two independent population samples: a discovery, family-based set of 211 trios and a replication set of 131 cases and 119 controls. A total of 14 genetic markers providing complete physical coverage of the candidate genes were genotyped using fluorescence-based platforms ABI 7500 or Thermo-Life QuantStudio 7 Flex. Statistical analysis was performed using the transmission disequilibrium test (for the family-based population) and multivariate logistic regression techniques. Statistically significant association with vitiligo was detected between rs6545836 of the *REL* gene and rs2069763 of the *IL2* gene in the family-based sample. Both association signals were concentrated among patients displaying autoimmune comorbidity and non-segmental vitiligo, consistently with our hypothesis. Evidence for validation was detected for rs2069763 in the case-control sample. Our findings suggest *REL* and *IL2* as new vitiligo susceptibility genes and reinforce the hypothesis of a shared genetic mechanism controlling vitiligo and other autoimmune diseases.

2325W

Non-MHC loci associated with psoriatic arthritis reveal evidence of a potential mechanism for joint involvement. M.T. Patrick¹, P.E. Stuart¹, K. Raja^{1,2}, G.R. Abecasis³, J.J. Voorhees¹, T. Tejasvi^{1,4}, J.E. Gudjonsson¹, V. Chandran^{5,6,7,8}, P. Rahman⁹, D.D. Gladman^{5,6,7}, R.P. Nair¹, J.T. Elder^{1,4}, L.C. Tsoi^{1,3,10}. *An international group of investigators studying psoriasis genetics.* 1) Department of Dermatology, University of Michigan, Ann Arbor, MI; 2) Morgridge Institute for Research, University of Wisconsin, Madison, WI; 3) Department of Biostatistics, Center for Statistical Genetics, University of Michigan, Ann Arbor, MI; 4) Ann Arbor Veterans Affairs Hospital, Ann Arbor, MI; 5) Department of Medicine, Division of Rheumatology, University of Toronto, Toronto, Ontario, Canada; 6) Centre for Prognosis Studies in the Rheumatic Diseases, Krembil Research Institute, University of Toronto, Toronto, Ontario, Canada; 7) Institute of Medical Science, University of Toronto, Toronto, Ontario, Canada; 8) Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada; 9) Memorial University, St. John's, Newfoundland, Newfoundland and Labrador, Canada; 10) Department of Computational Medicine & Bioinformatics, University of Michigan, Ann Arbor, MI.

The heritability of psoriatic arthritis (PsA) is higher than for psoriasis in general, yet no markers outside the major histocompatibility complex (MHC) have so far been found to distinguish PsA from cutaneous-only psoriasis (PsC) with genome-wide significance. Nevertheless, we showed in a recent meta-analysis of PsA vs. PsC associations, utilizing ~10 million markers from over 7,000 psoriatic patients across 6 cohorts, that non-MHC loci contribute significantly to the risk of psoriatic arthritis among psoriatic patients. We hypothesized that an integrative approach, utilizing additional chromatin accessibility, pharmacologic and transcriptomic data to study non-MHC genetic loci showing suggestive evidence of association ($p < 1 \times 10^{-4}$), could provide valuable information about the mechanisms contributing to the pathogenesis of PsA. We first demonstrated that these suggestively associated markers are significantly enriched with active regulatory elements (predicted using H3K27ac chromatin marks) for chondrogenic differentiated cells ($p = 3.99 \times 10^{-5}$, OR=3.09) and osteoblasts ($p = 3.33 \times 10^{-4}$, OR=2.66). Furthermore, genes close to these markers are targets of existing anti-PsA drugs (e.g. the NSAIDs diclofenac, indomethacin and piroxicam) and have been implicated in Wnt signaling. For example, three of the most significant distinct loci (encompassing *FZD4*, *FUT8* and *RUNX1*, respectively) overlap with active enhancer marks, are potential PsA drug targets, and are involved in Wnt signaling. Previous research has suggested dysregulation in Wnt signaling can lead to the excessive bone formation seen in PsA, and we found that expression levels of several genes participating in the Wnt signaling pathway, including *WNT5A*, *DKK2* and *FZD8*, are elevated (FDR $\leq 5\%$) in blood among psoriatic arthritis patients after they developed PsA, when comparing with pre-conversion ($p = 5.26 \times 10^{-4}$, OR=2.09). In fact, all six members of the Frizzled family of Wnt receptors that are differentially expressed in post-conversion blood manifest elevated, as opposed to decreased expression post-conversion ($p = 0.0225$). Our study identifies testable mechanisms by which non-MHC loci may impact PsA risk via their effects upon bone and joint physiology.

2326T

Association of MCP-1 and ACE genetic variants with osteoporotic population of Punjab: A pilot study. M. Singh, P. Bhardwaj, G. Singh, H.S. Sandhu, V. Sambyal, A.J.S. Bhanwer. Guru Nanak Dev University, Amritsar-143005, Punjab, India.

Osteoporosis is a systemic skeletal disease, characterized by low bone mass and micro-architectural deterioration of bone tissue. The condition predisposes to an increased risk of fracture, is a major health problem worldwide and called as "silent disease" because it progresses without symptoms until a fracture occurs. Besides environmental correlates, twin and family studies showed that bone mineral density (BMD) and bone turn over are under strong genetic control. Monocyte chemoattractant protein-1 (MCP-1) is a member of the C-C chemokine family, and a potent chemotactic factor for monocytes. MCP-1 has emerged as a potential genetic variant that plays a strong genetic role in skeletal homeostasis including acquisition and maintenance of bone mass. This study enrolled 221 subjects (116 osteoporotic cases and 105 healthy controls) from Punjab (North-West India). MCP-1 (rs3917887) Ins/Del polymorphism was genotyped and analyzed in all subjects using PCR. No significant association of rs3917887 polymorphism was observed with osteoporosis. Furthermore, the data was analyzed on the basis of gender. In males, D allele [$p=0.03$, $1.81(1.05-3.15)$] and dominant model (ID+II vs DD) [$p=0.02$, $2.48(1.1-5.62)$] conferred increased risk towards pathogenesis of osteoporosis. However, females revealed no association with the osteoporosis etiology. Though, osteoporosis is most commonly found in post-menopausal women because of oestrogen deficiency and is thought of as a women's disease. Nevertheless fractures or bone loss also occur in men and majority of males with fractures remain undiagnosed. The progression of osteoporosis in males is scanty but cannot be overlooked as observed in this study. However, more SNPs on larger data set will be replicated to establish better association of MCP-1 with osteoporosis. The financial assistance to M.S. by DST-PURSE, Gol is acknowledged.

2327F

Association of knee osteoarthritis with Pvull polymorphism and serum levels of Col2a1 gene. R.Nath. Srivastava¹, A.C. Sharma¹, S.R. Srivastava¹, S. Raj². 1) Orthopaedic surgery, King George's Medical University, Lucknow, Uttar Pradesh, India; 2) westminster college, Salt Lake City, UT, USA.

Objective: To investigate the relationship between Pvull single nucleotide polymorphism (SNP rs1635560) and knee osteoarthritis (KOA) in North Indian Population **Methods:** A total of 300 cases that confirmed radiographic KOA and equal number of age- sex matched healthy controls were enrolled. In this study cases consisted of men and women ≥ 40 years that fulfilled American College of Rheumatology (ACR) clinical and radiographic criteria for KOA. Venous blood samples were obtained from all cases as well as controls for genetic and expression analysis. Polymerase chain reactions were performed for genetic analyses using specific primer. In addition, the effect of serum levels of SNP rs1635560 on cases and controls were examined by enzyme-linked immunosorbent assay (ELISA) for expression analysis. **Results:** SNP rs1635560 of Col2a1 was associated with KOA ($P < 0.01$). TT genotype was significantly ($p < 0.01$) associated with KOA as compared to CC genotype, However CT was not associated ($p < 0.078$). In addition when alleles were compared, T allele was observed to be significantly associated with KOA ($P < 0.002$). The level of Col2a1 in serum significantly increased in KOA group, compared with those in control group ($P < 0.05$, χ^2), with 30.2 ± 7.8 ng/ml (Mean $\pm$ SD) in the OA group and 15.6 ± 4.2 ng/ml in the control group. **Conclusion:** Our data indicate that genetic variation in the Col2a1 gene is involved in the risk of KOA in North Indian population, confirming the results from previous studies on the potential importance of this gene in the pathogenesis of KOA.

2328W

Identification of novel risk regions from TWAS and cross-trait genetic correlations for Dupuytren's disease. M. Major¹, M. Ng², D. Furniss^{2,3,4}, B. Pasaniuc⁵, R. Ophoff^{6,7}. 1) Bioinformatics Interdepartmental Program, University of California Los Angeles, Los Angeles, CA; 2) Nuffield Department of Orthopaedics, Rheumatology, and Musculoskeletal Science, University of Oxford, Oxford, UK; 3) Department of Plastic and Reconstructive Surgery, Oxford University Hospitals NHS Foundation Trust, John Radcliffe Hospital, Oxford, UK; 4) NIHR Biomedical Research Centre, NDORMS, University of Oxford, Oxford, UK; 5) Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA; 6) Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA; 7) Center for Neurobehavioral Genetics, University of California Los Angeles, Los Angeles, CA, USA.

Dupuytren's disease is the most common inherited connective tissue disorder, characterized by progressive and irreversible fibroblastic proliferation affecting the hand's palmar fascia, with an onset typically in the sixth decade of life. Previously, 24 risk regions for Dupuytren's disease have been identified from genome-wide association studies (GWAS). This study sought to expand understanding of this disease by: (1) identifying more risk regions via a transcriptome-wide association study (TWAS) and (2) understanding the genetic relationship between Dupuytren's disease and 38 other complex traits. We performed TWAS using expression from 47 different tissues from 6 consortia spanning ~15,000 unique genes. A novel risk region was identified from the 43 significant gene-tissue models (at least 0.5 MB away from the known 23 risk regions from GWAS). Testing 38 traits with Dupuytren's disease using cross-trait LD Score Regression, we identified that BMI, HDL, and TG had significant genetic correlation with Dupuytren's disease. Furthermore, we found that our results were consistent with previous epidemiological findings that lower BMI increases risk for Dupuytren's disease. We investigated these findings further with these traits using p-HESS, which uses local SNP-heritability to quantify local genetic correlation. Using p-HESS we identified 8 regions significantly correlated with BMI and 3 regions significantly correlated with HDL. From this study we have identified a novel disease associated gene on chromosome 17. We also showed novel genetic links to BMI, HDL, and TG, with specific genetically linked regions found for BMI and HDL. Biologically, these regions need to be researched further to understand the broader implications..

2329T

Investigation of CD207 and atopic dermatitis: Follow-up of a GWAS locus. L. Paternoster, A. Budu-Aggrey, M.S Babaei, T. Battram. MRC Integrative Epidemiology Unit at the University of Bristol, Bristol, United Kingdom.

Introduction: A genomewide association study conducted by the EAGLE Eczema consortium in 2015 identified ten novel loci associated with atopic dermatitis (AD). One of the most interesting associations was at the 2p13.3 locus. The top SNP at this locus (rs112111458) is intergenic, but the nearby CD207 was a promising candidate for the causal gene. CD207 (langerin) is a transmembrane receptor, expressed in dendritic cells (e.g. Langerhan cells in the epidermis and mucosa) and is involved in antigen processing. **Aim:** We attempted to identify the causal gene at this locus using skin expression data and further sought to narrow the causal variant credible set at this locus. **Methods:** GTEx contains eQTL results from 414 sun-exposed and 335 non-sun-exposed skin samples. TwinsUK has skin RNAseq data, AD case/control information and genome-wide genotype data available for 672 individuals. Linear mixed models (in GEMMA) were used to test for cis-eQTL associations in TwinsUK between 25 genes and 5172 variants in the region. We also conducted eQTL analysis in controls only. We conducted colocalization analysis between the TwinsUK results and the GWAS results, to test if both signals were driven by the same underlying causal variant. We also undertook analysis to attempt to determine the causal variant at this locus, using GCTA-COJO and FINEMAP. **Results:** rs112111458 was a strong eQTL for CD207 in both GTEx skin samples ($p=7 \times 10^{-7}$ and 2×10^{-5}), with the AD risk allele associated with higher levels of CD207 expression. This SNP was not an eQTL for any other gene in GTEx. This was confirmed in the TwinsUK analysis ($p=4 \times 10^{-10}$). Colocalization analysis confirmed that the eQTL and GWAS results shared a causal variant (posterior probability (PP)= 93%). There was far less evidence for colocalization with any of the other 25 genes tested (next greatest PP was 15% for ANKRD53). The association between rs112111458 and CD207 expression was also present when restricting to controls only ($p=3 \times 10^{-10}$), suggesting that this association is unlikely to be driven by disease-status altering expression of this gene. Our fine-mapping analysis suggested that there was only one causal variant, but the 95% credible set contained 511 SNPs. However, the PP for rs112111458 (67%) was far greater than the next best SNP (rs6723629, PP=4%). **Conclusion:** Our results suggest that CD207 is likely to be the causal gene at the 2p13.3 locus. Our analyses were unable to determine the causal variant with any certainty.

2330F

Association of CCL2 gene polymorphisms and their serum levels with susceptibility to knee osteoarthritis in North Indian population. A.Chandra. Sharma¹, R.N. Srivastava¹, S.R. Srivastava¹, S. Raj². 1) Orthopaedic surgery, King George's Medical University, Lucknow, Uttar Pradesh, India; 2) Westminster College, Salt Lake city, Utah, USA.

Objective: To investigate the possible association between promoter polymorphisms rs1024611 and the polymorphism rs4586 in exon 2 of the chemokine (C–C motif) ligand 2 (CCL2) gene and knee osteoarthritis (KOA) in a North Indian population. **Methods:** DNA was obtained from 300 primary knee OA patients and 300 healthy controls. CCL2 genomic variants (rs1024611 and rs4586 polymorphisms) were detected by polymerase chain reaction restriction fragment length polymorphism (PCR-RFLP). In addition, the effect of serum levels of rs1024611 and rs4586 on cases and controls were examined by enzyme-linked immunosorbent assay (ELISA). **Results:** The rs1024611 A/G promoter polymorphism was associated with OA [genotype frequency, $p=0.007$; allele frequency, $p=0.01$, odds ratio (OR) =1.9, 95% confidence interval (CI) = 1.18 –3.07]. Significant association was observed between the G carrier of the rs1024611 A/G promoter polymorphism and primary knee OA patients ($p=0.01$, OR=1.7, 95% CI= 1.12 – 2.7). However, no significant difference was found in the rs4586 polymorphism. Haplotype frequency analysis revealed a significant difference ($\chi^2=8.01$, $p=0.02$). The CCL2 serum level of subjects with the G carrier (285.0 ± 86.5 pg/mL) of the rs1024611 A/G promoter polymorphism was statistically higher than that of subjects with the non-G carrier (160.5 ± 47.8 pg/mL). Further in relation with clinical severity of OA, we observed significant association of the G carrier of the rs1024611 A/G promoter polymorphism with both Visual analogue scale (VAS) ($p<0.01$) and Western Ontario and McMaster Universities (WOMAC) score ($p<0.05$). **Conclusions:** The G carrier of the rs1024611 A/G promoter polymorphism was found to be associated with primary knee OA, and could be a susceptibility factor in the development of primary knee OA in the North Indian population.

2331W

Targeted fine mapping and functional evaluation of chromosome 15q21-22 identifies AQP9 as a novel gene associated with keloid risk. T.L. Edwards¹, J.M. Keaton¹, J.N. Hellwege¹, S.M. Williams², D.M. Supp³, S.B. Russell⁴, D.R. Velez Edwards⁵. 1) Division of Epidemiology, Department of Medicine, Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN; 2) Division of Quantitative Sciences, Department of Obstetrics and Gynecology, Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN; 3) Division of Dermatology, Department of Medicine, Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN; 4) University of Cincinnati, College of Medicine Senior Investigator, Shriners Hospitals for Children - Cincinnati Burns Hospital; 5) Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland USA.

Keloids are benign dermal tumors that occur ~20 times more often in African versus European descent populations. While most keloids are sporadic, both familial aggregation and large differences in prevalence among populations support genetic risk. Prior studies identified variants at chromosome 15q21-22 associated with keloid risk in African and Asian ancestry populations without consensus for the causal variants. We conducted a fine mapping study using targeted next-generation sequencing of this region in 246 African-American samples using electronic health records ($n=119$ cases, 121 controls), and also functionally assessed keloid cellular phenotypes in culture. Common variant (MAC>20) and nonsynonymous gene-based rare variant associations within this region were evaluated adjusting for sex and the first 5 principal components from the LASER method. Genetically predicted gene expression (GPGE) association analyses were performed with S-PrediXcan to identify keloid gene-tissue combinations. The strongest single variant association was at rs72073238 ($P=2E-4$), odds ratio [OR]=2.4) near *POLR2M*, and a keloid risk-increasing association with increasing *POLR2M* GPGE was detected in transformed fibroblasts ($P=5E-3$). The *ALDH1A2/AQP9* common variant rs55803039 was also associated ($P=2E-4$), and *AQP9* was nominally significant in rare-variant gene-based SKAT analyses ($P=0.02$). Increasing *AQP9* GPGE was negatively associated with keloid risk in lymphocytes ($P=3E-3$). *AQP9* associations were robust to adjustment for BMI and the previously detected variants from this region, rs747722 (*MYO1E*) and rs8032158 (*NEDD4*). Characteristics of candidate genes identified from these and published association results were evaluated in keloid-derived and normal fibroblasts. We observed a 20-fold reduction in *AQP9* expression in keloid compared to normal fibroblasts ($P=7E-6$), which is consistent with the GPGE result. We observed an approximate 2-fold decrease in *IGFBP5* expression in keloid cells after silencing *NEDD4* ($P<0.05$), but no significant changes in the expression of other keloid markers. Silencing of either *NEDD4* or *MYO1E* decreased maximum culture density of both normal and keloid fibroblasts. These results suggest that nearest gene annotation for noncoding SNPs may not identify causal genes and that incorporating GPGE may improve gene identification. Our findings further support the chr15q21.2-22.3 region as containing keloid risk factors and that *AQP9* is a promising candidate.

2332T

Whole-genome sequencing and genome-editing implicate missense variants in *LRRK2* and *NOD2* in early-onset leprosy. M. Dallmann-Sauer^{1,2,3,4}, Y.Z. Xu^{1,2,3}, W. Correa-Macedo^{2,3,5}, A.L.F. Costa⁶, G.E. Cambri⁴, A. Alcaïs^{7,8,9}, J. Manry^{1,2,3}, V.M. Fava^{1,2,3}, C.M. Probst¹⁰, M. Mira⁴, E. Schurr^{1,2,3,4,5}. 1) Departments of Human Genetics and Medicine, Faculty of Medicine, McGill University, Montreal, Quebec, Canada; 2) Program in Infectious Diseases and Immunity in Global Health, Research Institute of the McGill University Health Centre, McGill University, Montreal, Quebec, Canada; 3) McGill International TB Centre, McGill University, Montreal, Quebec, Canada; 4) Core for Advanced Molecular Investigation, Graduate Program in Health Sciences, School of Medicine, Pontifícia Universidade Católica do Paraná, Curitiba, Paraná, Brazil; 5) Department of Biochemistry, Faculty of Medicine, McGill University, Montreal, Quebec, Canada; 6) State University of Piauí, Teresina, Piauí, Brazil; 7) Laboratory of Human Genetics of Infectious Diseases, Necker Branch, Institut National de la Santé et de la Recherche Médicale U1163, Paris, France; 8) University Paris Descartes, Imagine Institute, Paris, France; 9) St. Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, Rockefeller University, New York, NY, United States of America; 10) Laboratory of Functional Genomics, Instituto Carlos Chagas-ICC-FIOCRUZ, Curitiba, Paraná, Brazil.

Despite the availability of an effective treatment, leprosy caused by the human pathogen *Mycobacterium leprae* still affects over 200,000 new patients annually. Leprosy has a long incubation period; hence, cases with age-of-onset younger than 4 years are rare and represent an extreme end of the age distribution. Here, we studied monozygotic twins exhibiting clinical symptoms of leprosy at the age of 22 months. The early onset of disease in the twins raised the possibility of a strong genetic effect on leprosy susceptibility. Moreover, both the father and the paternal grandmother were affected by leprosy. This familial aggregation of cases further supported the hypothesis of strong genetic effects acting on leprosy susceptibility in the family. To investigate the genetic contribution to leprosy susceptibility in this family, we performed whole-genome sequencing in six family members. After variant calling, stepwise procedures of filtering and variant prioritization were applied to identify those variants that were most likely to be causal. A number of different approaches were used to determine co-segregation of coding variants with leprosy susceptibility in the family, taking into account age-at-diagnosis and possible model of inheritance. A total of 95 and 37 candidate variants were identified assuming recessive or dominant genetic models of inheritance of a putative leprosy susceptibility gene, respectively. Two missense variants – N551K and R1398H – homozygous in the twins located to *LRRK2*, a gene previously associated with susceptibility to leprosy phenotypes. Interestingly, these variants have been associated with protection from Parkinson's disease and Crohn's disease. In addition, the strong Crohn's disease risk variants P268S and R702W of *NOD2* were carried heterozygously by the twins and father suggesting a joint effect of homozygosity at *LRRK2* and heterozygosity at *NOD2* on leprosy risk. To follow-up on the genetic results, RAW macrophages were genome-edited by CRISPR/Cas9 to carry one or both *LRRK2* candidate variants. *M. leprae* triggered production of microbicidal reactive oxygen species in RAW cells was lower for the pS1398 variant and the double mutant compared to wild-type cells. These results implicate *LRRK2*, an established risk factor for Parkinson's disease and Crohn's disease, in leprosy susceptibility.

2333F

Genotype-first phenotyping of 32 patients with post-zygotic *GNAQ* or *GNA11* mutations. M. Jordan¹, V. Carmignac², A. Sorlin², P. Kuentz², C. Philippe², I. Investigators MUSTARD Cohort¹, P. Vabres^{1,2}. 1) CHU F Mitterrand, Dijon - FRANCE, France; 2) GAD team, INSERM UMR1231, Dijon, France.

Introduction: Post-zygotic mutations in *GNAQ/GNA11* genes account for mosaic conditions with vascular or pigmentary anomalies. We sought to delineate the phenotype of 32 patients harboring an activating post-zygotic mutation in *GNA11* or *GNAQ*. Method: All patients were studied because of cutaneous vascular anomalies suggestive of involvement of G-proteins. *GNAQ* and *GNA11* mutations were identified on ultradeep NGS on affected skin. A postzygotic *PIK3CA* mutation was ruled out in 9 patients. Cutaneous capillary malformations (CM) were subclassified as suggested by Happle : naevus flammeus, naevus roseus, cutis marmorata, and naevus anemicus. Patients with segmental asymmetry were classified as one of the six types defined by Oduber et al. Results: CM were observed in all patients, involving the face in 65.6% of cases (n=21), and were extensive with multiple affected segments in 71.8% (n=23). Naevus flammeus, segmental overgrowth, varicose veins (with deep vein thrombosis or venous agenesis) and macrocephaly were more common in patients with *GNAQ* mutations, whereas cutis marmorata (p=0.042), naevi anemicus (p= 0.03), and segmental hypotrophy were more frequent in patients with *GNA11* mutations. Other features did not predominate in either *GNA11* or *GNAQ* patients. Pigmentary anomalies were observed in 11 patients and associated with skin types III-IV (p=0.006). Neurological anomalies (epilepsy, mental retardation, or MRI brain anomalies) in 9 and glaucoma in 8. Six patients had high blood pressure (HBP) (p=0.003). Among them, renal artery dysplasia and renal hypoplasia were found. Discussion: Hence, in addition to well-known features associated with mosaic G-protein mutations (e.g. Sturge-Weber syndrome), we found previously undescribed manifestations: segmental overgrowth, arterial hypertension and renal artery or kidney anomalies. Risk of thrombosis, HBP and renal anomalies should be considered in the follow-up of these patients. We also found distinct clinical features possibly allowing to discriminate between *GNAQ* and *GNA11* mutations.

2334W**Multiscale causal network models identify VGF as an Alzheimer's disease gene.**

N.D. Beckmann¹, W.J. Lin^{1,2}, M. Wang¹, A.T. Cohain¹, A.W. Charney¹, M. Ehrlich¹, B. Zhang¹, S.R. Salton¹, E.E. Schadt¹. 1) Icahn School of Medicine at Mount Sinai, New York, NY; 2) Medical Research Center of Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, Guangdong, China.

Though discovered over 100 years ago, the molecular foundation of sporadic Alzheimer's disease (AD) remains poorly understood. To further characterize the complex nature of this disease, large scale data integration and advanced modeling are necessary. Here, using cutting edge bioinformatics approaches, we built multiscale causal networks on a large multi-omics dataset of AD. We integrated DNA variation, gene and protein expression into models enabling the detection and prioritization of *VGF* as an important key driver of the disease. To validate this at the functional and molecular level, we overexpressed *VGF* in 5xFAD mice. We found that *Vgf* overexpression partially rescued the AD phenotype in mice and that genes differentially expressed in this model were enriched for genes our human *VGF* subnetwork predicted would change in response to changes in *VGF*. Our findings support a causal, protective role for *VGF* in regulating AD.

2335T**Ascertainment and study design for the Resource for Early-Onset Alzheimer Disease Research (READR), a genomics resource for EOAD.**

G.W. Beecham¹, G. Garbiso¹, P. Baez², I. Martinez², L. Adams¹, T. Ayodele², M. Jean-Francois¹, M. Pericak-Vance¹, M. Cuccaro¹, C. Reitz². 1) John P Hussman Institute for Human Genomics, University of Miami, Miami, FL; 2) Taub Institute for Research on Alzheimer Disease, Columbia University, New York, NY.

Background: Genomic studies of Alzheimer disease (AD) have primarily focused on non-Hispanic White (NHW) participants affected by the late-onset form of the disease (AAO>65), or the study of early onset AD (EOAD; onset age <=65) cases from families showing Mendelian inheritance patterns associated with mutations in the *APP*, *PSEN1* and *PSEN2* genes. However, mutations in these three genes explain ~10% of EOAD cases, and among mutation carriers age at onset is often highly variable. To begin addressing this issue, we began the *Resource for Early-onset Alzheimer Disease Research* (READR) initiative to assemble and assess a large set of well-characterized multiplex EOAD families without known EOAD mutations from various ethnic groups as a resource for genomic studies. Methods: We have begun ascertainment of multiplex (2 or more affected) EOAD families using two methods. First, we utilized our existing community networks to identify families. This approach involves community outreach and collaborations with local neurology clinics. Our second approach involves collaborating with patient and caregiver support groups to identify families with positive family history who are interested in participating in genomics research studies. All families are being assessed using extensive neurological, medical and batteries neuropsychiatric batteries for deep cognitive endophenotyping. Results: Through our multipronged outreach approach we have identified nearly 175 possible/definite qualifying families (~50% non-Hispanic White, ~30% Hispanic, ~20% African American), validating our ascertainment methods. 40 of these families have already received whole-genome sequencing in a first batch; a second batch of 60 families is currently being prepared for sequencing. The remaining 75 families have been identified and contacted and will need to be thoroughly screened and ascertained. Through additional planned outreach efforts we expect to ascertain another eligible 20-30 families, yielding upwards of 200 families in total. Conclusion: We have established READR as a genomics resource for EOAD research. The substantial increase in sample size over existing EOAD efforts and the inclusion of minority populations will significantly enhance the ability to identify causative variants, validate variants across ethnic groups, and clarify the distinction of unexplained EOAD from EOAD with known mutations and late-onset AD.

2336F

An Alzheimer's disease stem cell model of a rare *SORL1* mutation. *H.N. Cukier^{1,2,3}, F.S. Johnson¹, C. Garcia-Serje¹, J.M. Vance^{1,2,3}, M.L. Cuccaro^{1,3}, M.A. Pericak-Vance^{1,2,3}, D.M. Dykxhoorn^{1,3}.* 1) John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL; 2) Department of Neurology, University of Miami, Miami, FL; 3) Department of Human Genetics, University of Miami, Miami, FL.

While genetics has been shown to play an integral role in Alzheimer's disease (AD) risk for over two decades, understanding the underpinning mechanisms which cause disease is still ongoing. We propose to focus this study on *SORL1*, a gene that has been implicated in both early and late onset AD. We have identified two siblings with late onset AD who carry a single base pair deletion (g.4292delG) that is predicted to result in a frameshift alteration and early termination of the protein (p.Cys143Ser*1). To understand the mechanism by which the *SORL1* alteration may be acting to contribute to AD risk, induced pluripotent stem cells (iPSC) lines were developed to examine cellular consequences in neuronal tissue, which can otherwise only be collected post-mortem. Peripheral blood mononuclear cells (PBMCs) were extracted from the whole blood of two siblings bearing the *SORL1* change, as well as two aged, and ethnically matched control individuals. iPSC reprogramming was performed using Sendai virus. Each iPSC line generated was validated for pluripotency through immunocytochemical staining and shown to be negative for any large-scale chromosomal abnormalities via karyotyping. These iPSC lines will be promoted to generate excitatory neurons through the overexpression of *NGN2* and create a disease relevant tissue to assess. General characterization of the cellular viability and morphology will be performed. Alzheimer-specific phenotypes such as the levels of secreted beta amyloid and intercellular whole and phosphorylated tau will also be measured. The *SORL1* protein has been shown to act in APP trafficking and processing; therefore, this will be assessed in aged neurons. Utilizing patient-specific iPSC lines will permit assessing the cellular and molecular consequences of genetic alterations that may contribute to AD risk. We will present the morphological and cellular phenotypes of these lines carrying a potential risk factor for AD.

2337W

Identification of novel noninvasive disease biomarkers for susceptibility to Alzheimer's disease progression. *S. Esmaeeli Nieh¹, Q. Guo², A. Vasanthakumar¹, J.W. Davis¹, K. Idler¹, R. Comley², J. Waring¹.* 1) Genomics Research Center, AbbVie, North Chicago, IL; 2) Translational imaging, AbbVie, North Chicago, IL.

Introduction Alzheimer's disease (AD) is a devastating neurodegenerative disorder. AD patients display hippocampal atrophy, memory impairment, and cognitive decline. Despite research efforts and a substantial unmet medical need, no effective cure has been identified and treatment remains symptomatic. Early diagnosis is essential to start treatment in a timely manner, and in treating the disease more effectively. However, reliable early molecular diagnosis has been a challenge, and current methods are invasive, time-consuming, and/or expensive. Objective In this study we sought to integrate DNA methylation data and brain endophenotypes within the very well-characterized Alzheimer's disease Neuroimaging Initiative (ADNI) dataset to identify novel noninvasive disease biomarkers for susceptibility to disease progression. Method We performed cross-sectional analysis on a cohort of ADNI patients who had both methylation and MRI data. We included a total of 260 samples comprised of 83 cognitively normal (CN), 140 mild cognitive impairment (MCI), and 37 Alzheimer's disease (AD) patients. Total gray matter and hippocampal volumes were used as endophenotypes for the association analysis. FreeSurfer was used to extract total gray matter and hippocampal volumes, as well as estimated total intracranial volume (eTIV). Differentially Methylated Positions (DMPs) were analyzed using routinely applied DNA methylation analysis pipelines. To study associations between peripheral blood methylation markers and brain endophenotypes, generalized linear mixed modeling regression was applied; these analyses were conducted using eTIV and age as co-variates. Results and Conclusion The cross-sectional analysis of differential DNA methylation yielded 483, 139, and 299 DMPs for the AD vs CN, AD vs MCI, and MCI vs CN comparisons respectively. We found DNA methylation changes in genes associated with neurodegeneration (e.g. *BDNF*), and in some novel genes associated with neurodevelopment and synaptic function (e.g. *IQSEC1*). In this study we integrated the DNA methylation dynamics with brain endophenotypes. Initial analysis of the association between DMPs and total gray matter and hippocampal atrophy revealed correlations with genes known to be involved in neuronal function. These findings may support the identification early biomarkers of disease progression, and shed light on the biological underpinnings of that progression, potentially helping identify novel drug targets.

2338T

Whole-exome sequencing (WES) analysis of Alzheimer's disease in non-*APOE4 carriers.** K. Fan, S. Rosenthal, E. Feingold, F. Demirci, O. Lopez, M. Kamboh. University of Pittsburgh, Pittsburgh, PA.

The genetics of late-onset Alzheimer's disease (AD) is complex due to the heterogeneous nature of the disorder. *APOE**4 is the strongest genetic risk factor for AD. Genome-wide association studies (GWAS) have identified more than 20 additional loci, each having relatively small effect size. Including *APOE**4, the known loci explain only about 30% of the genetic variance and thus much of the AD genetic variance remains unexplained. In an effort to identify some of the missing heritability of AD, we have performed whole-exome sequencing (WES) analysis focusing on non-*APOE**4 carriers from two WES datasets. The first dataset was an AD case-control sample (n=773) biased against *APOE**4 carriers. After quality control measures and removing subjects with *APOE**4, the non-*APOE**4 sample consisted of 372 AD cases (mean age-at-onset (AAO): 75.95 ± 7.09 years, 62% female) and 348 controls (mean age: 78.59 ± 4.72 years, 66% female). The second dataset was derived from the Alzheimer's Disease Sequencing Project (ADSP) that was downloaded from dbGaP (phs000572.v7.p4) and the non-*APOE**4 sample comprised 2,113 cases (mean AAO: 76.43 ± 7.98 years, 57% female) and 5,139 controls (mean age: 85.68 ± 3.93 years, 61% female). Following individual WES analysis, we performed WES meta-analysis for overlapping markers present in both datasets. Among the top five SNPs, two were from known AD loci (*APOE* and *MS4A6A*). The well-established protective *APOE*/rs7412 SNP coding for *APOE**2 was the top hit (odds ratio (OR)=0.39; *P*=1.34E-23). *MS4A6A*/rs12453 was the third most significant SNP (OR=0.82; *P*=1.16E-05). The other three SNPs were from three novel loci. The second most significant association was seen on chromosome 8 in *PLEC* (OR=1.45; *P*=3.18E-06) followed by *ABCD3* (OR=5.65; *P*=4.38E-05) on chromosome 1 and *STK33* (OR=2.87; *P*=7.08E-05) on chromosome 11. All three SNPs from these novel loci were synonymous mutations, and the top *PLEC* SNP was cis-eQTL for multiple genes in the GTEx data in the *PLEC* gene region, indicating its important biologic significance. In addition to affecting *PLEC* RNA expression in whole blood (*P*=3.17E-04) and brain hippocampus (*P*=3.02E-02), the top *PLEC* SNP was eQTL for nearby *PARP10* in the blood (*P*=3.62E-06) and for *SLC52A2* in brain frontal cortex (*P*=9.61E-04). Further characterization of the genes that are affected by this SNP may help to shed light on the roles of these genes in AD etiology. This study was supported by NIH grants AG030653 and AG041718.

2339F

Mitochondrial variants and haplogroups associated with late-onset Alzheimer's disease identified by whole exome sequencing. J.J. Farrell¹, X. Zhang^{1,2}, C. Zhu¹, R. Mayeux³, J.L. Haines⁴, M.A. Pericak-Vance⁵, G. Schellenberg⁶, K. Lunetta², L.A. Farrer^{1,2} on behalf of the Alzheimer's Disease Sequencing Project (ADSP). 1) Section of Biomedical Genetics, Boston University School of Medicine, Boston, MA, USA; 2) Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA; 3) Columbia University, New York, NY, USA; 4) Case Western Reserve University School of Medicine, Cleveland, OH, USA; 5) University of Miami Miller School of Medicine, Miami, FL, USA; 6) University of Pennsylvania, Philadelphia, PA, USA.

Background: Numerous studies suggest that mitochondrial dysfunction is involved in Alzheimer disease (AD) pathogenesis. **Methods:** The Alzheimer's Disease Sequencing Project (ADSP) performed whole-exome sequencing (WES) of 5,519 AD cases and 4,917 cognitively normal elderly controls of European ancestry (EA) and 218 cases and 177 controls of Caribbean Hispanic (CH) heritage. After mapping to the reference MT genome (NC_012920), MT variants were called using the haploid option of GATK HaplotypeCaller. Low-quality, multi-allelic and monomorphic variants were excluded using a QC metric based on comparison with MT variants called in the Alzheimer Disease Neuroimaging Initiative WGS dataset and the 1000 Genomes project. HaploGrep2 software was used to classify mtDNA Haplogroups with Phylotree 17. The association of AD with MT single nucleotide variants (SNVs) with minor allele count (MAC) ≥10 and call rate ≥ 0.80 (802 SNVs in EA, 135 SNVs in CH) and with mtDNA haplogroups was evaluated in EAs and CHs separately. In gene-based tests, only SNVs with minor allele frequency (MAF) <0.05 were considered, and genes with ≥2 SNVs and cumulative MAC ≥10 were tested. Analyses were performed using single variant score tests and gene-based (SKAT-O) tests in seqMeta implemented in the R package, controlling for age, sex, and principal components of ancestry. Results were combined across ethnicities using meta-analysis. Bonferroni-corrected thresholds were used to determine study-wide significance (SWS) for single variant (*P*=6.2×10⁻⁶) and gene-based (*P*=3.8×10⁻³) associations. In EAs, association of AD with 47 mtDNA haplogroups was tested using logistic regression. **Results:** After variant-level and sample-level QC, 4,220 variants and 10,610 samples remained for association analyses. A missense SNV in *MT-ND4L* which encodes NADH dehydrogenase subunit 4L involved in ubiquinone activity and oxidoreductase activity (26:10733:C:T, rs28709356, Asp88Glu, MAF=0.002) was associated with AD risk in EAs at nearly the SWS level (OR=7.52; *P*=7.3×10⁻⁶). One SWS gene-based association finding was identified with *MT-TI* (*P* = 3.0×10⁻³). None of the mtDNA haplogroups were nominally associated with AD. **Conclusions:** A set of high quality MT variants and mtDNA haplogroups were called from the large ADSP sequencing data. We identified significant association of AD risk with a rare SNV in *MT-NDL4* and the *MT-TI* gene. A replication study of independent cohorts from the extension of ADSP WGS data is now underway.

2340W

Convergence of biological pathways implicated by transcriptomic analysis in a multi-ethnic Alzheimer disease cohort. A.J. Griswold^{1,2}, S.K. Sivasankaran¹, O.K. Gardner¹, F. Rajabli¹, B.W. Kunkle¹, K. Hamilton¹, J. Jaworski¹, W.S. Bush³, E.R. Martin^{1,2}, G.W. Beecham^{1,2}, G.S. Byrd⁴, J.L. Haines³, M.A. Pericak-Vance^{1,2}. 1) John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL; 2) Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miami, FL; 3) Institute for Computational Biology and Department of Population & Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH; 4) Department of Biology, North Carolina A&T State University, Greensboro, NC.

Dozens of genetic variants conferring risk for and protection from Alzheimer disease (AD) have been identified. However, understanding the role of DNA variation directly on disease mechanisms, particularly in under-represented ethnic groups, is lacking. Because expression studies in readily available material, such as blood, can recapitulate transcriptional changes occurring in neuronal tissues, our goal was to identify transcriptomic changes related to AD from peripheral blood collected from 234 African American (AA) (115 AD, 119 cognitively normal, age matched) and 241 non-Hispanic Whites (NHW) (121 AD, 120 cognitively normal, age-matched). We performed RNAseq of all protein coding genes and identified differentially expressed genes between AD cases and cognitively normal individuals within each ethnicity adjusting for known covariates. A total of 419 genes (291 up-regulated, 127 down-regulated) and 490 genes (353 up-regulated, 137 down-regulated) were differentially-expressed in the AA and NHW datasets, respectively. Among genes previously implicated by genome-wide association studies in AD, two were differentially expressed in NHW (*SPI1* and *EPHA1*) and one in AA (*CR1*). Interestingly, only 16 genes were commonly differential in both ethnicities including upregulation of genes involved in Ab processing (*MMP9* and *APMAP*) and immune modulation (*LTB4R* and *PGLYRP1*). Despite the majority of significantly altered genes differing between ethnicities, pathway enrichment analysis revealed that immune response and RNA processing related pathways were enriched in both datasets. These results demonstrate that analysis of blood can reveal aspects of AD pathophysiology without the use of more invasive methods (e.g. PET scans and lumbar punctures). Furthermore, these data support the hypothesis that there is genetic heterogeneity for AD across racial/ethnic groups but a convergence of underlying processes representing a shared etiology. Studies such as this that identify common mechanisms across ethnically diverse individuals in a genetically heterogeneous disorder such as AD are critical to translating previous discoveries into diagnostic or therapeutic interventions.

2341T

Population-based genome-wide association study of cognitive decline in older adults free of dementia. M. Kamboh, Q. Yan, J. Beer, K-S. Fan, X. Wang, F. Demirci, B. Snitz, C-C. Chang, E. Feingold, M. Ganguli. University of Pittsburgh, Pittsburgh, PA.

Genome-wide association studies (GWAS) have identified over 20 risk loci for late-onset Alzheimer's disease (AD), but common variants at these loci account for only ~30% of the AD genetic variance. Alternative approaches are needed to identify additional genetic factors for AD. An attractive option is the use of AD-related quantitative endophenotypes. In this study, we used decline in cognitive function as a relevant endophenotype that is derived from different cognitive domains having high heritability estimates. We used longitudinal data on 5 cognitive domains (memory, executive function, language, attention and visuospatial ability) from two population-based cohorts, MYHAT (Monongahela-Youghiogheny Health Aging Team) and MoVIES (Monongahela Valley Independent Elders Survey), and conducted GWAS meta-analysis to: a) examine the association of previously reported AD loci with cognitive decline, and b) identify novel loci that affect cognitive decline. The MYHAT sample consisted of 767 participants with mean age of 77.1±7.3 years (61% women, all whites) and the MoVIES sample comprised 378 participants with mean age of 70.1±4.3 years (67% women, all whites). For decline over time in each cognitive domain, we normalized intra-individual slopes within each cohort, accounting for baseline age, sex and years of education. Normalized slope for each domain was used in individual GWAS analysis after including top principal components as covariates followed by GWAS meta-analysis. Among the known AD loci, *APOE*4* showed the strongest association with memory ($P=7.18E-06$) followed by language ($P=1.65E-04$). Two other known loci showed associations with more than one domain at $P<1E-03$, including *EPHA1* with visuospatial and memory ($P=1.24E-05$, $5E-04$), and *ABCA7* with executive function, language, attention and visuospatial ($P=3-7E-04$). A novel locus at genome-wide significant level was detected in *WDFY2* on chromosome 13 with attention ($P=3.37E-08$). In the GTex eQTL analysis, the genome-wide significant SNP was associated with RNA expression levels of *WDFY2* in several brain regions: cerebellar hemisphere ($P=1.07E-04$), cerebellum ($P=6.92E-04$), hippocampus ($P=2.18E-03$) and cortex ($P=2.29E-02$) as well as in whole blood ($P=4.41E-05$). In summary, our meta-GWAS analysis on >1,100 non-demented elderly subjects has identified a novel genome-wide significant locus for the attention domain. This study was supported by NIH grants AG023651, AG07562, AG030653 and AG041718.

2342F

Chromatin accessibility profiling in hiPSC-derived neurons identifies functional noncoding GWAS risk variants of Alzheimer's disease. A. Kozlova^{1,2}, S. Zhang^{1,2}, H. Zhang¹, M. Streit¹, J. Shi³, A.R. Sanders^{1,2}, P.V. Gejman^{1,2}, G. Thinakaran⁴, J. Duan^{1,2}. 1) Center for Psychiatric Genetics, NorthShore University HealthSystem, Evanston, IL; 2) Department of Psychiatry and Behavioral Neurosciences, University of Chicago, IL; 3) Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD; 4) Departments of Neurobiology, Neurology and Pathology, University of Chicago, IL.

Alzheimer's disease (AD) is the most common cause of dementia in the aged. Genome-wide association studies (GWAS) of the AD have identified more than 20 risk loci including the APOE locus. However, each GWAS locus typically spans several genes and many equally associated genome-wide significant SNPs; it thus remains challenging to identify which risk SNPs/genes are likely to be causal. Most GWAS risk variants are noncoding and likely affect gene transcription. A major determinant of transcription is chromatin accessibility (openness) to transcription factors. Here, we consider the allele-specific open chromatin (ASoC) assay as a functional readout of the regulatory variation by directly comparing within the same sample the effects of the two alleles of a heterozygous SNP on the measurements of open chromatin. Using human induced pluripotent stem cells (hiPSCs) as a cellular model, we carried out an open chromatin profiling by ATAC-seq in hiPSC-derived neural progenitor cells (NPCs) and NGN2-induced excitatory neurons (iNs) of 20 individuals with unknown AD status. We found that ~6% of the heterozygous SNPs (~100 K) showed ASoC in iNs and NPCs. ASoC variants were found significantly enriched (~4-fold) for AD GWAS risk variants, and can readily inform putatively regulatory risk variants at three AD loci (*CLU*, *BIN1*, and *CD2AP*). Although these ASoC variants were identified in NPCs and iNs, we have found that the mRNAs and the encoded proteins of *CLU*, *BIN1*, and *CD2AP* are also expressed in non-neuronal cell types such as microglia and oligodendrocytes in AD postmortem brains. To systematically examine their cell-specific gene expression, we used hiPSCs to derive other cell types including GABAergic neurons, microglia, and astrocytes. We found that hiPSC-derived neurons secreted Ab and accumulated Tau as neurons matured. By using CRISPR/cas9 epigenome editing, we are in the process of interrogating which genes are regulated by the putatively functional AD risk variants that exhibited ASoC. We are also extending the ASoC assay to other AD-relevant cell types (microglia and astrocytes) to identify which other AD GWAS risk variants are likely to be regulatory and causal. Our study provides a novel approach for identifying functional noncoding AD GWAS risk variants in hiPSC models. Identifying regulatory AD-risk SNPs and their regulated genes, together with cellular phenotypic characterization, will help bridge the AD GWAS findings to disease biology.

2343W

Independent associations of the *TOMM40* and *APOE* polymorphisms with risks of Alzheimer's disease. A. Kulminski, Y. Loika, K. Arbeev, O. Bagley, S. Ukraintseva, A. Yashin, I. Culminkaya. Biogerontology of Aging Research Unit, Social Science Research Institute, Duke University, Durham, NC.

Previous research reports that the *TOMM40* (Translocase of Outer Mitochondrial Membrane 40) rs2075650 SNP (single nucleotide polymorphism) and the *APOE* (Apolipoprotein E) ϵ 4-allele-coding rs429358 SNP are strongly associated with Alzheimer's disease (AD) when they are analyzed separately. Because these SNPs are in moderate linkage disequilibrium in Caucasians ($r^2 \sim 0.5$), the association of rs2075650 with AD is often attributed to linkage disequilibrium of rs2075650 with rs429358. This conclusion is supported by showing pivotal role of rs429358 in the association with AD in conditional statistical analyses using models with both these SNPs included. We performed meta-analysis of separate and additive associations of rs2075650 and rs429358 with risks of AD in a sample of N=2,518 AD cases and N=15,946 non-cases of Caucasians ancestry from four studies (the Framingham Heart Study, the Cardiovascular Health Study, the Health and Retirement Study, and the National Institute on Aging Late-Onset Alzheimer Disease Family Study) using the Cox proportional hazard regression model adjusted for age and sex. Minor alleles of rs2075650 and rs429358 SNPs were strongly associated with increased risks of AD in additive genetic models, $\beta=0.82$, SE(standard error)=0.04, $p=1.08 \times 10^{-100}$ and $\beta=1.02$, SE=0.04, $p=3.11 \times 10^{-164}$, respectively, when they were analyzed separately. Conditional analysis with both rs2075650 and rs429358 SNPs in the model identified small-effect association of rs2075650 with the increased risk of AD, $\beta=0.14$, SE=0.06, $p=1.00 \times 10^{-2}$, which was independent of strong association of rs429358 with AD, $\beta=0.92$, SE=0.05, $p=9.39 \times 10^{-85}$. Meta-analysis using genotypic model with both rs2075650 and rs429358 SNPs identified 4-fold larger risk for carriers of the rs2075650 minor allele homozygotes ($\beta=0.39$, SE=0.13, $p=2.84 \times 10^{-3}$) compared to heterozygotes ($\beta=0.10$, SE=0.07, $p=1.28 \times 10^{-1}$) and 1.5-fold larger risk for carriers of the rs429358 minor allele homozygotes ($\beta=1.67$, SE=0.12, $p=1.24 \times 10^{-43}$) compared to heterozygotes ($\beta=1.06$, SE=0.07, $p=4.81 \times 10^{-46}$). In addition to confirming more pronounced role of rs429358 compared with rs2075650 in the associations with risk of AD, our analyses identified independent roles of these variants from the *TOMM40* and *APOE* genes in AD and suggested that carriers of the rs2075650 minor allele homozygous genotypes could be at nonlinearly increased risk of AD.

2344T

The Alzheimer's disease sequencing project (ADSP) data update 2018.

Y. Leung¹, A.B. Kuzma¹, K.M. Faber², W.J. Salerno³, L.B. Cantwell⁴, N. Gupta⁵, R. Fulton⁷, C. Dalgard¹², O. Valladares¹, B. Vogel¹, E. McGrath-Martinez¹², E. Appelbaum⁷, S.H. Choi², K.L. Hamilton-Nelson¹¹, Y. Zhao¹, D. Muzny⁶, L. Qu¹, D. Reyes⁹, J. Waligorski⁷, J. Farrell⁴, A.C. Naj¹, J.C. Bis⁸, A.L. DeStefano², S. Seshadri¹⁰, E. Boerwinkle⁶, G.D. Schellenberg¹, T.M. Foroud³, L.-S. Wang¹, Alzheimer's Disease Sequencing Project Data Flow Work Group. 1) Penn Neurodegeneration Genomics Center, Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA; 2) School of Public Health, Boston University, Boston, MA; 3) Department of Medical and Molecular Genetics, Indiana University, Indianapolis, Indiana; 4) School of Medicine, Boston University, Boston, MA; 5) Broad Institute of Harvard and MIT, Cambridge, MA; 6) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 7) The Genome Institute, Washington Univ., St. Louis, MO; 8) Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA; 9) Taub Institute for Research on Alzheimer's Disease, Columbia University, New York; 10) Department of Neurology, Boston University School of Medicine, Boston, MA; 11) John P Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL; 12) The American Genome Center, Collaborative Health Initiative Research Program, Uniformed Services University, Bethesda MD.

Background. The Alzheimer's Disease Sequencing Project (ADSP) was established in 2012 as a key initiative to meet the goals of the National Alzheimer's Project ACT to prevent and effectively treat Alzheimer's disease (AD) by 2025. Developed jointly by National Institute on Aging (NIA) and the National Human Genome Research Institute (NHGRI), the aims of the ADSP are to: 1) identify new genomic variants contributing to increased risk of developing late-onset AD, 2) identify new genomic variants contributing to protection against developing AD, 3) provide insight as to why individuals with known risk factor variants escape from developing AD, and 4) examine these factors in multi-ethnic populations as applicable in order to identify new pathways for disease prevention. **Methods.** The ADSP Data Flow Work Group (DFWG) and the NIA Genetics of Alzheimer's Disease Data Storage Site (NIAGADS) support data production for the Genome Center for AD (GCAD), sharing and management for the ADSP, and facilitation of data access to the community. Samples are contributed by the AD Genetics Consortium (ADGC) and Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium (CHARGE). Sample plating and shipping is coordinated by the National Cell Repository for AD (NCRAD). Whole-exome and whole-genome sequencing data were generated by four sequencing centers, mapped to GRCh38 and processed by GCAD, and quality controlled by the ADSP. **Results.** The ADSP has completed whole-exome sequencing of 10,939 unrelated cases and controls, whole-genome sequencing of 854 members from 159 families, and whole-genome sequencing of 3094 unrelated cases and controls. These genomes along with 809 ADNI genomes will be available for qualified access in summer 2018 through the NIAGADS Data Sharing Service (DSS). Approved investigators can log in using eRA Commons authentication and download raw sequencing files through Amazon S3 and quality controlled VCFs and phenotype files directly through the Portal. As of June 2018, the Follow-up Study (FUS) phase has generated an additional ~2000 genomes and has ~14,000 ethnically diverse genomes planned for the next four years. **Conclusion.** The DFWG provides support to all ADSP work groups for data related issues, coordinates with NIAGADS, GCAD, and sequencing centers for data transfers, and reviews, posts, and notifies members of new results generated by other work groups. Additional information can be found on the ADSP website (www.niagads.org/adsp).

2345F

Use of a genetically-based biomarker risk algorithm to design Alzheimer's disease clinical trials with outcomes based on neurocognitive measures.

M.W. Lutz^{1,2}, D.K. Burns³, C.L. Dunn⁴, J.S. Cranston⁴, D.J. Hatch^{2,5}, A.M. Saunders³, G. Potter^{2,5}, J.R. Burke^{1,2}, B.L. Plassman^{1,2,5}, K.A. Welsh-Bohmer^{1,2,5}. 1) Department of Neurology, Duke University School of Medicine, Durham, NC; 2) Duke Bryan Alzheimer's Disease Research Center, Duke University School of Medicine, Durham, NC 27705; 3) Zinfandel Pharmaceuticals, Chapel Hill, NC 27517; 4) Duke University, Durham, NC 27708; 5) Department of Psychiatry, Duke University School of Medicine, Durham NC 27710.

The efficiency of clinical trials focused on the prevention of Alzheimer's disease (AD) is greatly increased by enrichment strategies that identify individuals likely to convert from normal cognition to AD or mild cognitive impairment (MCI) due to AD. A genetic biomarker risk algorithm (GBRA) has been developed to identify individuals at high risk of converting from cognitively normal status to MCI due to AD in a 5-year period. The GBRA is based on age and two genetic variants that are easily genotyped in blood: *APOE* and *TOMM40* '523. The GBRA is designed to allow study enrichment for cognitively normal individuals, between the ages of 65-83 years. The present report evaluates the GBRA's performance in predicting cognitive change in a community based AD prevention screening study. Of note, this length of follow-up is comparable to the length of AD prevention trials, 3-5 years. This trial duration is necessary for adequate power to detect a difference in effect between placebo and an intervention, e.g. the TOMMORROW trial (ClinicalTrials.gov identifier, NCT 01931566). GBRA performance was evaluated using data from two studies: the Duke Prevention Screening Study and Database/Repository (PREPARE) and The Measurement to Understand Reclassification of Disease of Cabarrus/Kannapolis (Murdock) study. Subjects were stratified into high or low risk groups at baseline. To account for differences in cognitive change based on age, sex and education, Montreal Cognitive Assessment (MoCA) and Trail making test (TrailsB) scores were standardized for these covariates using National Alzheimer's Coordinating Center UDS3 norms. For PREPARE/Murdock, individuals who had baseline and follow-up neurocognitive evaluations were included, (N=1255, mean age 67.7 years, sd 7.9, mean follow up time 3.7 years, sd 1.2). At baseline, a total of 949 individuals were classified as low risk and 306 as high risk. A statistically significant effect was observed for the GBRA association with change in MoCA scores (p=0.03) and with change in TrailsB scores (p=0.02). For both cognitive measures, direction of change was consistent with the risk category. When used in the context of a clinical trial using neurocognitive measures to assess MCI due to AD, the GBRA has strong potential to significantly reduce the numbers of study subjects required to accumulate sufficient conversion events.

2346W

Brain expression of vascular endothelial growth factor genes relates to cognitive performance. E.R. Mahoney¹, L. Dumitrescu¹, A.L. Jefferson¹, D.A. Bennett², J.A. Schneider², T.J. Hohman¹. 1) Vanderbilt Memory and Alzheimer's Center, Department of Neurology, Vanderbilt University Medical Center, Nashville, TN; 2) Rush Alzheimer's Disease Center, Department of Neurological Sciences, Rush University Medical Center, Chicago, IL.

Recent evidence suggests vascular endothelial growth factor (VEGF), an angiogenic factor, protects against neurodegeneration and cognitive impairment. However, the VEGF family includes multiple ligand and receptor genes that differentially influence angiogenesis and neurodegeneration. This study aims to identify which VEGF ligand and receptor genes are associated with cognitive performance. Autopsy data from 531 participants (normal cognition, n=180; mild cognitive impairment, n=148; Alzheimer's disease [AD], n=203) were acquired from the Religious Orders Study and Rush Memory and Aging Project. VEGF expression levels were obtained from the prefrontal cortex of postmortem brain tissue and quantified using RNAseq. RNA was isolated, reverse transcribed, and normalized as part of the Accelerating Medicine Partnerships AD project. A global cognitive composite score was calculated across 19 neuropsychological tests spanning multiple cognitive domains. Associations between the expression of 65 VEGF isoforms and longitudinal cognition were evaluated using mixed effects regression models co-varying for age and sex. Four genes, including six isoforms, were associated with longitudinal cognitive decline: *FLT4* (MIM 136352; FLT4-207 β =-0.163, $p<0.001$; FLT4-201 β =-0.053, $p=0.002$), *VEGFB* (MIM 601398; VEGFB-202 β =-0.004, $p<0.001$; VEGFB-201 β =-0.001, $p=0.02$), *NRP1* (MIM 602069; NRP1-204 β =0.148, $p=0.02$) and *FLT1* (MIM 165070; FLT1-201 β =-0.005, $p=0.03$). All but VEGFB-201 were differentially expressed in AD cases relative to controls ($p<0.01$). However, co-varying for diagnostic status and expression of cell-specific markers did not alter associations. When restricting to cognitively normal individuals, higher levels of VEGFA-209 (MIM 192240; β =0.11, $p<0.001$) were associated with better cognitive performance, consistent with previous proteomic findings. In summary, high levels of *VEGFB*, *FLT4*, and *FLT1* are associated with worse cognitive performance, perhaps due to their known upregulation in response to neural injury. In contrast, higher levels of *VEGFA* and *NRP1*, which act through kinase domain receptor (*KDR*), are associated with better cognitive performance, perhaps due to the angiogenic and neuroprotective properties of this pathway. These findings suggest that brain expression of VEGF ligand and receptor genes relates to cognitive performance and may have relevance in neurodegeneration. Future work will evaluate VEGF ligand and receptor genes in relation to neuropathology.

2347T

CD33 and Alzheimer's disease: Functional and association analysis of a frame shift variant. A. Mitchell¹, C. Zhang¹, S. McGrath¹, H. Crehan¹, T. Faust¹, D. Felsky², P. de Jager², M. Matijevic¹, J. Arai¹, M. Kinnunen³, A. Palotie³, J. Hutz¹, Industry Partnership for Human Genetics (IPHG). 1) Eisai, Andover, MA; 2) Columbia University, New York, NY; 3) FIMM - Institute for Molecular Medicine, Helsinki, Finland.

CD33 (siglec-3) is a transmembrane receptor expressed on the surface of microglia and peripheral monocytes. The minor allele of a non-coding variant near CD33 (rs3865444) has been associated with reduced risk of Alzheimer's disease (AD). This association appears to be driven by increased skipping of CD33 exon two, which eliminates the protein's extracellular sialic acid binding domain and is thought to confer loss of function. Thus, reduced CD33 expression or function may be protective for AD. We sought to functionally characterize this and other CD33 variants in order to glean more insight into the role of CD33 in AD and to further elucidate the potential protective mechanism. We performed a pQTL study using plasma and CSF from 91 individuals, testing for association between each genotyped or imputed variant within 1 MB of the CD33 gene and soluble CD33 (sCD33) protein level in each fluid. We found that the protective allele at rs3865444 was associated with reduced sCD33 in both fluids (β = -0.53, $p = 8.7 \times 10^{-7}$ in plasma). In addition, we identified a 4 bp deletion in CD33 (rs201074739) with non-Finnish European population frequency 0.02 that was associated with sCD33 in both fluids. This association remained highly significant after adjusting for rs3865444 (β = -1.4, $p = 3.1 \times 10^{-11}$). Because frame shift leads to a premature termination codon in an early exon, it is expected that mRNA carrying the deletion allele will be subject to nonsense mediated decay. To investigate this, we isolated peripheral monocytes from a volunteer who is homozygous for the deletion allele and found no surface expression of CD33 on monocytes in that sample. If reduced CD33 expression or function reduces risk of AD, the deletion allele would be expected to confer a protective effect on carriers. To test this, we performed a meta analysis of 8 studies including a total of 25,651 individuals. We found that the deletion allele is associated with a 20% reduction in AD risk (with $0.04 < p < 0.27$, depending on the model used). In ongoing analyses, we are performing functional studies using PBMC-derived monocytes from volunteers with all three genotypes, as well as in cell lines lacking CD33 or expressing CD33 with reference sequence, with exon two deleted, or the carrying rs201074739 deletion allele. We will present results from a range of functional studies and from our meta analysis of rs201074739 and AD.

2348F

Transcriptomic analysis of mouse models based on candidate genes associated with Alzheimer's disease. R.S. Pandey¹, L. Graham², A. Uyar¹, C. Preuss², G.R. Howell¹, G.W. Carter². 1) The Jackson Laboratory For Genomic Medicine, Farmington, CT; 2) The Jackson Laboratory, Bar Harbor, ME.

New genetic and genomic resources are identifying genetic risk factors for late-onset Alzheimer's disease (LOAD) and characterizing this common dementia at the molecular level. Experimental studies in model organisms can validate these associations and elucidate the links between specific genetic factors and transcriptomic signatures. However, most transgenic animal models are based on rare, early-onset AD genes which may not reflect the full transcriptomic signatures and complete neuropathology of LOAD. Animal models based on LOAD-associated genes are necessary to connect common genetic variation with LOAD transcriptomes, thereby providing novel insights into basic biological mechanisms underlying the disease. We performed RNA-seq on whole brain samples from a panel of six-month-old female mice, each carrying one of the following mutations: homozygous deletions of *APOE* and *CLU*; hemizygous deletions of *BIN1* and *CD2AP*; and a transgenic *APOEε4*. We also included a transgenic *APP/PS1* model for comparison to early-onset variants. Weighted gene co-expression network analysis (WGCNA) was used to identify modules of correlated genes and each module was tested for differential expression by strain. We then compared mouse modules with human postmortem brain modules from the Accelerating Medicine's Partnership for AD (AMP-AD) to determine the AD relevance of risk genes. Mouse modules were significantly enriched in multiple AD-related pathways, including immune response, inflammation, lipid-processing, endocytosis and synaptic-cell-functioning. Various modules were significantly associated with *APOE*, *APOEε4*, *CLU*, and *APP/PS1* mouse models. *APOE*, *APOEε4*, and *APP/PS1* driven modules overlapped with AMP-AD inflammation and microglial modules; *CLU* driven modules overlapped with synaptic modules; and *APP/PS1* modules separately overlapped with lipid-processing and metabolism modules. Furthermore, we found that immune/microglia related genes are up-regulated and synaptic genes are down-regulated in early-onset carriers. This study of LOAD mouse models provides a basis to dissect the role of AD risk genes in relevant AD pathologies. We determined that different genetic perturbations affect different molecular mechanism underlying AD, and mapped specific effects to each risk gene. Our approach provides a platform for further exploration into the causes and progression of AD by assessing animal models at different ages and/or with different combinations of LOAD risk variants.

2349W

Rare coding mutations in Alzheimer disease and other dementias. D.

Patel^{1,2}, J. Mez³, J. Chung^{1,2,7}, X. Zhang^{2,3}, R. Mayeux⁸, J.L. Haines⁹, M.A. Pericak-Vance¹⁰, G. Schellenberg¹¹, K.L. Lunetta¹, L. Farrer^{1,2,3,4,5,6}. 1) Bioinformatics Graduate Program; 2) Departments of Medicine (Biomedical Genetics); 3) Neurology; 4) Biostatistics; 5) Ophthalmology; 6) Epidemiology, Boston University Schools of Medicine and Public Health, Boston, MA; 7) Massachusetts General Hospital, Boston, MA; 8) Columbia University, New York, NY; 9) Case Western Reserve University School of Medicine, Cleveland, OH; 10) University of Miami Miller School of Medicine, Miami, FL; 11) University of Pennsylvania, Philadelphia, PA.

Background: Much of the unexplained heritability of late onset Alzheimer disease (AD) may be due to variants that are too rare to be detected in genome-wide association studies. **Methods:** In a sample of unrelated European ancestry participants in AD Sequencing Project (5617 AD cases, 4594 controls), variants identified by whole exome sequencing were culled from 110 genes previously associated with AD or other dementias, as well as genome-wide, and then filtered for predicted functional impact and occurrence in AD cases, but not controls. **Results:** A total of 24 variants with moderate/high functional impact from 19 genes were observed in 10 or more AD cases. This includes a missense mutation (rs149307620, n=10) in *NOTCH3* for which several other coding mutations are associated with cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL), a diagnostically distinct disorder, marked by severe headaches in young adulthood and stroke and dementia later in life. Evaluation of clinical/autopsy data from 7 of these AD cases revealed classic AD symptoms with progressive memory loss, moderate to severe amyloid and tau pathology at autopsy and no evidence of stroke or severe microvascular disease. Protein modeling showed the AD *NOTCH3* mutation causes quantitative changes in H-bonding that impact the JAG1 binding site, whereas CADASIL *NOTCH3* mutations result in the gain or loss of Cysteine and di-sulfide bonds affecting overall protein structure. Further analysis revealed all 10 cases with rs149307620 share a 5 SNP haplotype with frequency of 15% in ADSP controls. We also identified in 4 AD cases the *TREM2* Q33X mutation that in homozygous form causes Nasu-Hakola disease, a rare disorder characterized by early-onset dementia and multifocal bone cysts, suggesting an intermediate inheritance model for the mutation with respect to dementia outcome. Other findings include an 8-SNP haplotype (0.3% in controls) spanning 12.9 kb that is shared by 10 AD cases and contains 3 rare mutations in *ABCD4* and a 12-SNP haplotype (0.1% in controls) spanning 77.6 kb that is shared by 8 of 10 AD cases and contains 4 rare mutations, two each in *CELSR1* and *GTSE1*. Notably, *ABCD4* encodes an ATP binding cassette transporter that is in the same family as well-established AD gene *ABCA7*. **Conclusions:** We identified several novel rare variants predicted to have high impact on protein structure which may alter AD risk and be potential therapeutic targets.

2350T

Use of brain homogenate RNA expression data to identify novel cell-type specific alterations in Alzheimer's disease. I.S. Piras¹, P.D. Coleman², M.J. Huentelman¹. 1) Neurogenomics Division, Translational Genomics Research Institute, Phoenix, AZ, 85004, US; 2) Biodesign Institute, Neurodegenerative Disease Research Center, Arizona State University, Tempe, AZ, 85287, US.

Cell type resolution gene expression changes are critical to assess in neurological disease due to the heterogeneity within the brain. However, this requires labor and cost intensive approaches (i.e laser capture microdissection, LCM) reducing the sample size. We describe herein a bioinformatics approach that is able to leverage expression profiling data from brain homogenate to derive cell type specific differential expression. Our approach is able to identify expression differences that were previously only seen using LCM, demonstrating the utility of our method to expand the knowledge of brain disease gene expression changes without the need for generation of additional cell specific data. We utilized an LCM-derived RNA expression database from mouse cortex (Zhang et al. J Neurosci 34, 11929-47) to create a cell type enrichment expression score for each gene in the transcriptome. The database includes RNA profiling of 22,458 genes in: astrocytes (A), neurons (N), microglia (M), endothelial (E), and Oligodendrocytes (O). We classified the DEGs from: middle temporal gyrus (MTG), temporal cortex (TCX), BA22, BA10, BA44, hippocampus (HIP), and cerebellum (CBE) in AD and controls. We detected 50 gene classes from 11,994 DEGs. The "Mixed" genes (non specific cell expression), were the most represented (67.7%), followed by N (9.1%), E (5.8%), M (5.7%), A (4.6%) and O (3.1%). "N" genes were mostly downregulated in AD, and "E" genes were mostly upregulated, especially in the temporal area. "M" genes were upregulated in the temporal area and BA44, and slightly downregulated in BA10, HIP and CBE. "N" genes were enriched for the "GABAergic" pathway (MTG, TCX, HIP). Specific pathways for E were: "Angiogenesis" and "Hemostasis" (MTG and TCX). "M" genes were enriched for Immune Pathways (MTG, TCX, BA44, CBE), "O" genes were enriched for Glycosaminoglycan metabolism. We detected classes of cell-specific genes functionally related, observing pathways not identifiable from the homogenate results. For instance, "N" genes were mostly downregulated, and enriched for "GABAergic" pathway. The upregulation of "M" genes and the enrichment in Immune System processes seems specific to temporal area and BA44. Our approach affords the advantage of cell type level interrogation of the transcriptome using bulk RNA profiling which lends itself to more cost and time effective studies that should have increased statistical power.

2351F

Whole genome sequencing-based rare variant analysis in 450 multiplex Alzheimer's disease families. D. Prokopenko¹, L. Bertram², S. Morgan³, S. Amberkar⁴, R. Kirchner⁴, I. Wolkers², S. Hong², B. Hooli¹, K. Mullin¹, J. Miller⁴, B. Chapman⁴, O. Hofmann⁵, C. Lange⁴, W. Hide², R.E. Tanzi¹. 1) Genetics and Aging Research Unit, Department of Neurology, Massachusetts General Hospital, Boston, MA; 2) Lübeck Interdisciplinary Platform for Genome Analytics, Institutes of Neurogenetics and Cardiogenetics, University of Lübeck, Lübeck, Germany; 3) Department of Neuroscience Sheffield Institute of Translational Neuroscience University of Sheffield, Sheffield, UK; 4) Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA; 5) Melbourne Centre for Cancer Research, The University of Melbourne, Melbourne, Australia.

Alzheimer's disease (AD) is one of the most common neurodegenerative disease in the world, being the 6th leading cause of death in the United States. It provides a huge burden on the healthcare system, and this burden is projected to grow in the following decades. Genome-wide association studies (GWAS) have been successful at identifying common and rare genomic loci associated with AD. Whole-genome sequencing (WGS) techniques allow a comprehensive analysis of the entire genome, specifically of rare variants, which cannot be well imputed. We sequenced 1,971 individuals from 450 multiplex AD families from NIMH and NIA cohorts as our discovery dataset. We further gained access to population-based AD sequenced ADNI and ADSP cohorts and used them as our replication dataset. We performed single variant analysis using the FBAT approach. We further performed a region-based analysis of rare variants, which were partitioned into nonoverlapping windows with a spatial clustering approach. In this analysis we specifically focused on variants with low minor allele frequency and their association to AD. Although none of our rare variants were genome-wide significant, we highlight 2 novel potential AD loci from single rare variant analysis and 5 novel loci from region-based analysis showing evidence of replication. Using pathway analysis we show that a subset of identified loci are involved in regulatory interactions with *APP* and *ADAM10*. Whole Genome Sequencing will serve as an important resource for identifying rare functional variants, although large sample sizes in family-based cohorts will be needed to finemap known signals.

2352W

Genome-wide linkage analysis of Caribbean Hispanic Puerto Rican families supports evidence of linkage to chromosome 9. *F. Rajabli¹, B.E. Feliciano-Astacio², K. Celis¹, K.L. Hamilton-Nelson¹, L.D. Adams¹, A.R. Betancourt², V. Rodriguez¹, H. Acosta³, A. Chinea², G.S. Bird⁴, J.M. Vance^{1,5}, M.L. Cuccaro^{1,5}, M.A. Pericak-Vance^{1,5}, G.W. Beecham^{1,5}* 1) John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL, USA; 2) Universidad Central del Caribe, Bayamón, PR, USA; 3) Clínica de la Memoria, San Juan, PR, USA; 4) Center for Outreach in Alzheimer's, Aging and Community Health at North Carolina A&T State University, Greensboro, NC, USA; 5) 8.Dr. John T. MacDonald Foundation Department of Human Genetics, Miller School of Medicine, University of Miami, Miami, FL, USA.

Background. The majority of genomic studies of late-onset Alzheimer's disease (LOAD) focus on non-Hispanic White (NHW) population. Genetic studies in Caribbean Hispanic (CHI) populations are limited, despite evidence of higher incidence rates of AD (compared to NHW) among these ethnic groups. Moreover, the ancestral heterogeneity (admixture) of CHI makes studies of CHI population critical in the discovery of ancestry-specific genetic factors in AD. In this study, we analyzed twenty-five multiplex CHI Puerto Rican (PR) families using genome-wide linkage analysis with the goal of identifying genomic regions segregating with a disease that are likely to contain risk genes for PR AD. **Methods.** Genome-wide array genotyping (with IlluminaMEGA and GSA arrays) of 95 individuals (69 affected) from 25 families was used for linkage analysis. Models assuming autosomal dominant inheritance included: 1) parametric two-point affecteds-only and age-dependent penetrance models, and 2) parametric multipoint affecteds-only and age-dependent penetrance models, and 3) nonparametric multipoint model. Linkage disequilibrium pruning for the multipoint models was done using the 1000 Genomes Project PUR panel frequencies as a reference. Linkage analyses were implemented in the MERLIN software. **Results.** We identified linkage region with the LOD scores exceeding 2.6. A locus at 9p21 produced a linkage LOD score of 2.6 across all families in the nonparametric multipoint model. The highest linkage peak was observed at SNP marker rs7853053. The region for this locus is 17 centimorgans (cM) wide (46.8-63.7 cM). The region overlaps a previously reported linkage peak in NHW families (Zuchner et al. 2008). Genes located in this region, such as CDKN2A, C9orf72 were found as AD associated genes in other studies. CDKN2A involved in the process of neurodegeneration in AD. On the other hand, regions previously identified in AD linkage scans (PSEN1, PSEN2, APP, MAPT, APOE) showed no evidence of linkage in these analyses. **Conclusions.** Linkage analysis of CHI PR families extended evidence for a LOAD susceptibility gene on chromosome 9 and confirmed previously reported linkage to 9p21.3 in NHW families. Further, we will evaluate CDKN2A, C9orf72 and other potential candidate genes using ancestry-aware mapping methods. .

2353T

RNA expression and DNA methylation in the Alzheimer's Disease

Neuroimaging Initiative (ADNI) data. *B. Riley-Gillis, S. Grosskurth, K. Idler, A. Vasanthakumar, J. Davis, J. Waring.* Genomics Research Center, AbbVie, North Chicago, IL.

The aim of this study was to integrate peripheral blood RNA expression & DNA methylation data within the well-characterized Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset to understand the relationship between the two data types & to identify potential blood biomarkers of Alzheimer's disease (AD). The ADNI dataset is comprised of AD patients as well as mild cognitively impaired (MCI) & cognitively normal (CN) patients tracked longitudinally with data collected for WGS, expression analyses, imaging, CSF biomarkers & clinical data such as diagnostic & cognitive assessments (eg. MMSE). RNA expression data generated on the Affymetrix U219 array was obtained from the ADNI data portal where it had been normalized & quality controlled prior to upload. Differential expression analysis was performed for the comparisons AD vs CN, AD vs MCI & MCI vs CN employing a linear model incorporating disease status, gender, age, education, race & APOE e4 state as covariates. Overall, we identified nominal differential expression with adjusted p-values not meeting significance (<0.1). However, at a transcriptome-wide level we do see that differential gene expression changes for AD vs CN & AD vs MCI are more similar than MCI vs CN, indicating that broad gene expression changes may be occurring in the AD patients relative to MCI & CN. To investigate the relationship of RNA expression with DNA methylation, we utilized data generated on the Illumina EPIC methylation array. Differential DNA methylation yielded 483, 139 & 299 differentially methylated positions (DMPs) for the AD vs CN, AD vs MCI & MCI vs CN comparisons, respectively. The DMPs included 357, 113 & 225 individual genes when annotated based on a threshold of 50kb from the CpG. The top DMPs were overlapped with the differentially expressed genes (raw p-value < 0.1) & we focused on genes where the RNA expression was up-regulated & the DNA methylation was down-regulated, & vice-versa. Preliminary exploration of the overlapping genes identified biological connections to brain function & development; ongoing efforts to fully understand the application to AD include association of identified differential gene expression & methylation changes with CSF biomarkers & cognitive scores. **Disclosures:** All authors are employees of AbbVie. The design, study conduct, & financial support for this research were provided by AbbVie. AbbVie participated in the interpretation of data, review, & approval of the publication.

2354F

Towards personalized health: Multimodal risk prediction models for dementia. *N. Shah¹, C. Swisher¹, N. Schenker-Ahmed¹, P. Garst¹, I. Shomorony¹, D. Tkach¹, C. Venter², D. Karow¹.* 1) Human Longevity, Inc., San Diego, CA; 2) J. Craig Venter Institute, La Jolla, CA.

Current models for predicting onset of dementia rely primarily on single modality data (e.g. magnetic resonance imaging, genotyping, or cognitive assessment). However, it is necessary to include multiple modalities including imaging, genetic and clinical biomarkers to accurately assess an individual's lifetime and short-term risk. Furthermore, it requires genotyping beyond a single gene, analyzing imaging features beyond a single structure, and further investigation into mitigating risk factors. We describe a method that utilizes polygenic risk score for Alzheimer's disease, and 77 features from structural T1-weighted magnetic resonance imaging (MRI) including surface area, thickness, and volume of the brain regions to train multimodal models to predict the likelihood of cognitive decline as well as lifetime risk of dementia. The cumulative hazard from the lifetime risk model is then integrated with population-based incidence rate for dementia to calculate annualized lifetime risk personalized to the age and sex of an individual. For both models, respectively, using MRI and WGS data combined improves performance (1: AUC=0.95, 2: AUC=0.92 within 3 years) in the prediction of dementia progression over either MRI (1: AUC=0.93 at the time of event and AUC=0.55 10 years prior to developing dementia, 2: AUC=0.90 within 3 years) or WGS (1: AUC=0.82, 2: AUC=0.75 within 3 years) alone. We validated the models using an independent cohort. In addition we show that our model outperforms commonly used markers to predict onset of dementia including amyloid plaque quantitative markers from PET scan in detection of dementia. Finally, we show how the results of these multimodal models would be visualized to help an individual achieve a better understanding of their risk.

2355W

Investigation for a protective variant for the *ApoEε4* allele in Alzheimer disease. *J.M Vance¹, F. Rajabli¹, B.E. Feliciano-Astacio², K. Celis¹, L.D. Adams¹, K.L. Hamilton-Nelson¹, A.R. Betancourt³, H. Acosta³, P. Bussies¹, T. Starks⁴, N. Hoffman¹, C. Reitz¹, R. Mayeux¹, G.S. Bird⁵, M.L. Cuccaro¹, J.L. Haines⁶, G.W. Beecham¹, M.A. Pericak-Vance¹.* 1) John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL, USA; 2) Universidad Central del Caribe, Caguas, PR, USA; 3) Clinica de la Memoria, San Juan, PR, USA; 4) Gertrude H. Sergievsky Center, Departments of Neurology, Psychiatry, and Epidemiology, College of Physicians and Surgeons, Columbia University, New York, NY, USA; 5) North Carolina A & T State University, Greensboro, NC, USA; 6) Department of Epidemiology and Biostatistics, Institute for Computational Biology, Case Western Reserve University School of Medicine, Cleveland, Ohio, USA.

Background: The risk for late-onset Alzheimer disease (LOAD) in *ApoEε4* carriers differs between ethnic groups, with Japanese>Europeans>African-Americans (AA) >Africans. We hypothesized two possibilities for this difference: 1) differences in overall (global) ancestry (GA) that also reflects environmental pressures 2) differences due to local ancestry (LA) around the *ApoE* region. To investigate, we assessed GA and LA in our LOAD datasets in AA and Caribbean Hispanics/Latinos (CHI) for association and interaction with *ApoEε4*. **Methods:** Genotyping data were available on 2,229 AA (811 cases, 1,418 controls) and 267 Puerto Rican (PR) CHI samples (169 cases, 98 controls). LA was calculated using SHAPEIT and RFMix. Analyses included an interaction analysis between *ApoEε4* and LA or GA using logistic regression and a haplotype-based analysis of LA using GEE. For sequence comparison, we used *ApoEε4* carriers in African (Yoruban), European (CEU) and Japanese genomic DNA sequence data (1000 Genomes). Differences were screened for potential functional effect (insilico: GTEx, ROADMAP) and invitro (massively parallel reporter assays (MPRA)). iPSC of each LA subtype are differentiated into neurons, astrocytes and microglia. **Results:** We identified a 2mB region of LA surrounding the *ApoEε4* allele. Both haplotype and interaction analyses were significant for LA, but were not significant for GA. Interaction results: *ApoEε4* x LA (AA<0.008, PR<0.04) and haplotype analysis showed that the CEU *ApoEε4* haplotype was significantly different from the African haplotype (PR<1.3e-08, AA<2.2e-16) for AD risk. We identified 548 variants with significant frequency (Bonferroni corrected) differences between CEU and African populations in the LA region. A strong peak of sequence differences (p<e-12) lies within the strongest topologically associated domain of *ApoE*. MPRA clones have been created across the LA region and are being screened for regulatory potential. We have identified over 100 AA *ApoEε4* brains for analysis as well. iPSC have been created for *ApoEε4/4* homozygotes of each of the LA subtypes. **Conclusion:** LA surrounding the *ApoEε4* allele is a primary contributor for population differences in LOAD risk. Identification of this variant could lead to therapeutic targets and provide greater insight into the mechanism in which *ApoEε4* leads to AD. Counseling of risk for *ApoEε4* carriers will differ depending on the origin of the patients *ApoEε4* allele.

2356T

Integrative functional genomics analysis of neuropathologic features of Alzheimer's disease and related cognition decline reveals novel insights on previously identified loci. J. Yang¹, S. Nagpal², M. Epstein¹, L. Tsoi³, M. Patrick⁴, A. Wingo⁵, T. Wingo⁵. 1) Department of Human Genetics, Emory University School of Medicine, Atlanta, GA; 2) Department of Bioinformatics, Georgia Institute of Technology, Atlanta, GA; 3) Department of Dermatology, University of Michigan School of Medicine, Ann Arbor, MI; 4) Department of Psychiatry, Emory University School of Medicine, Atlanta, GA; 5) Department of Neurology, Emory University School of Medicine, Atlanta, GA.

Alzheimer's Disease (AD) is a devastating neurodegenerative disease that has ~70% of risk believed to be genetic with polygenic architecture. Previous GWAS have identified 20 genetic risk genes of AD. However, most identified SNPs fall within the noncoding regions of the genome where the closest genes were reported as the genetic risk genes. Consequently, the molecular mechanisms underlying these identified risk genes are largely unknown. To provide further insights into the biology of these known loci, we conducted a novel integrative functional genomics analysis using the genotype and transcriptomic data from the Religious Orders Study and Memory Aging Project (ROSMAP) study, with 499 samples from the dorsolateral prefrontal cortex (dPFC) tissues. We first used the recently published Dirichlet Process Regression (DPR) method to estimate the genetically-regulated component of the gene expression respectively for all 20 loci. We took the assayed gene expression levels (adjusted for other non-genetic covariates) as the response variable and all SNPs within a 10⁶ base pair window of the gene boundary as predictors. DPR nonparametrically models the genetic architecture by assuming a Dirichlet process prior for the effect-size variance, which significantly improved the regression over the alternative PrediXcan method (average R²=16.7% vs. 4.4%). Then we respectively tested the association between the estimated genetically-regulated components and the AD-related phenotypes — AD diagnoses, cognition decline, and two neuropathological features (CERAD, Braak). Interestingly, we identified significant associations for *SORL1* with respect to all four studied phenotypes, demonstrating that *SORL1* is likely to affect cognition decline (p=0.07) and AD (p=0.043) through CERAD (p=0.048) and Braak (p=0.014). Particularly, *CEFL1* was identified significant for both CERAD (p=0.004) and AD (p=0.032), showing that *CEFL1* is likely to affect AD through CERAD. In addition, *UNC5C* was identified significant for both cognition decline (p=0.0053) and AD (p=0.023), showing that *UNC5C* is likely to affect AD through the cognition decline. In summary, our study provides novel transcriptomic insights for the known loci of AD. Moreover, our fitted DPR models can be used to impute the genetically-regulated transcriptomic components (in dPFC tissues) for samples with genotype data only, which will significantly increase the study sample size and help illustrate the genetic etiology of AD.

2357F

Genome-wide interaction study of brain and CSF measures of Alzheimer disease on memory impairment. C. Zhu¹, J. Chung¹, J. Mez², K. Lunetta³, G. Jun^{1,4}, L. Farrer^{1,3}, X. Zhang^{1,3}. 1) Section of Biomedical Genetics, School of Medicine, Boston University, Boston, MA, USA; 2) Departments of Neurology, Boston University School of Medicine, Boston, MA, USA; 3) Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA; 4) Neurogenetics and Integrated Genomics, Andover Innovative Medicines Institute, Eisai Inc, Andover, MA, USA.

Background: Although amyloid- β plaques (A β) and neurofibrillary tangles (tau) are pathological hallmarks of Alzheimer disease (AD), ~10% of clinically diagnosed AD cases do not have AD pathology. Furthermore, a certain percentage of subjects with AD-like pathology do not have clinical AD symptoms (e.g., memory loss). Genetic factors that affect brain structure and cognition in cognitively normal adults may also influence AD pathogenesis later in life. In this study, we sought to identify genetic modifiers of the influence of AD-related endophenotypes on memory impairment. **Methods:** We evaluated the interaction of logical memory test scores including immediate (LMI) and delayed (LMD) recall, with single nucleotide polymorphisms (SNPs) across the genome on AD-related brain imaging and cerebrospinal fluid (CSF) endophenotypes including hippocampal volume (HPV), amyloid- β_{42} (A β_{42}) and total/phosphorylated tau (t-/p-Tau) in cognitively normal (CN, n=315), mild cognitive impairment (MCI, n=611), and AD subjects (n=196) from the Alzheimer's Disease Neuroimaging Initiative (ADNI) study. LMI and LMD were considered as cognitive outcomes, and the interaction between SNP and endophenotypes were estimated after controlling for age, sex, and principal components of ancestry. HRC-imputed ADNI 1 (N=691) and ADNI GO/2 (N=431) data were analyzed separately. Results were combined across ADNI 1 and ADNI GO/2 using meta-analysis. $P < 5.0 \times 10^{-8}$ was used to determine genome-wide statistical significance for association. **Results:** In CN subjects, we identified significant interactions of *ATP8A2* SNP rs12385841 with A β_{42} on LMI score ($P = 1.08 \times 10^{-8}$) and rs77369553 located near *ZNF804A* with ptau on LMD score ($P = 4.66 \times 10^{-6}$). No significant interactions were observed in MCI subjects. In AD subjects, significant interactions (P is $1.3 \times 10^{-8} \sim 1.4 \times 10^{-9}$) were observed for ptau with SNPs in *PCSK5*, *UNC79*, *CDYL*, and *ACSS3* on LMD score, and for tau/A β_{42} with SNPs in *DZIP1* ($P = 1.3 \times 10^{-10}$) and *GPC5* ($P = 4.4 \times 10^{-8}$) on LMD score. In addition, we found consistent high expression patterns of *CDYL*, *DZIP1*, *GPC5*, and *UNC79* in human normal brains in GTEx. **Conclusions:** Our findings provide insight into the complex relationships between the amyloid and tau burden and memory impairment in stages preceding and after onset of AD symptoms. Further functional studies are required to elucidate the role of identified genetic variants in AD pathophysiology. A replication study using sequencing data is underway.

2358W

Association of genetic variants/predicted gene regulatory scores with imaging-derived phenotypes in multiple sclerosis patients. K. Kim¹, L. Madireddy¹, C. Cordano¹, R.G. Henry¹, A.J. Green¹, B.A. Cree¹, S.L. Hauser¹, J.R. Oksenberg¹, S.E. Baranzini^{1,2}. 1) Department of Neurology, University of California San Francisco, San Francisco, CA; 2) Institute for Human Genetics, University of California San Francisco, San Francisco, CA.

Genome-wide association studies have identified 200 loci in addition to those in the MHC that confer susceptibility to multiple sclerosis (MS). However, a significant proportion of the heritability of MS remains to be explained. Relevant neuroimaging-derived phenotypes greatly empower the genetic analysis of complex diseases as they provide quantitative and reproducible indicators of disease activity, effectively linking the genotype and clinical phenotype. In MS, brain, spinal cord, and retinal imaging represent valuable quantitative assessments of CNS integrity and enable tracking of disease progression. In this study, we collected longitudinal neuroimaging-derived metrics from magnetic resonance imaging (n=466) and optical coherence tomography (n=363) in UCSF MS-EPIC cohort. To calculate a patient-specific rate of change of longitudinal metrics, we performed a linear mixed effect analysis using random slope model. Then, we conducted association studies in a different level of biological hierarchy using MS-susceptible genetic variants and predicted regulatory effects for 10 different cell types computed from the publicly available gene regulatory dataset. We found 2 variants with a significant effect on either phenotype. Rs2364482 (*LOC105369628*) was associated with gray matter volume change, and rs1250568 (*ZMIZ1*) with macular thickness (FDR < 0.05). We also tested an association of genes in 10 different cell types with imaging-derived metrics using predicted regulatory effects as a gene level analysis. A higher regulatory score of *EMR1* was associated with reduced brain atrophy change in CNS cell, on the other hand, a higher score of *EYA2* was associated with increased atrophy change ($p < 0.001$). *IKZF1* and *CAMTA1* were significant in T cell for brain atrophy ($p < 0.001$). Interestingly, *CAMTA1* has been found significant association with amyotrophic lateral sclerosis by GWAS but has not been reported in MS. For retinal thickness change, *MERTK* was significant for inner plexiform layer thinning in T or Treg cells. It has been reported that higher expression of *MERTK* was associated with risk of developing MS. We established an analytical pipeline to investigate the influence of genomic variation on neuroimaging-derived metrics. Then, we found candidate genetic variants and genes associated with phenotypic change using the pipeline. Our findings will provide more accurate disease pathogenesis and biomarkers for different disease progression in MS patients.

2359T

A multiple sclerosis CD4+ T cell methylation QTL reference map identifies the proximal functional consequences of 18 MS susceptibility variants. T. Roostaei¹, H.U. Klein¹, N. Patsopoulos², H.L. Weiner², P.L. De Jager¹. 1) Center for Translational and Computational Neuroimmunology, Department of Neurology, Columbia University Medical Center, New York, NY; 2) Department of Neurology, Brigham & Women's Hospital and Harvard Medical School, Boston, MA.

Multiple sclerosis (MS) is a complex autoimmune disease of the central nervous system. To date, 140 non-MHC autosomal genetic loci are associated with MS risk. However, the functional consequences of the majority of these loci are not known. CD4+ T cells are strongly implicated in MS. Here, we sought to determine the effects of MS loci on DNA methylation of CD4+ T cells in MS patients. 180 relapsing-remitting MS patients were recruited. Imputation was performed on the whole-genome genotyping data. CD4+ T cells were isolated from frozen peripheral blood mononuclear cell samples. DNA methylation was measured using Illumina 850K array. After quality control, data from 156 individuals with European ancestry were used for analysis. Linear regression was performed to assess the association between genotypes and DNA methylation in cis ($\pm 1\text{Mb}$) using M values, accounting for the effects of covariates including age and sex. From 769,634 tested CpG sites, we found evidence for the influence of cis genetic effects on the methylation of 107,925 CpGs (FDR-adjusted $p < 0.05$). From the identified CpGs common between the Illumina 850K and 450K arrays (n=52,687), 64.8% were previously identified as having cis-methylation quantitative trait loci (mQTLs) in data from naïve CD4+ T cells of healthy participants (FDR-adjusted $p < 0.05$, BLUEPRINT epigenome project, performed using 450K array), therefore validating our findings. The top variants for 762 mQTLs were in LD ≥ 0.2 with 119 lead variants of MS-associated loci. Colocalization analysis was performed to find loci whose effects on MS risk and cis-DNA methylation are likely driven by a shared causal variant. 3 mQTLs showed strong evidence of colocalization (posterior probability > 0.95) with 3 MS loci, affecting methylation levels of 12 CpGs in the body of *TNFSF14*, *ANKRD55* and *IL6ST* and the transcription start site of *CYP24A1* genes. 19 additional mQTL effects were colocalized with 15 other MS loci with posterior probability > 0.8 . Of note, one of these mQTL effects was an association between rs4939490 and the methylation level of cg14130459 in the 1st exon of *SLC15A3*, suggesting a novel MS susceptibility gene that is influenced by this variant. Our new MS CD4+ T cell reference mQTL map therefore identifies the most proximal functional consequence of a large proportion of MS variants, refining the list of such loci with effects in healthy subjects (BLUEPRINT) and introducing novel loci for investigation in future functional studies.

2360F

A genetic modifier model for SUDEP in SCN1A driven epilepsy. P.J. Van-deventer, R. Domínguez Vidafía, A.R. Ghazi, E.S. Chen, A.M. Goldman, C.A. Shaw. Baylor College of Medicine, Houston, TX.

Sudden Unexpected Death in Epilepsy Patients (SUDEP) is a rare disorder that affects some individuals with SCN1A driven epilepsy as well as epilepsy patients with different genetic drivers. We have pursued a genetic modifier model of SUDEP in SCN1A cases. Under this model, concomitant variation – or genetic load – outside of SCN1A but in genes related to the SUDEP pathophysiology are expected to be elevated for individuals at high risk for SUDEP, while individuals that lack such variants are at lower risk. To pursue this hypothesis we have created a data set and computational approach to evaluate the modifier model. We identified individuals who underwent exome sequencing that have pathologic or potentially pathologic SCN1A variation from the Baylor Genetics laboratory who were also patients at Texas Childrens hospital. We also utilize exome data from SUDEP individuals with SCN1A driven epilepsy. Taken together, these data provide a set of positive control SCN1A SUDEP cases as well as a collection of test individuals that we can evaluate genomically and by chart review. The computational procedure considers the pathology and rate of occurrence of variation in genes related to the pathophysiology of SUDEP. The gene set used for load analysis was identified by expert review of the literature and publications of genetic models of SUDEP. The pathways and processes implicated include cardiac arrhythmia and spreading depression, as well as other phenotypes. Our preliminary results indicate significant evidence for increased load of Variants of Uncertain Significance (VUS) and potentially pathologic alleles in the candidate gene set among the SUDEP cases. The promising results of this study suggest a broader role for the modifier model in understanding the severity and variation in presentation observed in Mendelian disease. The results suggest that the spectrum between Mendelian and Complex disease could be viewed as a continuum, where the totality of alleles should be considered.

2361W

Genetic architecture of the human anterior commissure. H.H.H Adams¹, K. Wittfeld², N. Jahanshad³, G. Roshchupkin⁴, B. Mazoyer⁵, A. Beiser⁶, L. Launer⁷, J. Bis⁸, X. Jian⁹, M. Luciano¹⁰, S. Trompet¹¹, E. Hofer¹², L. Yanek¹³, P. Kolovos³, A. Mahfouz¹⁴, V. Van Deerlin¹⁵, R.A. Poot⁶, P. Nyquist¹³, R. Schmidt¹², J.W. Jukema¹¹, J. Wardlaw¹⁰, M. Fornage⁹, W.T. Longstreth Jr⁶, V. Gudnason⁷, S. Seshadri⁶, S. Debette⁵, W. Niessen¹, P. Thompson⁴, H. Grabe³, M. Ikram¹, CHARGE consortium. 1) Epidemiology, Erasmus MC, Rotterdam, Zuid-Holland, Netherlands; 2) Cell Biology, Erasmus MC, Rotterdam, Zuid-Holland, Netherlands; 3) German Center for Neurodegenerative Diseases (DZNE) Rostock/Greifswald, Greifswald, 17487, Germany; 4) Imaging Genetics Center, USC Mark and Mary Stevens Neuroimaging & Informatics Institute, Keck School of Medicine of University of Southern California, Los Angeles, 90292, USA; 5) UMR5296 University of Bordeaux, CNRS, CEA, Bordeaux, France, 146 rue Leo Saignat, 33076 Bordeaux 2 cedex; 6) Department of Neurology, Boston University School of Medicine, Boston, MA, USA; 7) Intramural Research Program, NIA, NIH, 7201 Wisconsin Ave, Suite 3C-309, Bethesda, MD 20892; 8) Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA USA, 1730 Minor Avenue / Suite 1360 / Seattle, WA 98101; 9) Institute of Molecular Medicine and Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX, USA; 10) Centre for Cognitive Ageing and Cognitive Epidemiology, Psychology, University of Edinburgh, Edinburgh, EH8 9JZ, UK; 11) Department of Cardiology, Leiden University Medical Center, Leiden, the Netherlands; 12) Department of Neurology, Clinical Division of Neurogeriatrics, Medical University Graz, Austria, Auenbruggerplatz 22, 8036 Graz, Austria; 13) GeneSTAR Research Center, Department of Medicine, Johns Hopkins University School of Medicine, 1830 E Monument St Suite 8028, Baltimore, MD, 21287, USA; 14) Delft Bioinformatics Laboratory, Delft University of Technology, 2628 CD Delft, the Netherlands; 15) Center for Neurodegenerative Disease Research, Department of Pathology and Laboratory Medicine, Morris K Udall Parkinson's Disease Center Of Excellence, Institute on Aging, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA.

The anterior commissure is a nerve fiber bundle that interconnects the brain's hemispheres. It is important for higher-order cognitive functions and has a presumed role in neurodegeneration, but its genetic architecture remains elusive. Uncovering the genes underlying variation in the anterior commissure could aid understanding of its development and it potentially provides mechanistic insight into diseases that results from its dysfunction. We performed a two-stage genome-wide association study ($N = 18,828$) of the size of the anterior commissure and identified six independent variants at four loci (p -values ranging from 4.1×10^{-8} to 9.4×10^{-23}). Using *in silico* and *in vitro* approaches, including gene expression analyses and Targeted Chromatin Capture (T2C) in human neural progenitor cells, we mapped the loci to probable causal genes involved in axon guidance (*EPHA3* and *SEMA6A*), cognitive disorders (*CTNND2*), and growth factor signaling (*RIT2*). Also, brain-wide morphometry studies revealed distinct associations of the variants with connected grey matter regions in the brain. Genome-wide analyses revealed an enrichment for H3K4me1 peaks (marking enhancer sites), introns, and conserved sequences, as well as cell-type-specific annotations from the central nervous system and cardiovascular system. Furthermore, we identify pleiotropy between genes known to increase risk of neurodegenerative diseases and loci of the anterior commissure, including frontotemporal lobar degeneration gene *TMEM106B*. These analyses shed light on the genetic architecture of commissural tracts and establish the size of the anterior commissure as a relevant marker of neurodegeneration.

2362T

Multiple sclerosis genetic burden score in a systems biology study of MS patients from four countries. S.D. Bos^{1,2}, I.S. Brorson^{1,2}, E.A. Høgestøl^{1,2}, T. Berge^{2,3}, J. Saez-Rodriguez⁴, A. Uccelli⁵, F. Paul⁶, P. Villoslada⁷, H.F. Harbo^{1,2}. 1) Institute of Clinical Medicine, University of Oslo, Oslo, Norway; 2) Department of Neurology, Oslo University Hospital, Norway; 3) Department of Mechanical, Electronics and Chemical Engineering, Oslo Metropolitan University, Oslo, Norway; 4) Faculty of Medicine and University Hospital, University of Aachen, Aachen, Germany; 5) Ospedale Policlinico San Martino, Genova, Italy; 6) Department of Neurology, Charité – Universitätsmedizin Berlin, Berlin, Germany; 7) IDIBAPS, Neurosciences, Barcelona, Spain.

Introduction: Multiple Sclerosis (MS) is an auto-inflammatory disease which affects components of the nervous system. This results in neurological deficits of varying degree, depending on the affected site. In order to better study the complex interplay between genetics and the environment, which both are well-known risk factors a systems biology study (Sys4MS) was initiated in 2015. This study aims at applying systems medicine approach combining integrative omics, imaging and clinical data with computational tools, to develop personalized healthcare for people with MS. **Materials and Methods:** From four participating centres in total 328 MS patients and 90 healthy controls were included. DNA was extracted from whole blood for all participants and genotypes on the Illumina OmniExpress chip. Genotype data was then imputed, resulting in genotypes for 152 out of 200 MS-associated SNPs. Furthermore, variants at 17 out of 32 MS-associated HLA loci were imputed. The odds ratio was used to weight genotype on each of the 169 MS-associated loci, the sum of the genetic burden (MSGB) per individual was calculated. **Results:** The patients had a mean age of 41 years and 71% were women. The mean disease duration at the time of inclusion was 10 years (standard deviation 8 years). 83% of the patients had relapsing-remitting MS, whereas the remaining patients had primary progressive MS. The median expanded disability status scale (EDSS) score was 2.0 (range 0-8.0). The MSGB of patients was significantly higher as compared to the controls (respectively 4.23 vs 3.20 with a t-test p-value of 3.4×10^{-6}). We did not observe a difference in the MSGB between patients from the participating centres (Norway, Germany, Spain and Italy). **Discussion:** MS patients in this systems biology approach have a higher MSGB as compared to controls, in line with earlier studies. We did not see a significant difference in the MSGB between the participating centres, although we may lack the power to detect subtle differences between the MSGB between countries in our modest sample size. Further investigations of the role of the MSGB in detailed phenotypes collected in the Sys4MS study, such as the response to treatment and immune cell profiles remains to be investigated.

2363F

Neuropathological correlates and genetic architecture of microglial activation in elderly human brain. D. Felsky^{1,2}, T. Roostaei¹, N. Kwangsik³, S.L. Risacher³, E.M. Bradshaw^{1,2}, V. Petyuk⁴, J.A. Schneider^{5,6}, A. Saykin³, D.A. Bennett^{5,6}, P.L. De Jager^{1,2}. 1) Center for Translational and Computational Neuroimmunology, Dept. of Neurology, Columbia University Medical Center, New York, NY; 2) Program in Population and Medical Genetics, Broad Institute of MIT and Harvard, Cambridge, MA; 3) Indiana Alzheimer's Disease Center, Center for Neuroimaging, Department of Radiology and Imaging Sciences, Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, IN; 4) Pacific Northwest National Laboratory, Richland, WA; 5) Department of Neurology, Rush University Medical Center, Chicago, IL; 6) Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL.

Background: Microglia, the resident immune cells of the brain, have important roles in brain health. However, little is known about the regulation and consequences of microglial activation in the aging human brain. **Methods:** We assessed the effect of microglial activation in the aging human brain by calculating the proportion of activated microglia (PAM), based on morphologically defined stages of activation in four regions sampled postmortem from up to 225 elderly individuals. **Results:** We found that cortical and not subcortical PAM measures were strongly associated with β -amyloid, tau-related neuropathology, and rates of cognitive decline. Effect sizes for PAM measures are substantial, comparable to that of APOE $\epsilon 4$, the strongest genetic risk factor for Alzheimer's disease. Mediation modeling suggests that PAM accelerates accumulation of tau pathology leading to cognitive decline, supporting an upstream role for microglial activation in Alzheimer's disease. Genome-wide analyses identified a common variant (rs2997325) influencing cortical PAM that also affected *in vivo* microglial activation measured by positron emission tomography using [¹¹C]-PBR28 in an independent cohort. Finally, we identify overlaps of PAM's genetic architecture with those of Alzheimer's disease, educational attainment, and several other traits based on published GWAS summary data from over 1.7 million people. **Conclusions:** Active cortical microglia may cause worsening cognitive decline via an accumulation of tau pathology. We also identified and validated a novel genetic locus influencing postmortem microglial activation, providing strong biological rationale for further investigation. This locus may be clinically meaningful given its strong effect on *in vivo* TSPO imaging. Finally, genetic predisposition to microglial activation influences several other human traits including educational attainment, raising the intriguing possibility that genetic determinants of late life neuroinflammation may affect early life cognitive development.

2364W

Inositol deficiency induce autophagy impairing via PI3K/ Akt/ mTOR / p70S6K signaling. J. Guo¹, Z. Chen², H. Yue¹, S. Li¹, Y. An³, D. Luo¹, F. Zhang⁴, R. Zhong⁵, J. Lyu⁶, Y. Yang⁶, A. Chen⁶, L. Shi¹, X. Wang¹, F. Wang², Y. Bao¹, J. Zou¹, Z. Guan¹, T. Gao³, Z. Zhu¹, B. Niu¹, J. Huang¹, H. Wang⁶, T. Zhang¹, J. Wang¹, J. Wu⁸. 1) Beijing Municipal Key Laboratory of Child Development and Nutriomics, Capital Institute of Pediatrics, Beijing 100020, China; 2) State Key Laboratory of Genetic Engineering, Collaborative Innovation Center of Genetics and Development, School of Life Sciences, Fudan University, Shanghai, 200438, China; 3) Beijing Key Laboratory of Environmental & Viral Oncology, College of Life Science and Bioengineering, Beijing University of Technology, Beijing 100124, China; 4) Chinese Academy of Inspection & Quarantine, Beijing 100023, China; 5) National Center for Safety Evaluation of Drugs, National Institutes for Food and Drug Control, Beijing 100176, China; 6) Miami University, Department of Chemistry and Biochemistry, 651 E. High St. Oxford, OH 45056; 7) Institute for Nutrition and Health, Chinese Center for Disease Control and Prevention; 8) Department of Biochemistry & Immunology, Capital Institute of Pediatrics, Beijing 100020, China.

Periconceptional maternal myo-inositol (MI) deficiency is one of the crucial risk factors of neural tube defects (NTDs), but the exact mechanism remains poorly understood. In the present study, we sampled 80 NTD pregnancies and 80 controls from a high-prevalence area of Shanxi Province in northern China. We found that maternal MI deficiency was associated with an increased risk of NTDs by gas chromatography–mass spectrometry (GC-MS) analysis ($P < 0.05$). To further investigate the causality between maternal MI deficiency and the occurrence of NTDs, we established a maternal MI deficient mouse model using intraperitoneal injection of Li_2CO_3 , which acts as an inhibitor of inositol 1-phosphatase (IMPase) and a key enzyme for the synthesis and recycling of inositol in vivo. The MI level of the pregnant mouse decreased significantly after Li_2CO_3 injection, and reached to the lowest level at 8 hours after the injection (23.72mg/L vs 44.15 mg/L $P < 0.05$). Repeat trials showed that the average incidence of NTDs was 26.7%. The NTD phenotypes included encephalocele (85.8%), anencephaly (5.7%) and spina bifida (8.5%). Using the phospho-protein array (PEX100 and PAB216, Full Moon Biosystems, Inc.), we intended to find the possible differential phosphorylated proteins between the NTD and control in MI deficiency human and mouse respectively. We performed pathway analysis on the differentially phosphorylated proteins using KEGG databases of human NTDs with MI deficiency. The result displayed that PI3K-Akt signaling pathway (hsa04151) was the most enriching pathway. We then validated the proteins involved in the PI3K-Akt signaling pathway using a mouse phospho-protein array. A total of 21 proteins were found to be differentially phosphorylated (Fold change > 1.5), and 14 of them were differentially phosphorylated in both the human and mouse, including YWHAZ, AKT1, AKT2, BCL2, CCND1, NOS3, PTK2, IRS1, MTOR (mTOR), TP53, RPS6KB1 (p70S6K), PDPK1, PPP2CA and PTEN. Further pathway analysis on the 14 common differentially phosphorylated protein found that PI3K/Akt/mTOR/p70S6K pathway was the most disturbed signaling pathway induced by MI deficiency. Our study indicated that PI3K-Akt signaling pathway might be one of the important regulatory mechanisms of inositol deficiency-induced NTDs.

2365T

Genome-wide association study of febrile seizures identifies new loci implicating synaptic and fever response genes. L. Skotte¹, X. Liu¹, J. Grauholm^{2,3}, J.P.S. Fadista¹, V. Yakimov¹, V. Appadurai^{3,4}, M.S. Hildebrand⁵, T. Pers¹, B. Pasternak¹, A. Buil^{3,4}, F. Geller¹, S. Berkovic⁶, D.M. Hougaard^{2,3}, T. Werge^{3,4}, A. Hviid¹, M. Melbye^{1,6}, B. Feenstra¹. 1) Department of Epidemiology Research, Statens Serum Institute, Copenhagen, Denmark; 2) Danish Centre for Neonatal Screening, Department of Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark; 3) iPSYCH, Lundbeck Foundation Initiative for Integrative Psychiatric Research, Denmark; 4) Mental Health Center Sct. Hans, Mental Health Services, Capital Region Denmark, Roskilde, Denmark; 5) Epilepsy Research Centre, Department of Medicine, University of Melbourne (Austin Health), Victoria, Australia; 6) Department of Medicine, Stanford University School of Medicine, Stanford, California, USA.

Introduction: Febrile seizures affect 2-5% of children in western countries including Denmark with an incidence peak around 16 months of age. While usually benign, in a minority of cases febrile seizures are later followed by epilepsy. **Materials and Methods:** We extended our previous genome-wide association study on febrile seizures to a total of 4304 cases and 48290 controls. The analysis was based on 8.5 million variants imputed from the 1000 genomes project phase III reference panel, using logistic regression on imputed SNP dosages for association testing. Replication testing was carried out in independent Danish and Australian cohorts (N cases, N controls). **Results:** The 4 loci from the previous GWAS were confirmed, and we tried to replicate all 7 new loci with variants associated at $P < 5 \times 10^{-7}$. Six loci turned out to be consistently associated with febrile seizures (replication $P < 0.05$, combined $P < 5 \times 10^{-6}$). Subsequent genome-wide enrichment analysis suggested that pre-synaptic cytomatrix proteins were implicated in seizure etiology. The remaining loci suggested implication of prostaglandin signaling, endogenous interleukins, as well as GABAergic signaling. **Conclusion:** Our study identified novel genetic associations with febrile seizures implicating genes involved in neuronal excitability and fever response. **Grant support:** Lundbeck Foundation, Novo Nordisk Foundation, Oak Foundation and Carlsberg Foundation.

2366F

Cytoskeletal scaffolding protein WHIRLIN loss of function mutation in familial frontotemporal dementia. Y. Baradaran-Heravi^{1,2}, L. Dillen^{1,2}, C. Koçoglu^{1,2}, C. Van Broeckhoven^{1,2}, J. van der Zee^{1,2} on behalf of the BEL-NEU Consortium and the EU EOD Consortium. 1) Neurodegenerative Brain Diseases Group, Center for Molecular Neurology, VIB, Antwerp, Belgium; 2) Laboratory of Neurogenetics, Institute Born-Bunge, University of Antwerp, Antwerp, Belgium.

Background: Unresolved frontotemporal dementia (FTD) and early-onset dementia (EOD) families are systematically collected within the context of the BELNEU and EU EOD Consortia. **Materials and Methods:** Whole-exome sequencing (WES) was performed in a Belgian autosomal dominant dementia family with 5 patients over 3 generations. The index-patient was diagnosed with FTD at the age of 53 years, the other patients from previous generations were diagnosed with Alzheimer's disease or dementia unspecified. DNA for WES was available for 3 patients over 2 generations. Additional DNA samples of 2 unaffected relatives were used for variant co-segregation analysis. Shared, rare (MAF<0.01%), heterozygous variants with predicted protein-modifying effect were prioritized for follow up. Multiplex amplification resequencing was used to validate selected variants and perform candidate gene screening in unrelated FTD patients (n=397) and control individuals (n=838). **Results:** WES of 3 affected relatives revealed 89,732 shared variants. Data analysis as outlined prioritized 24 variants, of which 16 had a CADD score >20. Based on functional relevance in neurodegeneration, type of mutation, protein expression in brain and absence from the healthy relatives, a frameshift mutation (p.Thr119Profs*22) in *WHRN* was retained as the most promising candidate variant and gene. The generated premature stop codon is located in exon 1 and predicted to trigger nonsense-mediated RNA decay (NMD). Full exonic resequencing of *WHRN* identified a series of missense mutations in both patients and controls. Missense variants detected in patients often mapped to functional protein domains (PDZ1 and PDZ2 domains), whereas in controls all variants mapped outside the functional domains. This enrichment was significant (p=0.0018, SKAT-O). **Conclusion:** We identified a co-segregating frameshift mutation in the *WHRN* gene on chr9q32 in an autosomal dominant FTD family. Screening of *WHRN* in unrelated patients and controls, revealed a number of missense mutations affecting the PDZ domains - involved in anchoring receptor proteins to the cytoskeleton and hold together signalling complexes at cellular membranes - potentially also resulting in loss of protein function. The effect of the LoF mutations will be investigated in patient-derived biomaterials and the mutation screen expanded to a larger patient cohort to provide further supportive evidence for causal implication of *WHRN* in FTD and neurodegenerative dementias.

2367W

Inositol deficiency induce autophagy impairing via PI3K/ Akt/ mTOR / p70S6K signaling. J. Wang¹, J. Guo¹, Z. Chen², H. Yue¹, S. Li¹, Y. An³, D. Luo¹, F. Zhang⁴, R. Zhong⁵, J. Lyu⁶, Y. Yang⁵, A. Chen⁶, L. Shi¹, X. Wang¹, F. Wang¹, Y. Bao¹, J. Zou¹, Z. Guan¹, T. Gao³, Z. Zhu¹, B. Niu¹, J. Huang⁷, H. Wang⁸, T. Zhang¹, J. Wu⁸. 1) Capital Institute of Pediatrics, Beijing, China; 2) State Key Laboratory of Genetic Engineering, Collaborative Innovation Center of Genetics and Development, School of Life Sciences, Fudan University, Shanghai, 200438, China; 3) Beijing Key Laboratory of Environmental & Viral Oncology, College of Life Science and Bioengineering, Beijing University of Technology, Beijing 100124, China; 4) Chinese Academy of Inspection & Quarantine, Beijing 100023, China; 5) National Center for Safety Evaluation of Drugs, National Institutes for Food and Drug Control, Beijing 100176, China; 6) Miami University, Department of Chemistry and Biochemistry, 651 E. High St. Oxford, OH 45056; 7) Institute for Nutrition and Health, Chinese Center for Disease Control and Prevention; 8) Department of Biochemistry & Immunology, Capital Institute of Pediatrics, Beijing 100020, China.

Myo-Inositol (MI) is a small organic molecule which plays important roles in a diverse array of cellular activities. MI deficiency is related with the etiology of many diseases. Previous studies have underlined the vital role autophagy plays during neural tube development, while the PI3K/Akt/mTOR/p70S6K signaling pathway are primarily involved in the regulation of autophagy. In the present study, using the commercially available PI3-kinase ELISA kit, we detected the activity of the PI3-kinase MI deficiency NE-4C cells which were treated with 6.25 μ M Li₂CO₃ (the inhibitor of the key enzyme for MI synthesis and recycling). The results showed that PI3K activity was significantly higher in the Li₂CO₃-treated group ($P<0.05$), and the increased kinase activity was abolished by additional treatment with MI ($P<0.05$). Then in vitro enzymology experiment was performed, in which immunoprecipitated PI3K was treated with a gradient of inositol from 1 to 20 mg/L. We found that PI3K activity decreased dramatically with the increasing MI concentration in a dose-dependent manner when the inositol concentration was below 5 mg/L. The PI3K/Akt/mTOR/ p70S6K signaling pathway was significant activated in MI deficient NE-4C cells compare with the controls using the method of western blot. The level of LC3B and the number of autophagic vacuoles were analyzed by western blot and electron microscopy. We found that a significant impaired autophagy when the NE-4C cells were treated by Li₂CO₃, and the effects were abrogated by combining Li₂CO₃ with MI or signaling inhibitor (Ly294002 and Rapamycin). Which suggested that MI deficiency induced autophagy impairment through the Akt/mTOR/p70S6K signaling. We next validate the above result in our previous established MI deficiency mouse model, and found PI3K activity in the MI deficient NTD mouse model was significantly elevated in heart, neural and placenta tissues 8 hours after Li₂CO₃ injection. Furthermore, the maternal serum MI levels correlated with the PI3K activity in the mouse embryonic neural (MEN) tissue ($P<0.05$). The activity of Akt/mTOR /p70S6K signaling and the LC3B level were detected by western blot in MEN tissue. The result was coincide with that in NE-4C cells that the activity of Akt/mTOR /p70S6K signaling was over activated accompanied by the reduced level of LC3B. These results indicated that inositol deficiency induce autophagy impairing via PI3K/Akt/ mTOR/p70S6K signaling.

2368T

RAD9B variants identified in neural tube defects. Y. Lei¹, X. Cao¹, S. Wadhwa¹, R. Finnell^{1,2}. 1) Center for Precision Environment Health, Department of Molecular & Cellular Biology and Medicine, Baylor College of Medicine, Houston, TX.; 2) Collaborative Innovation Center for Genetics & Development, School of Life Sciences, Fudan University, Shanghai, China.

Neural tube defects (NTDs) which are the result of the failure of neural fold fusion during the first 4 weeks of embryonic development, are among the most severe of all human structure malformations. The global average incidence of NTDs is 0.5-1 per 1,000 newborns. Both the genetics and environmental factors contribute to the etiologies of NTDs. Although over 400 genes have been identified as contributing to NTDs in the mouse, the genetic etiology of human NTDs remains unclear. To date, there is no published genome wide association study on NTDs. No large scale (sample size >1000) whole exome sequencing or whole genome sequencing on NTDs have been performed let alone published. In our phase I whole genome sequencing of 100 NTD samples, we identified two predicted to be loss of function (LoF) variants (Chr12: 110960044 delA; Chr12: 110968402 insC) and one predicted-to-be-damaging missense variant (p.Ser354Gly) in RAD9B and when compared with the EXAC and the genomAD database, there are significant ($p < 0.05$) LoF variants enrichment in NTDs compared to public controls. In a phase II validation study, we re-sequenced 96 NTDs and identified three missense RAD9B variants (p.Ser10Gly, p.Gln146Glu and p.Gly221Arg). Two of the variants were predicted-to-be-damaging by SIFT software. RAD9B is a homolog of RAD9, which is an integral element of the PCNA-like HUS1-RAD1-RAD9(9-1-1) complex, that plays a role in genotoxin-induced CHK1 activation. Previous studies in mice found that heterozygotes of Rad9b mutant mouse exhibit abnormal neural tube closure, which is consistent with our finding that heterozygous RAD9B LoF variants are associated with NTDs. To explore the underlying mechanism by which RAD9B LoF variants contribute to NTDs, we performed functional analyses of the identified RAD9B variants in our WGS data. Two LoF variants disrupted RAD9B nucleus localization as well as DNA repair foci formation under oxidative stress. In a RAD9B knock out assay, we found that LoF variants could not counteract RAD9B siRNA effect on cell proliferation. In conclusion, our study indicates that RAD9B variants could contribute to the etiology of NTDs by adversely affecting normal RAD9B function. Our study is the first one to report RAD9B variants in human disease.

2369F

A Truseq RNA exome capture method for two different brain regions of restless legs syndrome patients. F. Sarayloo^{1,2}, A. Dionne-Laporte², D. Spiegelman², P.A. Dion^{2,3}, G.A. Rouleau^{1,2,3}. 1) Human Genetics, McGill University, Montreal, Quebec, Canada; 2) Montreal Neurological Institute, McGill University, Montreal (QC), Canada; 3) Department of Neurology and Neurosurgery, McGill University, Montreal (QC), Canada.

Introduction: Restless legs syndrome (RLS) is a prevalent disorder and individuals who are affected experience unpleasant sensations (mainly in their legs while at rest during evenings or nights). Since only movements can partially relieve these sensations, these individuals often refer to themselves as “nightwalkers” since they typically stand up and walk during the night to get rid of their symptoms. Approximately 77% of RLS cases have a family history, thus suggesting an important contribution of genetic factors. Unfortunately, the genes thus far associated with RLS explain less than 10% of the heritability. Despite these genetic advances, the molecular mechanisms underlying RLS remain to be established. **Methods:** Our study proposes to use transcriptomes (RNA-Seq) to explore dysregulations associated with the transcription factor encoded by *MEIS1*; the gene showing the strongest association with RLS. In 2008, a risk haplotype of *MEIS1* was reported to lower its RNA and protein expression in lymphoblastoid cells and specific brain region of RLS cases. We have generated RNA-Seq data from two human cell lines (SKNSH a neuroblastoma cell line and HEK293) across which we both over-expressed or knocked out *MEIS1*. The differential expression analysis was made using the edgeR algorithm and a list of dysregulated genes was delineated. To follow up and validate the dysregulation of these genes in RLS, we have flash-frozen brain tissues obtained from RLS cases. RNA was prepared from 30 thalamus and 31 pons samples of patients harboring the risk or non-risk alleles of *MEIS1*, as well as patients with other RLS risk alleles. We used the Truseq RNA exome capture method, as it is known to improve the quality of RNA-Seq libraries that are prepared from samples that were frozen for a long period. **Outcome:** Our study will result in a precious list of genes dysregulated in RLS, some of which may have already been reported to be involved in the disease. Other genes might represent novel candidates that should be examined for the presence of risk variants through genotyping (e.g. burden study) in unrelated cases; or through high-throughput sequencing in multiplex families. Overall, this study will have an important impact on our understanding of RLS.

2370W

Cerebral palsy gene discovery based on whole exome sequencing of multiplex consanguineous families. S. Bakhtiari^{1,2}, A. Tafakhori³, SC. Jin⁶, X. Zeng⁶, M. Corbett⁴, J. Liu^{1,2}, J. Heim¹, MM. Hockley⁶, B. Norton^{1,2}, H. Magee^{1,2}, C. Van-Eyk¹, RP. Lifton⁶, AH. MacLennan⁷, J. Gecz⁷, S. Padilla-Lopez^{1,2}, K. Bilguvari¹, H. Darvish⁶, MC. Krue^{1,2}. 1) Cerebral Palsy & Pediatric Movement Disorders Program, Barrow Neurological Institute, Phoenix Children's Hospital, Arizona, USA; 2) Departments of Child Health, Neurology, and Cellular & Molecular Medicine, University of Arizona College of Medicine, Phoenix, Arizona, USA; 3) University of Arizona College of Medicine, Phoenix, Arizona, USA; 4) Department of Genetics, Yale University, Connecticut, USA; 5) Department of Medical Genetics, Semnan University of Medical Sciences, Semnan, Iran; 6) Laboratory of Human Genetics and Genomics, Rockefeller University, New York, USA; 7) Robinson Research Institute, The University of Adelaide, Adelaide, Australia; 8) Department of Neurology, School of Medicine, Imam Khomeini Hospital and Iranian Center of Neurological Research, Tehran University of Medical Sciences, Tehran, Iran.

Cerebral Palsy (CP) is a common neurodevelopmental disorder which impairs movement as a result of disruptions to the developing brain. CP has been attributed to several environmental conditions including preterm delivery, asphyxia, pre-/perinatal infection and pre-/perinatal stroke. However, in about one third of CP cases, none of these risk factors are present ('cryptogenic'). Our recent analyses in sporadic cryptogenic CP cases have demonstrated that CP is genetically heterogeneous. Here, we performed whole exome sequencing in eleven multiplex, consanguineous families in whom affected children met international consensus criteria for CP with a nonprogressive clinical course and onset before age 3. Most affected individuals were non-ambulatory (Gross Motor Functional Classification Scale 4-5/5). We focused our analysis on deleterious homozygous missense (with MetaSVM = D or CADD $\geq$ 20), non-frameshift insertions/deletions and loss of function (nonsense, canonical splice site, frameshift) variants. Variants segregating in affected siblings and absent in healthy family members were prioritized. In all, we identified 10 putatively causal genes, including *SLC25A23*, *RNF170*, *AP4M1*, *SIL1* (2 families), *ALG2*, *ASNS*, *MAN2B1*, *CACNA1A*, *ALS2*, and *HPRT1*. Among these genes, the association of *SLC25A23* with human disease is novel. We also report the new association of biallelic variants in *RNF170* with a cerebral palsy phenotype (heterozygous mutations in this gene cause neuroaxonal dystrophy affecting the dorsal columns; #MIM 608984). Clinical follow-up revealed that some patients experienced symptom progression from the time of initial ascertainment while others did not. Our findings indicate that studying multiplex, consanguineous families may be a valuable approach for cerebral palsy gene discovery, particularly in severe Mendelian forms of the disorder, but careful phenotyping is crucial in order to differentiate genetic forms of CP from progressive disorders. Although analysis of sporadic cases previously revealed an enrichment of deleterious *de novo* and X-linked variants, the current analysis also indicates a role for biallelic variants in CP.

2371T

Utility of massively parallel targeted resequencing data for detection of copy number variants in developmental and epileptic encephalopathies. M.G. Mehafeey¹, B. Strobel¹, H. Kaur¹, I. Phelps¹, J. Nguyen¹, E.A. Fuerte¹, A.L. Schneider², G. Hollingsworth², K. Boysen², D. Doherty¹, I.E. Scheffer², H.C. Mefford¹. 1) Department of Pediatrics, University of Washington, Seattle, WA; 2) Epilepsy Research Centre, Department of Medicine, Austin Health, The University of Melbourne, Heidelberg, Victoria, 3084, Australia.

Objective: Copy number variants (CNVs) are regions of the DNA that have either been duplicated or deleted with respect to the reference sequence. CNVs are defined as being > 1kb in length and can be up to dozens of megabases in size. CNVs can be benign population variants or may be disease-related. CNVs are most reliably detected using chromosome microarrays, but many efforts have been made to identify CNVs reliably from massively parallel sequence data. In the developmental and epileptic encephalopathies (DEEs), clinically detectable pathogenic CNVs account for 5-10% of cases. Smaller, intragenic CNVs may not be detected by traditional array methods, but are likely to explain a subset of cases. We have generated targeted sequencing data for a large cohort of patients with unsolved DEE. Here we present an approach using targeted massively parallel sequencing data to identify potentially pathogenic CNVs. **Methods:** We performed targeted capture using molecular inversion probes (MIPs) or single molecule MIPs (smMIPs) followed by sequencing of 267 known and candidate genes for DEE in a cohort of 1071 probands with DEEs. After BWA alignment, read depths for each resequencing probe for all samples were analyzed. Z-scores were calculated to give a unique deviation metric from the normal sampling distribution for each individual over each targeted resequencing probe as previously described (PMID 23160955). We used a test set (53 genes, 823 samples) to ensure we could accurately detect two previously validated CNVs (4.85Mb del and 4.05Mb dup). In-house scripts were then written to automatically scan the remaining genes for areas with 4 or more consecutive significant z-scores ($x > 1.5$ for duplications or $x < -1.5$ for deletions). These areas were also given a CNV score based on a summation of logistic function values from key metrics across the CNV being evaluated to create a graphical output of the results. **Results:** We identified 261 predicted rare or novel CNVs in 193 individuals ranging from ~500 bp (single exon deletion) to 46 Mb (multiple gene duplication). Ongoing custom array studies to validate putative CNVs will determine the accuracy of our calling algorithm. This method has the potential to identify small, intragenic CNVs that are missed by traditional array methods, increasing the efficiency of detecting pathogenic sequence and copy number variants using a single assay.

2372F

A genome-wide DNA methylation analysis of CpG-polymorphisms reveals a genetic association between a chromosome 6 locus and age of onset in *C9orf72* carriers. M. Zhang¹, R. Ferrari², M.C. Tartaglia^{1,3,4}, J. Keith⁵, C. Sato¹, J. Robertson¹, L. Zinman⁵, E. Rogaeve¹, the International FTD-Genomics Consortium. 1) Tanz Centre for Research in Neurodegenerative Diseases, University of Toronto, Toronto, Canada; 2) Department of Molecular Neuroscience, Institute of Neurology, University College London, London, UK; 3) Krembil Neuroscience Center, University Health Network Memory clinic, Toronto Western Hospital, Toronto, Canada; 4) Department of Medicine, Division of Neurology, University of Toronto, Toronto, Canada; 5) Sunnybrook Health Sciences Centre, University of Toronto, Toronto, Canada.

The G₄C₂-repeat expansion in *C9orf72* is the most common known cause of amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). High phenotypic heterogeneity of *C9orf72* patients includes a wide-range in age of onset (AO), but AO modifiers are largely unknown. AO could be influenced by environmental and genetic factors both of which may trigger DNA methylation (DNAm) changes at CpG-sites. CpGs are the most mutable sites in the human genome because methyl-C can spontaneously deaminate to T (e.g., 35% of all coding mutations occur at CpG-sites). Allele-specific DNAm is largely attributed to CpG-SNPs, which have often been detected within promoter regions, transcription factor binding sites and DNase I hypersensitive sites, thus regulating the level of gene expression. We tested the hypothesis that AO in *C9orf72* patients is associated with some common single nucleotide polymorphisms (SNPs) causing a gain or loss of CpG-sites (CpG-SNPs) and thus resulting in DNAm alterations. First, we estimated the association between AO in *C9orf72* patients (n=46) and DNAm levels at all 7603 common CpG-SNPs available on the 450k BeadChip, which was followed by a genetic association study of the discovery (n=144) and replication (n=187) *C9orf72* cohorts. We found that AO was reproducibly associated with SNPs within a 124.7kb linkage disequilibrium block tagged by top-significant SNP and containing two overlapping genes (*LOC101929163* and *C6orf10*). A meta-analysis of 331 *C9orf72* carriers revealed that every A-allele of top-significant SNP reduced the Hazard ratio by 30% (p=0.0002); and the median AO of AA-carriers was 6 years later than GG-carriers. Gene expression studies of frontal cortex tissues from 25 ALS autopsy cases revealed that the G-allele of the top-significant SNP is associated with increased brain expression of a nearby gene *HLA-DRB1* (involved in initiating immune responses), while the A-allele is associated with their reduced expression. Our findings suggest that carriers of GG-genotype (linked to an earlier AO in *C9orf72* patients) might be in a more pro-inflammatory state (e.g., by microglia) than AA-carriers. Further investigating the functional links within the *C6orf10/LOC101929163/HLA-DRB1* pathway will be critical to better define age-dependent FTD-ALS pathogenesis.

2373W

Whole exome sequencing in epilepsy reveals significant gene burden of ultra-rare deleterious variants. Y. Feng^{1,2} on behalf of the EPI25 Collaborative. 1) Stanley Center for Psychiatric Research, Broad Institute, Cambridge, MA, USA; 2) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA.

Whole exome sequencing (WES) in rare, severe epilepsies have successfully identified gene targets that opens the opportunity for development of individualized treatments. For common epilepsy types, while family studies and genome-wide association studies have revealed a consistent genetic influence, the number of specific gene associations remains limited. As epilepsy is a highly heterogeneous disorder, large and well-phenotyped samples are required for the characterization of genetic contribution across epilepsy syndromes. In light of this, the Epi25 Consortium was formed, an international collaboration of multiple patient cohorts that aims to elucidate the genetic risk underlying rare and common epilepsies through WES. In our year 2 effort, WES data were generated for 1453 patients with epileptic encephalopathy (EE), 4443 with genetic generalized epilepsy (GGE), and 5486 with non-acquired focal epilepsy (NAFE). For comparison, 17668 controls without any known neuropsychiatric condition were aggregated from independent cohorts. All samples were jointly called at the Broad Institute and went through rigorous quality-control procedures, with ancestry matched and exome capture difference minimized. Overall burden analysis showed a significant enrichment of ultra-rare protein-truncating variants (PTVs) in loss-of-function intolerant genes for all three subtypes compared to controls (ORs=1.44, 1.38, 1.24, p-values=1e-07, 5e-12, 1e-06 for EE, GGE, and NAFE). An increased burden of ultra-rare PTVs and damaging missense variants in epilepsy patients was also observed in missense-constrained genes, brain-specific genes, known EE genes, and genes associated with intellectual disability. Gene-based burden tests re-captured several established epilepsy genes in EE patients, including *SCN1A*, *STXBP1*, and *KCNB1* (ORs=11.6, 18.4, 27.6, p-values=4e-09, 3e-06, 2e-05). No genes reached exome-wide significance for the two common epilepsies. However, for GGE, some channel or transporter genes were among the strongest associations, including *SLC6A1*, *CACNA1G*, and *GABRA1* (ORs=24.6, 4.2, 21.5, p-values=9e-05, 3e-04, 3e-04); for NAFE, two familial focal epilepsy genes also came up at the top (*LG11* and *DEPDC5*; ORs=11.35 and 3.53, p-values=4e-04 and 2e-03). These results confirm a crucial role of ultra-rare deleterious variants in both common and rare epilepsies and provide a strong replication to previous gene findings in the largest epilepsy sequencing study to date.

2374T

A microRNA-328 binding site in *PAX6* discriminates between rolandic and other epilepsies. *N. Panjwani¹, F. Lin¹, L. Addis^{2,3}, S. Tang^{2,4}, D.K. Pal^{5,6}, L.J. Strug^{1,7,8}.* 1) Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada; 2) Department of Basic and Clinical Neuroscience, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, SE5 9RX, UK; 3) Neuroscience Discovery Research, Eli Lilly and Company, Erl Wood, Surrey, GU20 6PH, UK; 4) Evelina London Children's Hospital, London, SE1 7EH UK; 5) King's College Hospital, London, SE5 9RS UK; 6) Northwick Park Hospital, Middlesex, HA1 3UJ UK; 7) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, ON, Canada; 8) Biostatistics Division, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada.

Introduction Rolandic epilepsy (RE) is a neurodevelopmental disorder characterized by seizures and centroparietal sharp waves (CTS) on electroencephalogram (EEG). Fine-mapping of the CTS linkage region identified a variant in the 3'-UTR of *PAX6* (rs662702) known to disrupt microRNA-328 binding (Panjwani et al, 2016). The impact of this variant on other epilepsies, however, is not known. We investigated the relationship between rs662702 and Electrical Status Epilepticus during Slow Sleep (ESES), Myoclonic Astatic Epilepsy (MAE), Childhood Absence Epilepsy (CAE) and Juvenile Absence Epilepsy (JAE), and we replicated in an independent RE sample. **Methods** N=177 individuals with RE, ESES, or MAE were genotyped on the InfiniumOmni2.5 BeadChip and 126 patients with CAE and JAE on the HumanOmniExpress BeadChip. 263 passed QC and clustered with Europeans. We used the European unrelated subset of the 1000 Genomes Project (phase 3) genotyped on the HumanOmni2.5 as controls (N=522), and logistic regression to test for association between rs662702 and each epilepsy type. The rs662702 SNP was also assessed in a prediction model to discriminate between RE and other epilepsy diagnoses adjusted for sex and age of first seizure onset.

Diagnosis	CC	CT	TT	Total	MAF	Odds Ratio	P-value
1KG EUR Controls	462 (0.885)	56 (0.107)	4 (0.008)	522	0.061	-	-
RE Discovery (Panjwani et al, 2016)	115 (0.757)	31 (0.204)	6 (0.039)	152	0.141	1.97	1.50E-03
RE Replication	19 (0.63)	11 (0.37)	0	30	0.183	3.31	8.30E-04
ESES	66 (0.94)	4 (0.06)	0	70	0.029	0.47	0.14
MAE	36 (0.88)	5 (0.12)	0	41	0.061	0.98	0.96
CAE	84 (0.913)	7 (0.076)	1 (0.011)	92	0.049	0.84	0.62
JAE	27 (0.90)	3 (0.10)	0	30	0.05	0.83	0.75

Results In the replication RE sample the frequency of the rs662702 T allele is elevated (replication odds ratio=3.31; P=8.3e-4), whereas for all other epilepsies tested no significant difference from the 1KG control population was observed (P>0.05). A training set comprised of 60% of the RE, MAE and CAE samples and five-fold cross-validation was used to define a model comprised of rs662702 genotype, sex and age of first seizure onset, to discriminate RE from the other epilepsy diagnoses. External validation of the model on 91 (40%) of the remaining patients yielded an AUC of 0.690. **Discussion** The frequency of the T allele in rs662702 is elevated in two independent RE samples, while the frequency in the other epilepsy types tested was similar to the 1KG controls, suggesting specificity for RE. A model that includes this SNP shows promise as a diagnostic tool for RE.

2375F

Genetic generalized epilepsies and parasomnias are linked to *CLCN2*, *EFHC1* and *GABRA1* genes in Colombian families. *N. Pineda Trujillo¹, D. Cornejo-Sanchez¹, J. Carrisoza-Moog¹, R. Solarte-Mila², D. Cabrera-Hemer², C. Gomez-Castillo³, R. Thomas⁴, S.-M. Leal⁵, W. Cornejo-Ochoa².* 1) Grupo Mapeo Genético, Facultad de Medicina, Universidad de Antioquia. Medellín 050010470, Antioquia, Colombia; 2) Grupo Pediciencias, Facultad de Medicina, Universidad de Antioquia. Medellín 050010470, Antioquia, Colombia; 3) IPS Univesitaria, Universidad de Antioquia. Medellín-Colombia; 4) Institute of Neuroscience, Newcastle University, United Kingdom; 5) Center for Statistical Genetics, Columbia University, New York, USA.

The genetic architecture of focal and genetic generalised epilepsies (GGE) remain cryptic. Endophenotypes, such as sleep-disorders, represent an attractive way of delineating the genetic factors. Epilepsy GWAS and exome sequencing studies have identified common and rare variants of importance. Epilepsy GWAS and exome sequencing studies have identified common and rare variants of importance. Here, we test the contribution of genetic variants in *CLCN2*, *EFHC1* and *GABRA1* genes in a set of extended Colombian families with seizures and sleepwalking (SW) or sleep paralysis (SP), or both. These families presented with GGE and at times focal seizures were also present, in combination with either SW or SP, or both. The three genes coding-region (*CLCN2*, *EFHC1* and *GABRA1*) was sequenced by Sanger in 60 affected individuals. After selecting tagSNPs, 39 variants were typed in 723 individuals (283 affected individuals) in 155 families. Genotypes were obtained by KASP methodology (LGC genomics). Statistical analyses included H-WE, non-parametric linkage and association tests, combining all convulsive affection status with/without sleepwalking and sleep paralysis. No discrimination between focal or generalized seizures were considered. Twenty-three out of 39 SNPs were successfully typed. All of the SNPs were found in H-WE (p>0.05/23). No Mendelian inconsistencies were found. After base calling rates criteria, 143 families were analysed. Suggestive association was found for SNP rs11920716 (p=0.0063).When SW was added to epilepsy, significant linkage findings arose. Thus, thirteen gene variants at *CLCN2* (p=0.0003), including rs2293603, rs92449317, rs11656822, rs11920716, rs9820367 among others; rs41275339, rs4260711, rs12187575, rs6867604, rs2279020 at *GABRA1* (p=0.0007) and rs3804505, rs2065144, rs62407897, rs1266787, rs17851770 at *EFHC1* (P=0.0001) were, respectively, linked to this phenotype. Extended phenotype (epilepsy + SW+ SP) was even more significantly linked to these gene variants (P< 0.0003) only for genes *CLCN2* and *EFHC1*.We suggest that it may be valuable to reappraise the importance of *CLCN2* and *EFHC1* in special populations, such those with sleepwalking and sleep paralysis comorbidities. This work also demonstrates the value of investigating families with epilepsy from outside of traditional European populations.In conclusion, our results suggest that *CLCN2* and *EFHC1* genes are linked to GGE + SW + SP; whilst *GABRA1* gene is linked to GGE + SW in Colombian families.

2376W

Over 70% of pathogenic in Czech epilepsy cohort is propagated into protein domains. D. Stanek¹, P. Lassuthova¹, L. Sedlackova¹, J. Neupauerova¹, K. Sterbova², A. Rebelo³, D. Bis³, C. Saghira³, S. Zuchner³, P. Seeman^{1,2}. 1) DNA laboratory, Department of Pediatric Neurology, 2nd Faculty of Medicine, Charles University in Prague and University Hospital Motol, Prague, Czech Republic; 2) Department of Pediatric Neurology, 2nd Faculty of Medicine, Charles University in Prague and University Hospital Motol, Prague, Czech Republic; 3) Department of Human Genetics and John P. Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, FL 33136, USA;.

Introduction: Many studies that variation in conserved protein domains are less tolerated and thus are more likely connected to a clinical phenotype. These potentially significant consequences of DNA variation on domain structure and function makes them a valuable target for genomic studies. In the previous project, we created the database of protein domains published in PubMed, mapped to a chromosomal location. Our lab has long-time experience with NGS analysis in patients with neurodegenerative disorders. We are continuously implementing new approaches to optimize analysis workflow for NGS data. **Methods:** We performed analysis on 60 variants, which were evaluated as pathogenic or likely pathogenic according to ACMG criteria, gathered from the cohort of patients with epileptic encephalopathies. Variants were annotated by our custom database (manuscript in preparation), that provides information about all protein domains/features published in PubMed mapped on its chromosomal location. With this kind of information, we are able to determine whether is variant translated into the specific domain and how a change in this domain can affect the function of a protein. **Results:** Out of 60 variants, 71% (43) was propagated into 94 protein domains. Most common were Ion channel domains (34 occurrences), binding domains (9 occurrences) and transmembrane regions (7 occurrences). These regions are very conserved and changes in these structures can be harming the whole organism. **Conclusion:** Our result showed, that most of the pathogenic variants are translated into protein domains. This information about protein domain can help experts in evaluation of NGS. On the other hand, we can see that many of these variants are not connected to any protein domain. Based on this fact we cannot decide pathogenicity on variant only based on these criteria. Supported by: AZV 15-33041, GAUK 388217 and NINDS R01NS075764.

2377T

Exome-wide burden analysis of loss-of-function variants in Parkinson's disease. B.I. Bustos¹, V. Escott-Price², N.M. Williams³, M.A. Nalls³, A. Brice⁴, J.C. Corvol⁵, S. Lesage⁶, T. Gasser⁶, J. Hardy^{7,8}, P. Heutink^{5,6}, N.W. Wood^{7,9}, A.B. Singleton³, H.R. Morris¹⁰, S.J. Lubbe¹, The International Parkinson's Disease Genomics Consortium. 1) Department of Neurology and Center of Genetic Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA; 2) Institute of Psychological Medicine and Clinical Neurosciences, Medical Research Council Centre for Neuropsychiatric Genetics and Genomics, Department of Psychological Medicine and Neurology, Cardiff University School of Medicine, Cardiff, UK; 3) Laboratory of Neurogenetics, National Institute on Aging, Bethesda, Maryland, USA; 4) Inserm U 1127, CNRS UMR 7225, Sorbonne Universités, UPMC Univ Paris 06 UMR S 1127, Institut du Cerveau et de la Moelle épinière, ICM, F-75013, Paris, France; 5) German Center for Neurodegenerative Diseases (DZNE), Tübingen, Germany; 6) Department of Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research, University of Tübingen, Tübingen, Germany; 7) Department of Molecular Neuroscience, Institute of Neurology, University College London, London, UK; 8) Reta Lila Weston Institute, University College London Institute of Neurology, Queen Square, London, UK; 9) UCL Genetics Institute, University College London, London, UK; 10) Department of Clinical Neuroscience, Institute of Neurology, University College London, London, UK.

Parkinson's disease (PD) is a complex motor neurodegenerative disorder, affecting ~1% of the population over 65 years old. Genome-wide genotyping and sequencing based association studies have discovered several genes and pathways associated with causality or risk. However, they only explain a small proportion of the disease heritability, therefore others remain to be identified. Here we present an exome-wide burden analysis in PD in order to detect a set of variants that could help us to identify novel gene-sets and pathways. We have analyzed the International Parkinson's Disease Genomics Consortium (IPDGC) exome-wide study comprising, after QC, 1,530 unrelated individuals (1,078 cases and 452 healthy controls). Annotation and functional prediction of variants was done with ANNOVAR. Whole-exome burden analysis for different frequency of variants was performed using logistic regression, modeling the allelic counts for each individual with disease status and adjustment for gender, first 5 principal components and sequencing coverage. We have found a significant burden of loss-of-function low frequency variants in cases vs controls (CADD phred score > 20, MAF < 0.05; $p=1.59 \times 10^{-4}$). The signal increases its significance for novel singleton variants ($p=1.29 \times 10^{-20}$, OR=1.125). We found a stronger effect within early onset PD cases (<40 years old age-at-onset, $p=1.88 \times 10^{-28}$, OR=1.192) and weaker for late-onset patients (>60 years old age-at-onset, $p=1.4 \times 10^{-4}$, OR=1.123). In a gene-set approach, we analyzed variants mapping to loss-of-function intolerant genes (ExAC pLI score > 0.9, N=3,043/18,224) and obtained an increased effect in all and early onset, compared to late-onset cases (OR=1.232, OR=1.349 and OR=1.180, respectively). Our results indicate that private loss-of-function variants are significantly enriched in PD patients, providing new insights into the genetic architecture of this disease. Expanded gene-set, pathway analysis and validation efforts are currently in progress in order to unravel underlying mechanisms of these variants and confirm our findings.

2378F

A first-degree family history of Alzheimer's disease and apolipoprotein E genotype influences cognition. J.S. Talboom^{1,2}, A.K. Haberg^{3,4}, M. De Both^{1,2}, M. Naymik^{1,2}, A.L. Siniard^{1,2}, L. Ryan^{2,3}, E.L. Glisky^{2,3}, C.L. Lewis^{1,2}, M.A. Fritz¹, M.J. Huentelman^{1,2}. 1) The Translational Genomics Research Institute, Phoenix, Arizona 85004 USA; 2) Arizona Alzheimer's Consortium, Phoenix, Arizona 85004 USA; 3) University of Arizona, Tucson, Arizona 85721 USA; 4) Norwegian University of Science and Technology, Trondheim 7491 Norway.

Background A first-degree family history Alzheimer's disease (FH) can be a proxy for heritable and non-heritable risk factors of dementia. However, the exact influence of FH on cognition across the lifespan is poorly understood. Further, the presence of FH specific interactions with genetics on cognition remains largely unexplored. We examined the influence of FH and a well-known Alzheimer's disease genetic risk factor, on paired associate learning (PAL).

Methods We developed a web-based PAL task (at www.mindcrowd.org) and tested over 75,000 individuals between the ages of 18-85. Next, we examined the apolipoprotein E (APOE) genotype, in over 500 FH positive individuals, via dried blood spot collection from the cohort. **Results** FH was associated with significantly decreased PAL performance. This difference was larger in participants under the age of 60. Propensity score matching analysis revealed an effect size of approximately half a word pair deficit in FH positive individuals. Further, FH positive carriers of the APOE E4 allele also demonstrated a higher risk for decreased PAL performance. **Conclusions** This study suggests that FH and APOE genotype are important factors that modify the trajectory of an individual's cognitive performance across their lifespan.

2379W

Genetic study in Taiwan's Hirschsprung disease. W. Yang^{1,2}, J.C. Chen¹, P.L. Chen². 1) Chang Gung Memorial Hospital, Linkou Branch, No.5, Fuxing St., Guishan Dist., Taoyuan City, Taiwan 333; 2) Graduate Institute of Clinical Medicine, National Taiwan University, No.7.Chung San South Road, Taipei 10002.

Introduction Hirschsprung's disease (HSCR, OMIM142623) is a disease caused by congenital absence of intrinsic ganglion cells in various length of the distal intestine and results functional bowel obstruction. HSCR has an incidence around 1 in 5000 to 1 in 10000 live births, differs with gender and race. Asian has the highest incidence, and male is affected 3.5 to 4 times more likely than female. According to the different length of aganglionic intestine, HSCR can be further divided into short (S-HSCR), long (L-HSCR), or total colon aganglionosis (TCA). S-HSCR accounts for more than 80% of the patients. Recurrence risk of a S-HSCR patient's sibling is 5% and increase to 10% to 20% for L-HSCR or TCA patients. We wish to find out the genetic variations of Taiwan's Hirschsprung disease and discuss the possibility to prevent disease pass down to the next generation. **Patient and Methods** Patients who was diagnosed as Hirschsprung's disease by pathology reports were included. A 31 genes NGS panel is designed according to previous literatures. All the genes are sequenced for the whole gene, except for NRG3. Due to the size limitation, NRG3 will be surveyed for the exons only. All the DNAs of the probands were send for NGS analysis to find any mutation genes in the panel. The mutations found in NGS were further confirmed by Sanger sequencing in all family members. **Results** Between 2014/07 to 2018/05, We included 42 families, 147 participants into our study. Half of the patients are L-HSCR, 8 are familial type and the male to female ration is 3 to 1. Mutations were found in 6 of the familial patients and 24 of the sporadic patients. The detection rate of a possibly pathogenic mutation is 75% in familial case and 63.16% for sporadic cases. *RET* is the most common found gene and was found in 11 patients. In the long segment group, 21.74%(5 patients) has either frameshift or stop codon. Three frameshifts located in *RET* exon 7 or exon 11, 1 frameshift in *L1CAM* and a stop codon in *NRG1* exon 2. For short segment HSCR patients, one patient had stop codon in *EDNRB* exon 7, and the other had a frameshift in *ZEB2* exon7. They were all heterozygous, de novo mutations. **Conclusions** All long segment HSCR patients should receive at least a panel-based work up for gene mutations. In vivo fertilization for the patients found a de novo obvious deleterious mutation, such as frameshift or stop codon on a HSCR causative gene, might help avoid pass down the disease to their off springs.

2380T

Meta-analysis of migraine with over 93,000 cases and 730,000 controls identifies 124 risk loci. H. Hautakangas¹, P. Gormley^{2,3}, A. Auton⁴, N. Litterman⁴, A. Palotie^{1,2,3,5,6}, M. Pirinen^{1,7,8} on behalf of the 23andMe Research team and International Headache Genetics Consortium (IHGC). 1) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; 2) Psychiatric & Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; 3) The Stanley Center for Psychiatric Research, The Broad Institute of MIT and Harvard, Cambridge, MA, USA; 4) 23andMe, Inc., Mountain View, CA, USA; 5) Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA; 6) Department of Neurology, Massachusetts General Hospital, Boston, MA, USA; 7) Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland; 8) Department of Public Health, University of Helsinki, Helsinki, Finland.

Migraine is a common brain disorder typically characterized by disabling episodes of severe headache accompanied by nausea and hypersensitivity to light, sound and smell. Its prevalence is around 15% worldwide and family and twin studies estimate a heritability of about 40%, pointing to a genetic component of the disorder. The largest genome-wide association study reported so far identified 38 loci and suggested that these loci are enriched for genes highly expressed in vascular and smooth muscle tissues (Gormley et al. Nat Gen 2016). In this meta-analysis, we increase the migraine sample size by 56% by adding to the earlier data 22,644 cases from 23andMe and 10,881 cases from the UK Biobank. This takes our total sample size to 93,199 cases and 733,977 controls. All the new samples have self-reported phenotype while a subset of the earlier data was ascertained through clinics. Using LD score regression (LDSC) we estimate that the genetic correlation between the new cohorts and the earlier data is above 0.99. Based on this observation we carried out a fixed-effects meta-analysis. We identified 124 genome-wide significant ($P < 5 \times 10^{-8}$) autosomal single-nucleotide polymorphisms (SNPs) that were at least 250 kb apart from each other. Of these 124 loci, 92 are reported here for the first time while 32 loci overlap with the previously reported 37 autosomal loci. To investigate the biology behind migraine, we evaluated whether the polygenic migraine signal was enriched near genes with activating histone marks specific to a certain tissue or cell type by applying LDSC to specifically expressed genes (LDSC-SEG). We report enrichment in six central nervous system cell types and four cardiovascular cell types as well as one cell type of the digestive system at a false discovery rate of 5%. Our results point to both neuronal and vascular cell types, consistent with earlier evidence that both systems are important in migraine, but they potentially relate to different subtypes of migraine. We used information on migraine with aura (6,332 cases) and migraine without aura (8,348 cases) to further stratify the risk loci by migraine subtypes available to us. To conclude, we report the largest migraine GWAS to date, with 92 new loci and enrichment of association signal in both central nervous system and cardiovascular system. These data provide unprecedented possibilities to further disentangle the complex and heterogeneous genetic background of migraine.

2381F

Genomic variants associated with cognitive impairment in Parkinson's disease. S. Chung¹, N. Choi¹, J. Kim², K. Kim¹, M. Kim³, Y. Kim¹, H. Ryu¹, K. Park¹. 1) Neurology, Asan Medical Center, Seoul, South Korea; 2) Neurology, Metro hospital, Anyang, South Korea; 3) Neurology, Bobath Memorial Hospital, Seongnam, South Korea; 4) Neurology, Kyungpook National University Hospital, Daegu, South Korea.

Background: Cognitive impairment is a frequent nonmotor symptom in Parkinson's disease (PD). There is substantial heterogeneity in the onset timing of symptom, severity, and long-term outcome of cognitive impairment in PD. Etiology of cognitive impairment in PD is still unclear and pathological features are highly variable. There is no convincing biomarker to predict the development of cognitive impairment in patients with PD. **Objectives:** We aimed to identify the genomic variants that are associated with cognitive impairment in patients with PD. **Methods:** Genomic data was produced in patients with PD (N=1,021), using the Korean Chip (K-CHIP), Affymetrix Axiom KORV1.1 (variants number of 827,400), which contains imputation genome-wide association study (GWAS) grid and other GWAS loci, functional variants of nonsynonymous exome, pharmacogenetics variants, variants in genes involved in absorption, distribution, metabolism and excretion (ADME) of drugs, and expression quantitative trait loci (eQTL). Genomic analysis was performed according to the scores of Mini-Mental State Examination (MMSE) and Montreal Cognitive Assessment (MoCA). **Results:** One upstream variant in Chromosome 17 (OR=0.54, CI=0.44-0.69, $P=1.45 \times 10^{-7}$) and one intronic variant in Chromosome 3 (OR=0.31, CI=0.20-0.50, $P=3.77 \times 10^{-7}$) were significantly associated with the lower scores of MMSE (score < 24) after Bonferroni correction. One intronic variant in Chromosome 10 (OR=0.53, CI=0.41-0.70, $P=4.56 \times 10^{-6}$) was significantly associated with the lower scores of MoCA (score < 23). Several variants showed associations with lower scores of MMSE or MoCA, but they did not remain significant after Bonferroni correction. **Conclusions:** This study identified new loci associated with lower scores of cognitive screening tests in patients with PD. Further studies are needed to confirm our findings using more comprehensive cognitive tests.

2382W

Deep mitochondrial sequencing in Parkinson's disease reveals possible genetic modifiers of reduced penetrance in Parkin/PINK1 mutation carriers. J. Trinh¹, A. Grunewald², K. Wasner³, A. Hicks³, P. Bauer⁴, S. Imhoff⁵, K. Kandaswamy⁴, N. Ouzren⁴, M. Werber⁴, M. Weiss⁴, A. Rolfs^{4,5}, P. Pramstaller⁴, P. Seibler¹, K. Lohmann¹, C. Klein¹. 1) Institute of Neurogenetics, Universität zu Lubeck, Luebeck, Germany; 2) Luxembourg Centre for Systems Biomedicine, University of Luxembourg; 3) EURAC, Bolzano, Italy; 4) Centogene AG, Rostock, Germany; 5) University of Rostock, 18147 Rostock, Germany.

Background: Biallelic mutations in *Parkin* and *PINK1* are fully penetrant and cause recessively inherited Parkinson's disease (PD). On the other hand, heterozygous mutations in these genes may be considered as a risk factor for PD or even act in a dominant manner with highly reduced penetrance. Since both *Parkin* and *PINK1* function in the removal of dysfunctional mitochondria, we hypothesize that mitochondrial DNA (mtDNA) mutations influence the age-dependent penetrance and thus the onset of disease. In other words, they may be regarded as a 'second hit' in affected *Parkin*/*PINK1* mutation carriers. **Method:** We performed deep mtDNA sequencing in 124 individuals with Illumina NextSeq in blood-derived DNA and brain tissue-derived DNA to assess mitochondrial mutational load including somatic mosaicism and individual mtDNA variants. Patients were recruited from Germany and Italy and comprised carriers of *PINK1* (n=89) and *Parkin* mutations (n=159), idiopathic PD patients (n=83) and controls (n=100). We also compared mtDNA variants between iPSC-derived neurons and post-mortem brain frontal cortex, occipital lobe, and substantia nigra of patients with *Parkin* mutations and controls. **Results:** A mean coverage of >10,000X was achieved with high sensitivity of detecting low level heteroplasmic (<15%) variants. *Parkin* mutation carriers have cumulatively more variants including both single nucleotide variants (SNV) and copy number variations (CNV) compared to controls and idiopathic PD (p<0.01). *PINK1* mutation carriers showed a non-significant trend (p=0.10). Comparing early and late onset *Parkin* mutation carriers, we identified a potential protective variant in *ATP6* (p.A177T). iPSC clones and differentiated neurons reveal a stable process in maintaining mtDNA integrity and postmortem brain tissues reveal differential cumulative mtDNA SNV and CNVs. **Conclusion:** *Parkin* and *PINK1* mutations may predispose to an increased mitochondrial mutational load. Replication and random segregation of heteroplasmic mitochondrial DNA may partially explain the variable expressivity (i.e. age at onset) and the reduced penetrance in (heterozygous) *Parkin*/*PINK1* mutation carriers.

2383T

Phenotype-genotype characterization of a new intellectual disability syndrome with acroosteolysis: New roles for TMEM41B protein. F. Pirozzi¹, A. Gevorgyan^{1,2}, S. Swagemakers^{3,4}, B.C.J. van der Eerden⁵, J.P.T. van Leeuwen⁵, A.R. Khan⁶, G. Neri⁷, P.J. van der Spek^{3,4}, P. Chiurazzi¹. 1) Center for Integrative Brain Research (CIBR), Seattle Children's Research Institute, Seattle (USA); 2) Department of Genetic and Genome Sciences, Case Western Reserve University School of Medicine, Cleveland (USA); 3) Department of Bioinformatics, Erasmus University Medical Center, Rotterdam (NL); 4) Department of Pathology, Erasmus University Medical Center, Rotterdam (NL); 5) Department of Internal Medicine, Erasmus University Medical Center, Rotterdam (NL); 6) Trinity Biomedical Sciences Institute, Trinity College Dublin, Dublin (IE); 7) Department of Medical Genetics, Catholic University of Rome, Rome (IT).

We describe a novel syndromic form of acroosteolysis (AOL, progressive osteolysis of hand and foot phalanges associated with recurrent ulcers of the fingers and soles), inherited as an autosomal recessive trait and characterized by intellectual disability, sensory neuropathy, hypogonadism and recurrent infections. Whole-genome sequencing, combined with SNP haplotyping of an affected family identified two variants residing in neighboring genes in 11p15.4 (*TMEM41B* and *DENND5A*) and resulting in two missense mutations (D107G and R772Q) of the respective proteins. Bioinformatic analysis of the two variants characterized the D107G variant in *TMEM41B* as likely pathogenic, while the R772Q variant in *DENND5A* is likely benign. *TMEM41B* encodes a SNARE-associated Golgi protein with a potential role in synapsis formation. This is the first reported association of variants of this gene with human disease. *DENND5A* encodes a DENN-domain-containing protein that functions as a RAB-activating guanine nucleotide exchange factor (GEF); it is also known as RAB6-Interacting Protein 1 (RAB6IP1) as it is recruited by RAB6 onto Golgi membranes. Frameshift mutations of *DENND5A* are associated to Early Infantile Epileptic Encephalopathy 49 (EIEE49). Using microscale thermophoresis of the wildtype and mutant *DENND5A*, we did not detect any differences in the ability to bind RAB6, potentially confirming the benign nature of this missense variant. Our experiments further showed that *TMEM41B* mRNA is markedly upregulated in patient-derived osteoclast cell cultures, while patients' blood mononuclear cell-derived osteoclasts exhibit significantly enhanced bone resorption activity, suggesting a causative role of the *TMEM41B* missense variant. In order to fully characterize the contribution of both the *TMEM41B* and *DENND5A* to the phenotype, we are using genome editing to generate induced Pluripotent Stem Cells (iPSCs) containing either one of the two homozygous mutations, or both. These iPSCs will then be differentiated in osteoclasts in order to investigate the bone resorption phenotype. At the same time, we will differentiate the mutant iPSCs to cortical neurons in order to elucidate each gene's contributions to the neuronal phenotype. Our study provides the first description of a novel disorder with a potential causative role for the *TMEM41B* gene. Our investigations will also help elucidate the physiological role of the *TMEM41B* protein in human disease for the first time.

2384F

Mitochondrial DNA variants and headache. S. Børte^{1,2}, B.S. Winsvold^{2,3}, L.G. Fritsche^{4,5}, I. Surakka⁶, J.B. Nielsen⁶, W. Zhou⁷, C.J. Willer^{6,7,8}, J.A. Zwart^{1,2,3}. 1) Research and Communication Unit for Musculoskeletal Health, Division of Clinical Neuroscience, Oslo University Hospital, Ullevål, Oslo, Norway; 2) Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway; 3) Department of Neurology, Oslo University Hospital, Oslo, Norway; 4) HUNT Research Centre, Department of Public Health and General Practice, Norwegian University of Science and Technology, 7600 Levanger, Norway; 5) Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, MI 48109, USA; 6) Department of Internal Medicine, Division of Cardiology, University of Michigan Medical School, Ann Arbor, Michigan, 48109, United States of America; 7) Department of Computational Medicine and Bioinformatics University of Michigan, Ann Arbor, Michigan, 48109, United States of America; 8) Department of Human Genetics, University of Michigan Medical School, Ann Arbor, Michigan, 48109, United States of America.

Background: Mitochondria are, through oxidative phosphorylation, the main source of energy in the cells. Even though mutations in mitochondrial DNA and mitochondrial dysfunction have been implied in migraine no mitochondrial-wide analyses have been performed on any headache type yet. We aimed to explore whether there are associations between mitochondrial DNA variants and different headache diagnoses. **Methods:** We used a sample of 57,924 genotyped individuals from the population-based Nord-Trøndelag Health Study (HUNT) from Norway. Headache diagnoses were obtained by questionnaires. Genotyping was performed using Illumina HumanCoreExome chip, including 369 mitochondrial variants. In the first approach, we called and imputed the variants as male X chromosome, using 2,202 sequenced individuals from the same HUNT population as reference panel. In a second approach, we analysed frequency of the B-allele, to account for heteroplasmy. In a third approach, we assigned and analysed major European haplogroups. **Results:** No single mitochondrial variants or any of the haplogroups were significantly associated with any headache type. **Conclusion:** No single mitochondrial variants or haplogroups were associated with any headache types, indicating that inherited mitochondrial mutations do not play a major role in the major primary headache diagnoses.

2385W

Variational autoencoder identifying clinically meaningful latent features in multiple sclerosis bases on blood gene expression and genetically regulated gene expression across tissue-types. L.J. Lancashire¹, A.C. Provost¹, P.L. De Jager², M. Haas¹, N.P. Daskalakis¹. 1) Cohen Veterans Bioscience, Inc., Cambridge, MA; 2) Columbia University.

Multiple Sclerosis (MS) is clinically heterogeneous and it has been challenging to predict progression or treatment response. A study of MS relapse included 362 patients that were untreated or treated with one first-line treatment at the time of blood sampling. Whole blood genome-wide gene expression was available together with genome-wide genotype data, which were used to predict genetically regulated gene expression using PrediXcan transcription imputation (average 5000 imputed genes across 48 tissues). A Variational autoencoder (VAE) was applied to the observed blood and imputed transcriptome based on genotype. VAE is a deep generative model that allows for automatic engineering of non-linear features while learning a reduced dimensional manifold capturing the gene expression in a way that can be interpreted. The most variable genes were selected by median absolute deviation. These input genes were encoded to 100 features and then decoded back to the original number of variable genes. Models were trained in the encoding stage using batch normalization, the Adam optimizer and rectified linear units. During the decoding stage a sigmoid activation function was used. We investigated state transitions from the latent dimensions of the model to determine the extent to which the approach learned a manifold representation that could be interpreted from a biological perspective. The latent features were able to capture i) patient gender, ii) previously discovered subtypes of MS (MSa/MSb) and interestingly were able to iii) stratify the cohort into distinct groups representing time to relapse. All of this was achieved in an unsupervised manner, meaning that the phenotypic information was not used in the modelling process. Further, the trained models weights were investigated to determine the contribution of the genes to each learned node in the encoded layer. Genes that contributed greater than 2 standard deviations above or below the mean within the nodes contributing to the three phenotypes above were selected for pathway enrichment using gene ontology terms. Here, the trained model captured relevant biological patterns and pathways. For example, genes driving gender were located on sex chromosomes, and those driving MS subtypes were overrepresented in immune signalling pathways, a finding previously reported in these groups. VAE show promise in identifying lower representations of high dimensional biological data and are interpretable.

2386T

Rare variant analysis of efavirenz central nervous system effects using whole exome sequencing data. Y. Veturi¹, D.W. Haas^{2,3}, Y. Bradford⁴, A. Verma⁵, S.S. Verma¹, J.J. Eron⁴, R.M. Gulick⁶, S.A. Riddler⁶, P.E. Sax⁷, E.S. Daar⁸, G.D. Morse⁹, E.P. Acosta¹⁰, M.D. Ritchie¹¹, *AIDS Clinical Trial Group*.

1) Department of Genetics, University of Pennsylvania, Philadelphia, PA., USA; 2) Department of Medicine, Vanderbilt University School of Medicine, Nashville, TN, USA; 3) Department of Internal Medicine, Meharry Medical College, Nashville, TN, USA; 4) Department of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; 5) Department of Medicine, Weill Cornell Medicine, New York City, NY, USA; 6) Department of Medicine, University of Pittsburgh, Pittsburgh, PA, USA; 7) Division of Infectious Diseases, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; 8) Department of Medicine, David Geffen School of Medicine at UCLA, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA, USA; 9) School of Pharmacy and Pharmaceutical Sciences, University at Buffalo, SUNY, Buffalo, NY, USA; 10) Department of Pharmacology and Toxicology, University of Alabama at Birmingham, Birmingham, AL, USA; 11) Penn Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, PA.

The anti-HIV drug efavirenz (EFV) commonly causes central nervous system (CNS) adverse effects (AEs). Analyses included patients who were randomized to EFV-containing regimens in four AIDS Clinical Trial Group (ACTG) protocols and provided genetic consent. Cases had new onset grade ≥ 2 (moderate or greater) EFV-consistent CNS AEs within 48 weeks of study entry while receiving EFV. Controls had no EFV-consistent CNS AEs while receiving EFV for ≥ 96 weeks. Whole exome sequencing data was collected from 310 individuals (70 cases, 150 controls) using Agilent Sure Select Human All Exon kit V6. We used BioBin software to bin rare variants (minor allele frequency $< 5\%$) by genes and tested for association between CNS AEs and binned variants using burden-based (logistic regression) and dispersion-based (sequence-kernel association test) methods with Madsen-Browning weighting. We adjusted for effects of age, sex, race and first 5 ancestry-derived principal components. We also applied different filters to bin rare variants based on functional annotation to detect loss-of-function and predicted deleterious non-synonymous variants using different scoring algorithms: SIFT 5.2.2, Polyphen 2, likelihood ratio test and MutationTaster. Fifteen genes known to be highly to moderately expressed in brain (based on Genotype-Tissue Expression Project portal) were nominally associated with EFV-consistent CNS AEs, including *DNER* (delta/notch-like epidermal growth factor repeat) on chromosome 2 (logistic regression p-value = $4.9E-5$), *NECAB2* (N-terminal EF-hand calcium binding protein 2) on chromosome 16 (SKAT p-value = $1.6E-4$), and *NKA1N4* (sodium/potassium transporting ATPase interacting 4) on chromosome 20 (SKAT p-value = $3.4E-4$). None were statistically significant after correcting for multiple testing. Our findings suggest possible novel associations with EFV CNS AEs.

2387F

The genetics of multiple sclerosis in the Northern Isles of Scotland.

C.L.K. Barnes¹, P.K. Joshi¹, T.S. Boutin², C. Hayward², H. Campbell², J.F. Wilson^{1,2}. 1) Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, UK; 2) MRC Human Genetics Unit, University of Edinburgh, Institute of Genetics and Molecular Medicine, Western General Hospital, Edinburgh, UK.

The prevalence of multiple sclerosis (MS) in the Northern Isles of Scotland is among the highest in the world, with Shetland having 229 cases / 100,000 individuals and Orkney having a significantly higher prevalence of 402 cases / 100,000 individuals¹. Although the environmental and genetic influences for MS are largely established, the explanation for this excess prevalence of MS in the Northern Isles, particularly Orkney, is not clear. Previous investigations into possible influences, including vitamin D deficiency² and homozygosity³ in Orkney, have been unable to provide an explanation. In 1981, Compston suggested⁴ that Orkney may have a higher frequency of common risk variants due to the founders having higher frequencies of these variants by chance. We aimed to explore this suggestion by comparing the common risk variants in Shetland and Orkney to those in mainland Scotland to assess if there was a difference in common risk variant frequencies, and if so, how that translated into disease prevalence. We calculated the frequency of common risk variants and polygenic risk scores (PGRS) in non-MS cases in Orkney and Shetland (using the ORCADES and VIKING cohorts) and compared these to non-MS cases mainland Scotland (using a subset of the Generation Scotland cohort). A logistic regression model was then used to assess the relationship between PGRS and MS. We found there was a significant difference between the mean PGRS in Orkney, Shetland and mainland Scotland, however this did not translate into a significant "real world" difference in prevalence that would be expected if common risk variants were the cause of the excess prevalence of MS in Orkney. Additionally, we investigated if Orkney and Shetland had any unique, undiscovered common risk variants by conducting a genome-wide association study in the populations, however no unique variants were discovered. This suggests that another cause, possibly rare variants, are the cause of excess prevalence of MS in Orkney. **References** 1. Visser, E. M. et al. (2012) A new prevalence study of multiple sclerosis in Orkney, Shetland and Aberdeen city. *J. Neurol. Neurosurg. Psychiatry*, **83**, 719–24. 2. Weiss, E. et al. (2016) Farming, Foreign Holidays, and Vitamin D in Orkney. *PLoS One*, **11**, e0155633. 3. McWhirter, R. E. et al. (2012) Genome-wide homozygosity and multiple sclerosis in Orkney and Shetland Islanders. *Eur. J. Hum. Genet.*, **20**, 198–202. 4. Compston, A. (1981) MULTIPLE SCLEROSIS IN THE ORKNEYS. *Lancet*, **318**, 98.

2388W

Blood-based RNAs distinguish cases from controls and indicate *C9orf72* carrier status in amyotrophic lateral sclerosis (ALS). M.T.W. Ebbert¹, M. Prudencio¹, Y. Song¹, B. Oskarsson¹, J. Wu², R. Rademakers¹, M. Benatar², J. Fryer¹, L. Petrucelli¹. 1) Neuroscience, Mayo Clinic, Jacksonville, FL; 2) Department of Neurology, Miller School of Medicine, University of Miami, Miami, FL.

Background. Neurodegenerative diseases like amyotrophic lateral sclerosis (ALS), Alzheimer's disease, and others are particularly challenging to address clinically because, once symptoms onset, the disease has ravaged affected tissues beyond the body's ability to function normally. Identifying blood-based biomarkers that distinguish cases and controls is an ideal starting point for identifying biomarkers that track with disease longitudinally. We have identified >300 RNAs in whole-blood across 59 cases and 30 controls that distinguish ALS cases from controls, and indicate *C9orf72* (C9) carrier status. **Methods.** We performed total RNA sequencing from whole blood across 29 C9 expansion carriers (C9+), 30 non-carriers (C9-), and 30 controls after rRNA and hemoglobin depletion. Individuals were clinically diagnosed with ALS, and were negative for *SOD1*, *TARDBP*, and *FUS* mutations. Groups were 55%, 53%, and 47% female for C9+, C9-, and controls, respectively, with median ages of 61, 61, and 56 years. Data were quality checked, aligned using STAR (hg38), and tested for differential expression (cases and controls) using DESeq2 after conditional quantile normalization. Similar analyses between C9+ and C9- were performed. Statistical models controlled for sex and age at collection. Collection site ($p=0.99$) and RNA RIN ($p=0.42$) were not statistically significant. **Results.** There were >300 (FDR 0.1) coding and non-coding differentially-expressed RNAs between cases and controls, and several between C9+ and C9-. Cases and controls were significantly different by hierarchical clustering (SigClust; $p=1.48e-4$) with a positive-predictive value of 86.5%. Combining RNAs with an AUC ≥ 0.8 ($n=8$), we attained $\geq 90\%$ sensitivity/specificity using a linear model (AUC=97%). For C9 status, we attained $\geq 85\%$ sens/spec (AUC=92%). Approx. 81% of cross-validation (100x, 20-fold) results had $\geq 50\%$ sens/spec, where 40% of results had 100% sens/spec. Top phenotype and gene ontology terms were statistically ($7.257e-11 < p \leq 0.01$) and biologically significant, including ontology terms specific to documented ALS phenotypes and pathways. **Conclusions.** While effective therapies are essential, a pre-symptomatic disease diagnostic is equally critical to intervene before symptomatic neurological damage occurs. Our results may have important implications for future ALS clinical diagnostics and therapy, and potentially other neurodegenerative diseases, by allowing earlier diagnosis.

2389T

Enrichment of rare protein truncating variants in patients with amyotrophic lateral sclerosis. S.M.K. Farhan^{1,2}, D.P. Howrigan^{1,2}, L.E. Abbott^{1,2}, A.E. Byrnes^{1,2}, C. Churchhouse^{1,2}, M.J. Daly^{1,2}, B.M. Neale^{1,2}. 1) Massachusetts General Hospital, Harvard Medical School, Boston, MA; 2) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard.

Amyotrophic lateral sclerosis (ALS) is a genetically and phenotypically heterogeneous neurodegenerative disease. Genetic variation is an important risk factor for ALS. Known familial ALS gene mutations are found in 5-10% of patients reporting a positive family history but also in ~10% of sporadic patients, further eliminating the distinction between familial and sporadic disease. To search for novel genetic risk factors underlying ALS, we aggregated exomes from 3,864 cases and 7,839 ancestry matched controls, generating the largest ALS exome-based case-control study to date. We complemented our analysis by leveraging allele frequencies from large external exome sequencing databases such as DiscovEHR (>50,000 samples) and the non-brain related subset of ExAC (>45,000 samples). We observed a significant excess of ultra-rare and rare protein-truncating variants (PTV) among ALS cases, which was primarily concentrated in genes intolerant of truncating mutation (constrained genes); however, a significant enrichment in PTVs does persist in the remaining exome. There was no enrichment of rare PTVs in genes known to confer risk to ALS, genes associated with clinically overlapping diseases, or genes in which their expression is specific to the brain. We conducted gene level analyses, which highlighted three known ALS genes (*SOD1*, *NEK1*, and *FUS*) as the most strongly associated. We also observe suggestive statistical evidence for multiple novel genes including *DNAJC7*, which is a highly constrained gene and a member of the heat shock protein family (HSP40). HSP40 proteins, along with HSP70 proteins, facilitate protein maintenance and quality control, such as folding of newly synthesized polypeptides, and clearance of degraded proteins. When these processes are not regulated, misfolding and accumulation of degraded proteins can occur leading to aberrant protein aggregation, one of the pathological hallmarks of neurodegenerative diseases. This work reveals new insights in to the underlying genetic risk factors for ALS that would not be possible if conducting each study individually, and hints to novel signals that may be further queried as more data are collected in the future.

2390F

Role of uric acid in multiple sclerosis risk and progression. *M. Montierth¹, J.C Denny², M.F. Davis¹*. 1) Department of Microbiology and Molecular Biology, Brigham Young University, Provo, UT; 2) Departments of Biomedical Informatics and Medicine, Vanderbilt University Medical Center, Nashville, TN.

Multiple sclerosis (MS) is a severe neurodegenerative that affects approximately 400,000 Americans. MS is also quite complex, with many genetic and environmental factors affecting disease risk and pathogenesis. Among many other factors, studies have indicated that oxidative stress plays a key role in the pathogenesis of MS. Uric acid, a natural scavenger of peroxynitrite, has been suggested as having a protective effect against MS. However, previous studies of uric acid in MS are affected by many confounding factors, and as such it is still difficult to draw conclusions regarding this association. In this study, we compared uric acid lab values between patients with MS and control patients in a large population of electronic medical records to better understand this difference. Additionally, Mendelian randomization studies can be used to lend power to epidemiological studies by using genetic variants as instrumental variables, and thus avoiding potential biases of an observational study. We performed a Mendelian randomization analysis to investigate the link between MS and uric acid. MS cases and controls were selected from the Vanderbilt University Medical Center's Synthetic Derivative (SD) database. The SD is composed of de-identified medical records used specifically for research purposes and includes more than 2 million individuals. Using lab values from 3803 individuals (500 cases, 3303 controls) with available uric acid tests, we determined that the mean uric acid level was significantly lower in patients with MS than in the control population (5.308 vs. 6.314, $p=1.64E-13$). From the SD database, 1,625 cases and 6,187 controls were genotyped. In these individuals, we used 37 SNPs previously associated with uric acid levels to generate a weighted genetic risk score (wGRS) for each individual. We then regressed MS case status on the uric acid wGRS. Preliminary results show that alleles associated with higher uric acid levels were associated with decreased MS risk. A more powerful follow up analysis will better characterize this relationship. We also created an MS wGRS using 110 published SNPs associated with MS and regressed uric acid lab values on the MS wGRS. This regression analysis found MS wGRS to not have a significant influence on serum uric acid level ($p=0.674$). In conclusion, we present evidence that suggests that SNPs associated with higher uric acid levels have a protective effect against MS.

2391W

Genetic study of longitudinal brain atrophy in Multiple Sclerosis shows susceptibility loci are not influencing progression. *P.G. Bronson¹, A.R. Pine¹, F.B.S. Briggs², A. Dhroso¹, B. Fisher¹, S. John¹, K. Estrada¹*.

1) Statistical Genetics & Genetic Epidemiology, Biogen, Inc, Cambridge, MA 02142; 2) Department of Population and Quantitative Health Sciences, School of Medicine, Case Western Reserve University, Cleveland, OH 44106; 3) Value Based Medicine, Biogen, Cambridge, MA 02142; 4) Translational Biology, Biogen, Cambridge, MA 02142.

Multiple sclerosis (MS), a demyelinating autoimmune disease, affects 2.5M people. Susceptibility genetics has identified >100 loci, but we have a limited understanding of how these risk loci influence disease progression. There have been few large studies of MS progression, and none have used well-annotated longitudinal data. Identifying genetic predictors of progression is important for drug discovery and treatment decision-making. We conducted a GWAS of longitudinal brain atrophy in 3,446 MS cases collected in 9 randomized clinical trials. Cases were relapsing remitting at onset (15% secondary progressive; 68% female; mean age 38 years). We examined three additional traits: expanded disability status scale (EDSS; N = 4,352), nine-hole peg test (9HPT; N = 1,890) and timed 25-foot walk (T25FW; N = 1,006). All traits were analyzed by annualizing the difference between first and last time point (mean time range = 88 weeks; range 1 to 3 years). For 9HPT and T25FW, we compared the extremes of the annualized distributions (top vs. lowest quartile). There were no genome-wide significant results. We identified 99 regions with $P < 1 \times 10^{-5}$ (43 PBVC, 32 EDSS, 13 9HPT, 11 T25FW); none were in the human leukocyte antigen (HLA) region. 12/111 susceptibility SNPs and 2/40 candidate progression SNPs had $P < 0.05$ in our study, with same direction of effect. The two candidate progression SNPs with $P < 0.05$ in our study were *RELN* (rs10487166*T; EDSS beta = -0.03, $P = 0.028$) and *CRTAC1* (rs11189446*T; PBVC beta = -0.049, $P = 0.014$). Both *RELN* and the alternate transcript of *CRTAC1* encode secreted extracellular matrix (ECM) proteins exclusive to the brain. The ECM is a functional, dynamic scaffolding that fills extracellular space, accounting for 15% of brain volume. In animal models, *RELN* overexpression boosts synaptic contacts and dendritic spine size. *CRTAC1* promotes axon growth by blocking interaction between a receptor for axon growth inhibitors (NRG1) and all four of its ligands (RTN4, MAG, OMG, TNFSF13B). In conclusion, this is the first large GWAS of progression to use well-annotated longitudinal data. Our study does not show an effect of susceptibility loci on progression. Interestingly, we have sufficient power to exclude HLA from influencing progression; in rheumatoid arthritis HLA is a predictor of progression. There was modest association with two candidate progression SNPs in *RELN* and *CRTAC1*, highlighting a potential role for these ECM proteins in MS progression.

2392T

Prioritization of causal variants possibly associated with hippocampal volume using a massively parallel reporter assay. Y. Cooper¹, J. Davis², A. Jasinska³, S. Service², S. Kosuri^{3,4}, N. Freimer^{1,2}, G. Coppola^{2,5}. 1) Department of Human Genetics, University of California, Los Angeles, Los Angeles, CA, USA; 2) Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, Department of Psychiatry and Biobehavioral Sciences, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA; 3) Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, CA, USA; 4) UCLA-DOE Institute for Genomics and Proteomics, Molecular Biology Institute, Quantitative and Computational Biology Institute, Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, Jonsson Comprehensive Cancer Center, University of California, Los Angeles, Los Angeles, CA, USA; 5) Department of Neurology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles CA, USA.

The hippocampus is a conserved vertebral brain structure critical for episodic memory, cognition, and spatial reasoning. Hippocampal volume has been associated with a variety of neuropsychiatric disorders including Alzheimer's disease, schizophrenia, and bipolar disorder, and is highly heritable, with heritability estimates between 0.5-0.9. Therefore a better understanding of the genetic underpinnings of this complex trait is warranted. Our group previously identified a locus in a large pedigree of Caribbean vervet monkeys associated with both hippocampal volume and the quantitative expression (eQTL) of two long non-coding RNAs. However, this locus spans an extended region of linkage disequilibrium (LD) and contains thousands of variants tightly correlated with the three lead single nucleotide polymorphisms (SNPs) identified. We therefore sought to fine-map this locus to prioritize variants for subsequent downstream modeling and functional validation. We utilized a Massively Parallel Reporter Assay (MPRA) which allows for high-throughput interrogation of the effects of genetic variation on gene transcription. We assayed 1000 variants prioritized by LD with our lead SNPs, ultimately narrowing down our search space to 20 putative causal variants. We are validating these findings in vervet IPSC-derived neurons using a luciferase assay to determine the effect of these variants in neurons. These results indicate that MPRA is a suitable approach for the fine-mapping of eQTL loci and prioritization of variants in genome sequencing studies. We plan to model *in vitro* the effects of these putatively causal variants to interrogate novel molecular mechanisms regulating hippocampal volume.

2393F

Associations of genetic risk variants for cognitive impairment and for Parkinson's disease (PD) with clinical features in PD. A. Deutschlander^{1,2}, T. Konno², A. Soto¹, R. Walton¹, A. Strongosky², M. Heckmann³, N. Diehl³, J. VanGerpen², R. Uitti², O. Ross¹, Z. Wszolek². 1) Human Genetics, Mayo Clinic Florida, Jacksonville, FL; 2) Neurology, Mayo Clinic Florida, Jacksonville, FL; 3) Division of Bioinformatics, Mayo Clinic Florida, Jacksonville, FL.

It was the aim of this study to analyse associations of genetic risk variants for cognitive impairment and for Parkinson's disease (PD) with clinical motor and non-motor features in PD including calculation of genetic risk scores. A large clinical heterogeneity is observed in PD (motor and non-motor features). Studies on associations of genetic susceptibility variants with PD features are limited and mostly analysed variants in only one gene (e.g., *SNCA*, *LRRK2*, *MAPT*, *GBA*). We included 862 Caucasian patients (550 male, mean AAO: 64 yrs) with PD, who had demographic and detailed clinical data available. Patients were classified into 4 distinct clinical subtypes based on their prominent motor feature. Genotyping was performed for 33 independent PD risk variants as identified in GWAS for PD and for several risk variants for cognitive impairment. Frequencies of variants were compared for motor and non-motor signs, subtypes, rate of progression, survival, and early cognitive decline. The tremor-dominant subtype was found most frequently, followed by akinetic-rigid, mixed, and gait difficulty subtypes. Orthostatic hypotension, rapid progression, visual hallucinations, resting tremor and REM sleep behavior disorder showed significant associations with PD risk variants. A polygenic risk score further correlated with PD features (e.g. age at onset of PD). The APOE4 genotype was significantly associated with risk for dementia; we present further data on genetic risk variants. Genetic susceptibility variants account for some of the clinical heterogeneity observed in PD in our large single site patient cohort. Clinicogenetic studies can provide a better understanding for this variability including parameters for prognosis. Further, a homogenous patient cohort, which takes genetic underpinnings of PD into account, is likely to be of importance for conducting clinical trials.

2394W

FUS p.P525L de novo mutation in sporadic juvenile ALS with aggressive progression and developmental delay. O. Goldstein¹, M. Gana-Weisz¹, B. Nefussy², S. Twito¹, B. Vainer², V.E. Drory^{2,3}, A. Orr-Urtreger^{1,3}. 1) The Genetic Institute, Tel Aviv Sourasky Medical Center, 6 Weizmann Street, Tel Aviv 64239, Israel; 2) Neuromuscular Diseases Unit, Department of Neurology, Tel Aviv Sourasky Medical Center, 6 Weizmann Street, Tel Aviv 64239, Israel; 3) Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel.

Amyotrophic lateral sclerosis (ALS) is a complex and heterogeneous neurodegenerative disease, characterized by loss of upper and lower motor neurons. Juvenile ALS is characterized by age at disease onset (AAO) before 25 years, with several genes identified as causative. More often than in non-juvenile ALS, these patients have a slower progression rate, and milder symptoms. Trio exome-sequencing (WES) of two unrelated juvenile ALS male patients with AAO at 18 and 19 years was done. Both patients had a fulminant disease that led to tracheostomy after only six months. They are still alive, locked-in and ventilated (for 1.5 and 5 years). None had history of ALS in their family, originated from mixed Ashkenazi-Balkan and Ashkenazi-North-Africa Jewish families. Both had a history of moderate learning and behavioral disabilities, and attended special classes for mentally disabled children, although functioned borderline well within a restricted social network. Trio analyses with 109X mean X-coverage included *de novo*-, rare-homozygosity-, and compound- heterozygosity-analyses. We identified the same *de novo* mutation in the Fused In Sarcoma (*FUS*) gene, p.P525L, in both. A TaqMan genotyping assay was used to genotype 3 ALS cohorts: early-onset (AAO $\leq$ 40) and AAO >40 years from Ashkenazi and North-Africa Jewish origins (61, 320, and 94, respectively). Carriers were confirmed by Sanger sequencing. Among 8 other juvenile ALS patients, one carried the same mutation. This sporadic male patient had an AAO at 24 years and required tracheostomy after eight months. His past medical history was remarkable for mild mental retardation and need for a special school for mentally disabled. Paternal samples were not available, but both parents were neurologically and mentally unremarkable at the time of their son's diagnosis. All other ALS patients tested (467 with AAO >25 years) were negative for this mutation, suggesting full penetrance at an early age. Interestingly, all three patients reported learning disabilities/mild mental retardation, supporting previous observations of a possible connection between mutations in *FUS* and early life cognitive dysfunction. Our findings suggest that the *FUS* p.P525L mutation should be screened in ALS patients with early AAO, aggressive disease course and sporadic occurrence, especially when ALS is accompanied by developmental delay.

2395T

Elucidating the role of genetic variants in known ALS-associated genes by means of targeted NGS analysis. G. Marangi¹, S. Lattante¹, P.N. Doronzio¹, S. Frangella¹, A. Conte², A.K. Patanella², G. Bisogni², M. Zollino¹, M. Sabatelli^{2,3}. 1) Institute of Genomic Medicine, Catholic University, Rome, Rome, Italy; 2) NEuroMuscular Omnicentre (NEMO), Serena Onlus Foundation, Policlinico universitario A. Gemelli; Foundation, Rome; 3) Institute of Neurology, Catholic University, Rome, Italy.

Amyotrophic lateral sclerosis (ALS) is a progressive, fatal neurodegenerative disease, whose etiology is not fully understood yet, but for which genetic variants in several genes have already been identified as causal factors in a consistent proportion of both familial and sporadic cases. The number of ALS-associated genes has been steadily increasing in recent years. Consequently, simultaneous screening of multiple genes is likely to be more efficient than gene-by-gene testing in order to identify possible causative variants in ALS patients. We designed a custom panel of 40 genes (including 5' and 3' UTRs) claimed to be associated with ALS and performed targeted next-generation sequencing in 500 consecutive Italian ALS patients (including about 10% familial and 90% sporadic cases) and 200 controls. We evaluated the following aspects of the results: 1) frequencies and relative impact of different categories of variants, including known pathogenic variants (found in 6% of cases), novel coding variants (12%), and variants with MAF <0.1% (28%); 2) frequencies and relative impact of variants in known genes, thus showing that patients were enriched for rare variants in previously neglected genes (i.e.: *OPTN*, *EPHA4*, and *MATR3* among the others); 3) co-occurrence of variants in specific genes; 3) association of certain variants with specific ALS phenotypes; 4) frequencies of variants in 5' and 3' UTRs of ALS-associated genes. Our panel has been proving to be a useful tool for evaluating and defining the relative contribution of genes, the effective role of specific coding variants and the possible involvement of variants in 5' and 3' UTRs in ALS. Furthermore, the integration of all genetic data we obtained allowed us to explore the oligogenic hypothesis in ALS etiopathogenesis, considering that we found a higher frequency than expected of patients carrying two or more rare variants in the studied genes.

2396F

GWAS of broadly defined 'self-reported migraine' identify risk loci for tension-type headache. *D.R. Nyholt.* Queensland University of Technology, Brisbane, Australia.

Migraine is one of the world's most prevalent, disabling, misdiagnosed, undertreated and underfunded disorders, and has an enormous socio-economic burden. The diagnostic criteria of the International Headache Society (ICHD) describe the two major varieties of migraine: i) *migraine without aura* (MO), and ii) *migraine with aura* (MA). MO is defined by at least five lifetime headache attacks (relaxed to two attacks for MA) lasting 4–72 hours, with at least two characteristics: unilateral location, pulsating quality, moderate or severe pain intensity, or aggravation by or causing avoidance of routine physical activity (e.g., walking or climbing stairs); and at least one of 'nausea and/or vomiting' or 'photophobia and phonophobia'. The most common MA requires headache, with or without MO characteristics, plus one or more aura symptoms of the following types: visual, sensory, speech and/or language, motor, brainstem, or retinal. Tension-type headache (TTH) is the most common form of headache and is defined by at least ten headaches lasting 30 minutes to 7 days, with at least two characteristics: bilateral location, pressing or tightening (non-pulsating) quality, mild or moderate intensity, or not aggravated by routine physical activity; and neither 'nausea and/or vomiting' and 'photophobia and phonophobia'. In 2016, the largest published genome-wide association study (GWAS) for migraine analysed 59,674 migraine cases and 316,078 controls, and reported 44 (34 new) risk variants for migraine. Subsequent tissue expression enrichment analysis of genes within 50 kb of credible SNP sets (i.e., having 99% probability of containing the causal SNP at each migraine locus) identified that vascular tissues were significantly enriched for expression of migraine-associated genes. Whereas the previously largest migraine GWAS of 23,285 migraine cases and 95,425 controls (published in 2013) utilised a majority (60%) of cases fulfilling ICHD criteria, the 2016 GWAS utilised more cases with broadly defined self-reported migraine status—in particular GWAS data from 23andMe contributed 30,465 self-reported cases which made up 51% of all migraine cases. Analysis of UK Biobank data indicates migraine GWAS utilising broadly defined 'self-reported migraine' identify risk loci for 'non-migraine' (TTH) headache. These findings indicate heterogeneous (and poor) diagnostic specificity has important implications for the interpretation of findings from large migraine GWAS.

2397W

Tissue specific gene enrichment analysis reveals aberrant physiological regulations in ischemic stroke. *H.N. Singh.* CSIR-Institute of Genomics & Integrative Biology, Delhi, Delhi, India.

Background: Ischemic stroke is a heterogeneous multifactorial complex neurological disorder. Treatment options are inadequate due to limited understanding of mechanisms that mediate neuronal injury. Therefore, pathway based genome wide expression analysis could be significant to understand the disease etiology and development of future medications. **Aim:** To assess expression patterns of groups of genes that share common biological function, chromosomal location, or regulation as well as identification of potential therapeutic targets. **Materials & Methods:** The microarray dataset GSE22255 was retrieved from the NCBI/GEO database, which includes mRNA expression data of peripheral blood mononuclear cells of 20 controls and 20 ischemic stroke patients. The bioconductor "affy" package was used to calculate mRNA expression and further pairwise ttest was applied to find the significant differential expression ($p < 0.01$). Gene Set Enrichment Analysis was performed on deregulated genes and further drug targets were identified using Reactome FI Cytoscape Tool. **Results:** Out of 117 genes, majority of genes were found to be upregulated (67.5%) and showed 68.6% functional enrichment with disease phenotypes (Figure 1A). However, 32.5% genes were upregulated with 31.4% correlation in controls (Figure 1B). Additionally, homeostatic process ($ES = -0.31$; $NES = -0.89$; $NOM\ p\text{-val} = 0.602$) and cell development pathway ($ES = -0.31$; $NES = -0.75$; $NOM\ p\text{-val} = 0.813$) were identified as most dysregulated in the disease conditions (Figure 2B). The GSEA analysis suggested three markers (*SLC18A2*, *SLC12A5*, *RHAG*) from homeostatic process and five markers from cell development pathway (*ETV5*, *PARVA*, *SLC12A5*, *RHAG*, *BMPRI1B*, *ROBO2*) are significant for the disease. Further, *SLC12A5* and *RHAG* were found as potential druggable target for stroke. **Conclusions:** Pathway based expression and gene set enrichment analysis showed cellular homeostasis and cell development pathways are mostly affected in ischemic stroke conditions. Seven markers and two druggable targets were identified, however further experimental validations are required.

2398T

Mutations in *VPS13C* in a Belgian cohort of Parkinson disease patients.

S. Smolders^{1,2}, D. Crosiers^{1,2,3}, P. Cras^{2,3}, C. Van Broeckhoven^{1,2}, BELNEU Consortium. 1) VIB Center for Molecular Neurology, University of Antwerp, Antwerp, Belgium; 2) Institute Born-Bunge, University of Antwerp, Antwerp, Belgium; 3) Department of Neurology, Antwerp University Hospital, Edegem, Belgium.

Background *VPS13C* was first identified as a novel risk gene for Parkinson Disease (PD) in a genome-wide association meta-analysis for PD (Nalls et al. Nature Genetics, 2014) and later as the fourth gene causative for autosomal recessive early onset PD (Lesage et al. Am J Hum Genet, 2016), alongside *PARK2*, *PINK1* and *PARK7*. **Materials and Methods** Targeted resequencing of all 86 exons and exon-intron boundaries of *VPS13C* was performed in an extended Belgian PD cohort of 641 unrelated patients (mean onset age 62.8 ± 11.5 years; 42.7 % female), 23.6 % of whom exhibit a familial history of PD. We focused on rare variants with an apparent impact on functional protein levels i.e. non-synonymous coding and splice site variants. **Results** We identified 1 patient with a homozygous missense variant, 1 patient with a homozygous splice site variant and 6 patients carrying 2 compound heterozygous missense variants in *VPS13C*. The overall mutation frequency in our Belgian PD cohort is 11.72 %. The homozygous p.A444P patient carrier had an onset age of 40 years and no known family history of PD. This p.A444P variant was present heterozygous in 2 other unrelated PD patients but absent in the Belgian control cohort of 1,181 unrelated, age- and geographically matched individuals. Since endogenous protein expression in lymphoblast cells of a heterozygous p.A444P carrier was reduced by 40 % (Philtjens et al., unpublished data), we expect a loss of protein expression by > 50 % in the homozygous carrier. A homozygous splice-site mutation, c.4166-8C>T, was identified in a patient with PD dementia at the age of 59 years. The effect on splicing of exon 37 is yet unknown. Of interest is the identification in a patient of heterozygous compound missense variants in *PARK2* and *VPS13C*, p.P437L/p.L239S, underscoring a role for multiple genetic factors in PD etiology. **Discussion** Our results confirmed *VPS13C* as a novel risk gene for PD. *Cis/trans* configuration for putative compound heterozygous mutations and the effect of novel splice site variants need to be investigated. Patient-derived biomaterials will be used to study the pathological relevance of the *VPS13C* mutations. Oligogenic inheritance of *PARK2/VPS13C* mutations could be explained by their essential roles in common mitochondrial quality control pathways.

2399F

Longitudinal small RNA and whole transcriptome sequencing for over 1,500 individuals across 4,600 whole-blood samples as part of the Parkinson's Progression Markers Initiative.

D.W. Craig¹, M.R. Cookson², I. Violich¹, E. Hutchins², E. Alsop¹, E.M. Sullivan¹, J. Mote¹, S. Levy¹, A. Keller², K. van Keuren-Jensen². 1) Department of Translational Genomics, University of Southern California, Los Angeles, CA; 2) Translational Genomics Research Institute, TGen; 3) National Institute of Aging, Bethesda, MA; 4) HudsonAlpha Institute, Huntsville, AL; 5) Saarland University, Center for Bioinformatics.

The Parkinson's Progression Markers Initiative (PPMI) is a landmark observational clinical study to comprehensively evaluate cohorts of significant interest using advanced imaging, biologic sampling and clinical and behavioral assessments to identify biomarkers of Parkinson's disease progression. We report on development of a comprehensive RNA resource from whole blood samples that can be accessed and utilized by the PD research community. Approximately 4,600 samples from 1,500 individuals enrolled within PPMI cohorts were expression profiled profiles using two assays. These data complement existing whole-genome sequencing data and other genomic profiles on the same subjects, and will provide the community with a resource for development of novel integrative analytical approaches. We detail overall quality and utility of the resource, specifically examining informativeness of these whole-blood samples towards a neurological phenotype. We detail quality control and a pilot study comparing multiple platforms for a series of covariates. We describe finally development of a portal allowing end-users to interactively explore different facets of the data, both allowing for comparisons across phenotypic and clinical metrics. Together this portal and rich multi-omic data resource provides researchers an unparalleled resource for development and discovery of biomarkers predicting progression and/or suggesting new disease-modifying therapies.

2400W

Genetic etiology and prediction of the pharmacotherapy response

status in epilepsy. C. Leu^{1,2}, A. Avbersek¹, D. Speed³, A. Ingason⁴, S. Peter⁵, R. Krause⁶, M. McCormack^{6,7}, N. Delanty^{6,8}, G.L. Cavalleri^{6,8}, J. Craig⁹, C. Depondt¹⁰, M. Johnson¹¹, B.P.C. Koeleman⁷, H. Lerche¹², T. Marson¹³, T. O'Brien^{14,15}, J.W. Sander^{1,16,17}, G.J. Sills¹⁸, P. Striano¹⁹, F. Zara²⁰, S.M. Sisodiya^{1,17}, *EpiPGX Consortium*. 1) Department of Clinical and Experimental Epilepsy, NIHR University College London Hospitals Biomedical Research Centre, UCL Institute of Neurology, London WC1N 3BG, UK; 2) Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, Cleveland, Ohio, United States of America; 3) UCL Genetics Institute, University College London, London, UK; 4) deCODE Genetics, Reykjavik, Iceland; 5) Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg; 6) Department of Molecular and Cellular Therapeutics, Royal College of Surgeons in Ireland, Dublin, Ireland; 7) Department of Medical Genetics, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands; 8) The FutureNeuro Research Centre, RCSI, Dublin, Ireland; 9) Department of Neurosciences, Acute and Unscheduled Care, Belfast Health and Social Care Trust, Belfast, UK; 10) Department of Neurology, Hôpital Erasme, Université Libre de Bruxelles, Brussels 1070, Belgium; 11) Division of Brain Sciences, Imperial College Faculty of Medicine, London, United Kingdom; 12) Department of Neurology and Epileptology, Hertie Institute for Clinical Brain Research, University of Tübingen, Germany; 13) Department of Molecular and Clinical Pharmacology, University of Liverpool, Division of Neurology, The Walton Centre NHS Foundation Trust, The Clinical Sciences Centre, Lower Lane, Liverpool, UK; 14) Department of Medicine and Neurology, Royal Melbourne Hospital and University of Melbourne, Parkville, Vic., Australia; 15) Departments of Neurosciences and Neurology, The Central Clinical School, The Alfred Hospital, Monash University, Melbourne, Vic., Australia; 16) Stichting Epilepsie Instellingen Nederland (SEIN), Heemstede, Netherlands; 17) Chalfont Centre for Epilepsy, Chalfont-St-Peter, Bucks, UK; 18) Department of Molecular and Clinical Pharmacology, Institute of Translational Medicine, University of Liverpool, Liverpool, United Kingdom; 19) Pediatric Neurology and Muscular Diseases Unit, Department of Neurosciences, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health, University of Genoa, Genoa, Italy; 20) Laboratory of Neurogenetics and Neuroscience, Institute G. Gaslini, Genoa, Italy.

Rationale: Pharmacoresistance and comorbid neurological disorders (NDs) are two key features of epilepsy, contributing to the negative impact of epilepsy on the quality of life, yet our understanding of the genetics behind both features is limited. Pharmacoresistance may be of genetic etiology, but the exact degree of genetic contribution is unknown, with no robust evidence for association with any single gene. Whilst genetic overlap between NDs and epilepsy has been quantified epidemiologically, the role of comorbid NDs to genetic analyses and treatment solutions in epilepsy remains elusive. **Methods:** To identify risk factors for pharmacoresistance, a genome-wide association study (GWAS) was performed in 2354 individuals with pharmacoresistant and 1580 with pharmacoresponsive epilepsy of European ancestry. The nucleotide polymorphism (SNP) based heritability of pharmacoresistance was estimated in the same sample using LDSC. A linear mixed model GWAS (as implemented in BOLT-LMM) was performed using SHAPEIT/IMPUTE2 imputed Illumina OmniExpress SNP-chip data. To evaluate the predictive power of genetic risk factors for comorbid NDs, we used published ND-GWAS summary statistics in PLINK and R, following best-practices for allelic scoring. **Results:** The GWAS did not reveal any genome-wide significant associations with pharmacoresistance in epilepsy. In the same sample, the estimated observed-scale single nucleotide polymorphism (SNP) heritability h^2_{SNP} was 0.22 (standard error=0.13). Among four comorbid NDs (Alzheimer's disease, schizophrenia, major depressive disorder, attention deficit hyperactivity disorder) and two traits (educational attainment and neuroticism), genetic risk factors for Alzheimer's disease significantly predicted pharmacoresistance in epilepsy patients with $P=1.3 \times 10^{-4}$, explaining 0.05% of the trait variance. Interestingly, the best prediction was achieved, when considering only strong associations with Alzheimer's disease with $P < 10^{-6}$. **Conclusions:** This study provides evidence for a common genetic contribution to pharmacoresistance in epilepsy as a phenotype and implicates comorbid NDs in the biology of pharmacoresistance. Specifically, common variation associated with Alzheimer's disease might help predict each epilepsy patient's response to pharmacotherapy, potentially adding to the body of evidence increasingly implicating inflammatory processes in epilepsy.

2401T

HLA risk alleles and CYP2C9 variant as predictor to prevent phenytoin hypersensitivity in Asians. S. Su^{1,2}, C. Chen^{2,3}, W. Chung^{1,2,3}. 1) Whole-Genome Research Core Laboratory of Human Diseases, Chang Gung Memorial Hospital, Keelung, Taiwan; 2) Department of Dermatology, Drug Hypersensitivity Clinical and Research Center, Chang Gung Memorial Hospital, Linkou, Taiwan; 3) College of Medicine, Chang Gung University, Taoyuan, Taiwan.

Importance: Phenytoin remains one of the most commonly used drugs to treat epilepsy. Although being effective in preventing provoked seizures, phenytoin can cause life-threatening, severe cutaneous adverse reactions (SCARs) like Stevens-Johnson syndrome (SJS), toxic epidermal necrolysis (TEN), and drug reaction with eosinophilia and systemic symptoms (DRESS). **Objective:** To implement a preemptive genetic test that comprises multiple predisposing alleles for the prevention of phenytoin hypersensitivity reactions, especially for SCARs. **Participants:** Three sets of patients with phenytoin-related SCARs and drug-tolerant controls, including 128 cases (65 SJS/TEN, 63 DRESS) and 376 controls from Taiwan, 129 cases (67 SJS/TEN, 62 DRESS) and 195 controls from Thailand, and 9 SJS/TEN and 94 controls from Japan, were enrolled for this study. **Main Outcomes and Measures:** Carrier rates of HLA-B alleles and a known genetic risk for phenytoin-related SCAR, *CYP2C9*3*, were examined in Taiwanese cohorts to select associated alleles for composing a multiplex genetic testing. The clinical validity of the proposed test was further assessed by constructing receiver operating characteristic (ROC) curves and performing meta-analysis. **Results:** In addition to *CYP2C9*3*, we found that *HLA-B*13:01*, *HLA-B*15:02* and *HLA-B*51:01* were significantly associated with phenytoin hypersensitivity reactions with distinct phenotypic specificities (*HLA-B*13:01* with DRESS; *HLA-B*15:02* with SJS/TEN; *HLA-B*51:01* linked to all phenotypes). Strikingly, we showed an increase in predictive sensitivity of concurrently testing *CYP2C9*3/HLA-B*13:01/HLA-B*15:02/HLA-B*51:01* from 30.5 to 71.9% for selecting the individuals with the risk of developing phenytoin-SCAR in Taiwanese cohorts, accompanied by a specificity of 77.7% (combined sensitivity, 64.7%; specificity, 71.9% for three Asian populations). Meta-analysis of the four combined risk alleles showed significant associations with phenytoin-SCAR in three Asian populations. **Conclusions and Relevance:** Combining the assessment of risk alleles of HLA and *CYP2C9* potentiated the implementation of predictive genetic tests to prevent phenytoin hypersensitivity in Asians.

2402F

Association of common and rare *GCH1* variants with Parkinson's disease in European and Ashkenazi Jewish populations. U. Rudakou^{1,2}, J.A. Ruskey^{2,3}, J.P. Ross^{1,2}, L. Krohn^{1,2}, S.B. Laurent^{2,3}, D. Spiegelman^{2,3}, C. Liong⁴, L. Greenbaum^{5,6}, G. Yahalom⁷, E.A. Fon^{2,8}, Y. Dauvilliers⁹, R.N. Alcalay^{4,10}, S. Hassin-Baer⁶, N. Dupré^{11,12}, Z. Gan-Or^{1,2,3}. 1) Department of Human Genetics, McGill University, Montreal, QC, Canada; 2) Montreal Neurological Institute, McGill University, Montreal, QC, Canada; 3) Department of Neurology and Neurosurgery, McGill University, Montréal, QC, Canada; 4) Department of Neurology, College of Physicians and Surgeons, Columbia University Medical Center, New York, NY, USA; 5) Joseph Sagol Neuroscience Center, Sheba Medical Center, Tel Hashomer, Ramat Gan, Israel; 6) Department of Neurology, Chaim Sheba Medical Center, Tel-Hashomer, Ramat-Gan, Israel; 7) Parkinson Disease and Movement Disorders Clinic, Chaim Sheba Medical Center, Tel-Hashomer, Ramat Gan, Israel; 8) Department of Experimental Medicine, McGill University, Montréal, QC, Canada; 9) National Reference Center for Narcolepsy, Sleep Unit, Department of Neurology, Gui-de-Chauliac Hospital, CHU Montpellier, University of Montpellier, Inserm U1061, Montpellier, France; 10) Taub Institute for Research on Alzheimer's Disease and the Aging Brain, College of Physicians and Surgeons, Columbia University Medical Center, New York, NY, USA; 11) Division of Neurosciences, CHU de Québec, Université Laval, Québec City, QC, Canada; 12) Department of Medicine, Faculty of Medicine, Université Laval, Québec City, QC, Canada.

Background: *GCH1* encodes the enzyme GTP cyclohydrolase 1, which is essential for dopamine synthesis in nigrostriatal cells. Loss-of-function mutations in *GCH1* lead to a severe reduction of dopamine synthesis in nigrostriatal cells and are the most common cause of DOPA-responsive dystonia (DRD). Co-occurrence of DRD and Parkinsonism has been reported in families with *GCH1* mutations, and a study on sporadic Parkinson's disease patients (PD) demonstrated an increased frequency of pathogenic *GCH1* mutations that were previously reported to cause DRD. *GCH1* is also implicated in genome-wide association studies of PD. While several studies replicated the association between rare or common *GCH1* variants and PD, other studies failed to replicate this association. **Methods:** First, *GCH1* and its 5' and 3' untranslated regions were sequenced in 1113 PD patients and 1147 controls, followed up by sequencing of additional 565 PD patients and 575 controls from Israel (92% Ashkenazi Jewish). Molecular inversion probes (MIPs) were designed in our lab for targeted capture of the regions of interest. To examine the association of rare *GCH1* variants with PD, optimized sequence kernel association test (SKAT-O) was performed using R. Common variants were analysed with logistic regression. **Results:** In the first cohort, three variants (p.R184C, p.M221T and p.K224R) that are considered to be pathogenic in DRD were found in five (~0.5%) PD patients, and none in controls (SKAT-O, $p=0.024$). One variant (p.V204I) that was previously reported to have conflicting interpretation of pathogenicity in DRD was found only in 3 controls (~0.3%). A common variant, rs841, was associated with reduced PD risk in our data (OR=0.71, 95% CI=0.61-0.83, $p=1.24 \times 10^{-4}$). This variant was associated with increased expression of *GCH1* in brain areas relevant for PD in the genome-tissue expression (GTEx) portal. The results of the analysis of the Israeli cohort are forthcoming. **Conclusions:** The association between *GCH1* variants and an increased risk for PD reported in this study provides further evidence that rare, pathogenic *GCH1* variants should be considered as a risk factor for Parkinson's Disease. Furthermore, common variants in the *GCH1* locus also affect the risk for PD, and the mechanism underlying this effect should be further studied. The study of the Israeli cohort will show if *GCH1* variants are associated with PD in a population of different ethnicity.

2403W

Modeling the function of Parkinson's disease risk alleles using multiplexed differentiation of iPSC-derived dopaminergic neurons. J.L. Jerber^{1,2}, D. Seaton³, D. Pearce², W. Chen⁴, O. Stegle⁵, F. Merkle^{5,6}, D. Gaffney². 1) Open Targets, Wellcome Genome Campus, Hinxton, CB10 1SD, UK; 2) Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, CB10 1SA, UK; 3) EMBL-EBI, Wellcome Genome Campus, Hinxton, CB10 1SD, UK; 4) Biogen, Cambridge MA 02142, USA; 5) Metabolic Research Laboratories and Medical Research Council Metabolic Diseases Unit, Wellcome Trust-Medical Research Council Institute of Metabolic Science, University of Cambridge, Cambridge, CB2 0QQ, UK; 6) The Anne McLaren Laboratory, Wellcome Trust-Medical Research Council Cambridge Stem Cell Institute, University of Cambridge, Cambridge, CB2 0SZ, UK.

Parkinson's disease (PD) is the second most common neurodegenerative disorder, affecting >3% of the population over the age of 75. PD is a multifactorial disease with genetic, age and environmental factors all playing a major role in disease pathogenesis. From the genetic standpoint, PD heritability is complex and Mendelian inheritance only account for 5-10% of the cases. Genome-wide studies (GWASs) have, to date, identified >41 PD risk loci however, the functional basis underlying risk mainly remains unknown. Functional characterisation of PD risk loci is challenging due to the inaccessibility of the affected cell types. Previous studies have been based on whole-tissue samples obtained from post-mortem donors, composed of many cell types with unknown specificity to the disease. Here, we used hiPSC technology to generate functional midbrain dopaminergic neurons (mDA), the cell type preferentially lost in this disease, in lines from 100 healthy donors derived by the Human Induced Pluripotent Stem Cell Initiative (HIPSCI). Our study used a "multiplexed" design, with lines from different donors pooled, differentiated and deconvoluted using single cell RNA-seq. This enabled us to control for between-differentiation batch effects, and efficiently collect samples across 4 differentiation time points, including a stimulation with the PD-associated compound, rotenone. We show that the representation of lines from different donors is maintained over the course of the differentiation, and identified different dopaminergic-like subpopulations and enrichment of Parkinson associated genes was clearly observed in a limited set of cells. Our study provides a novel framework for the discovery of novel disease pathways and for the efficient functional characterisation of disease risk loci across scores of cell subtypes and states in a single model system.

2404T

Association of IL-16 gene with ischemic stroke in Sudanese patients:

A cross-sectional study. H. Ahmed^{1,2}, A. Gassoum², O.A Seidi², M. Arbab^{2,3}, S. Eldeaf², H. Musa¹. 1) Faculty of Medical Laboratory Sciences, University of Khartoum, Sudan; 2) National Center of Neurological Sciences, Sudan; 3) Faculty of Medicine, University of Khartoum, Sudan.

Background: Stroke is the second leading cause of death after ischemic heart disease; IL-16 is identified as a pro-inflammatory cytokine that is a key element in the ischemic cascade after cerebral ischemia. **Objectives:** The aim was to study the association of IL-16 gene with ischemic cerebral stroke in Sudanese patients. **Methods:** A cross-sectional study was conducted for ischemic stroke patients at the National Center for Neurological Sciences (NCNS), Khartoum, Sudan, from March 2016 to September 2017. Demographic and clinical data was collected using structural questionnaire, hematological data including platelets count and D-dimer were estimated. Genomic DNA was extracted from blood, IL-16 gene was amplified using polymerase chain reaction (PCR). Thereafter, IL-16 PCR products were sequenced. **Results:** A 52.7% of the study population were males, 44% were at 65-82 year age and 29% were at 48-64 year age. Hypertension and heart disease were significantly ($P < 0.05$) associated with ischemic stroke, and 83% of ischemic stroke patients have D. dimer ≥ 0.3 mg/l ($P = 0.000$). High level of D. dimer was detected in 35.6% of hypertensive ischemic stroke patients, and 8.9% of heart diseases patients. The polymorphism rs11556218 T/G was detected in 75% of the patients (50% having TG allele, while 25% have GG allele). Mutation in IL-16 gene was associated with ischemic stroke risk factors including D. dimer ($OR = 3.5$) and hypertension ($OR = 1.6$). Moreover, it has significant relationship with ischemic stroke ($OR = 3$). **Conclusion:** These findings indicated that the rs11556218 T/G can be used as indicator for ischemic cerebral stroke in Sudanese population. **Key words:** ischemic stroke, Sudanese patients; D. dimer, IL-16, SNP.

2405F

Functional effects of damaging mutations in *RAI1* in neurons and astrocytes from patient-derived induced pluripotent stem cells. J.L. Cross¹, S. Detera-Wadleigh¹, A.C.M. Smith², F.J. McMahon¹. 1) National Institute of Mental Health, National Institutes of Health, Bethesda, MD; 2) National Human Genome Research Institute, National Institutes of Health, Bethesda, MD.

Retinoic Acid Induced I (*RAI1*) encodes a transcriptional regulator involved in circadian rhythms and neuronal differentiation. *RAI1* is highly expressed in brain, and mutations or structural variants involving the gene cause neurodevelopmental disorders such as Smith-Magenis Syndrome (SMS) and Potocki-Lupski Syndrome (PTLS). Most SMS cases carry a ~3.7 MB deletion on chromosome 17p11.2 that involves *RAI1* and up to 14 additional genes, but point mutations in *RAI1* are sufficient to cause SMS in about 10% of cases. Clinical features of SMS include obesity, craniofacial abnormalities, sleep cycle disturbance, and intellectual disability; some patients also manifest mood disorders and autism spectrum symptoms. Several studies have investigated *RAI1* deletions in mice, but very few studies have examined the functional effects of *RAI1* mutations in patient-derived cells. We have generated induced pluripotent stem cells (iPSCs) from SMS patients with point mutations in *RAI1* to study the impact of these mutations at the cellular level. Two cell lines were derived from patients with missense mutations in exon 3, resulting in premature truncation of the protein. The third patient line carries a frameshift in the first 548 bases, resulting in a severe phenotype. All three mutations are predicted to disrupt the nuclear localization tag within the C-terminal end of RAI1 protein. We have differentiated the iPSCs into neural precursor cells, neurons, and astrocytes and our ongoing studies compare cellular phenotypes in these patient-derived and matched control cell lines. In addition to standard measures of neuronal growth and morphology, bulk RNA-seq, and ATAC-seq, we are using immunocytochemistry to investigate the sub-cellular localization of the RAI1 protein. Planned studies will also investigate expression of circadian pathway genes. We hypothesize that the *RAI1* mutations causing SMS will affect the sub-cellular localization of RAI1 and disrupt expression of genes involved in circadian rhythms, neuronal differentiation, and other key pathways, resulting in cellular phenotypes that may model aspects of SMS, bipolar, and autism spectrum disorders in vitro. Such cellular models may provide insights into mechanisms involved in SMS and other neuropsychiatric disorders, pointing toward novel treatments.

2406W

APOE, Alzheimer's disease, and Hispanic populations. *E.E. Blue¹, A.R.V.R. Horimoto², E.M. Wijsman^{1,2,3}, T.A. Thornton².* 1) Division of Medical Genetics, University of Washington, Seattle, WA; 2) Department of Biostatistics, University of Washington, Seattle, WA; 3) Department of Genome Sciences, University of Washington, Seattle, WA.

Although the relationship between *APOE* and both Alzheimer's disease (AD) risk and age-at-onset (AAO) is well-established in cohorts with European ancestry, there is debate regarding its effects in other ancestry groups. We compared the effects of *APOE* on AD risk and AAO among individuals with European and Hispanic ancestry from the National Institute on Aging's Late-Onset Alzheimer's Disease and the National Cell Repository for Alzheimer's Disease (NIALOAD) study with the Columbia University Study of Caribbean Hispanics with Familial and Sporadic Late Onset AD (CU Hispanics). Both Hispanic data sets represent admixed individuals, with proportions of African, European, and Native American ancestry varying between cases and controls and near *APOE*. Survival analyses revealed *APOE* had significant effects on AAO of AD across data sets. The effect of *APOE* genotype on AAO of AD among Europeans was nearly twice that observed among Hispanics. The count of *APOE* E4 alleles significantly increased the hazard of AD all data sets ($p < 7 \times 10^{-5}$), while the count of *APOE* E2 alleles significantly decreased the hazard of AD in the Europeans and CU Hispanics ($p < 2 \times 10^{-16}$) and approached significance in the NIALOAD Hispanics ($p = 0.13$). *APOE* proved to be an important covariate for genome-wide association studies (GWAS). Across all data sets, *APOE* adjustment significantly improved model fit for AAO association testing ($p < 0.01$), but not case-control analyses. The effects of *APOE* on GWAS results differed dramatically between European and Hispanic data. Although all GWAS adjusted for population structure and relatedness, the correlation between p-values from models with and without *APOE* adjustment was much weaker among the European data vs. the Hispanic data for both AD risk ($r^2 = 0.49$ vs. $r^2 = 0.97$ or $r^2 = 0.81$) and AAO ($r^2 = 0.58$ vs. $r^2 = 0.87$ or $r^2 = 0.83$). GWAS without *APOE* adjustment found significant association between the *APOE* region and both AD risk and AAO in the Europeans and CU Hispanics, and significant association between AAO of AD and 11p11.2 ($p = 4.9 \times 10^{-8}$) in the Europeans and 14q24.3 ($p = 1.33 \times 10^{-10}$) in the CU Hispanics. The association between AAO of AD and 11p11.2 and 14q24.3 were robust to *APOE* adjustment and correspond to genes implicated in AD risk or pathology. Our results indicate that ancestry informs the association between *APOE* and both risk and AAO of AD, and that GWAS results in samples with European ancestry are sensitive to *APOE* adjustment.

2407T

Understanding genetic regulatory mechanisms of neurodegenerative and psychiatric disorders by cell-type specific deconvolution of gene expression profile in brain. *L. He^{1,4}, Y. Park^{1,4}, J. Davila-Velderrain^{1,4}, H. Mathys², L. Hou^{1,4}, P. De Jager^{2,4}, L. Tsai^{2,4}, M. Kellis^{1,4}.* 1) Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, MA; 2) Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology; 3) Department of Neurology, Columbia University Medical Center, New York, NY, USA; 4) Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA.

Cell-type specific RNA-seq information can help reveal regulatory mechanisms underlying complex traits and human disease, but is not readily available across large disease cohorts. To overcome this challenge, we developed a new method for deconvolution of bulk RNA-seq data using single-cell RNA-seq information for 80,000 cells in human frontal cortex across 48 AD and non-AD individuals, and bulk RNA-seq information across 482 individuals from the ROSMAP cohort. We focused on six major cell types, including excitatory and inhibitory neurons, oligodendrocytes, astrocytes, microglia and oligodendrocyte progenitor cells. We observed a significant reduction in excitatory neurons for AD ($p = 10^{-6}$) and age ($p = 10^{-6}$), and a converse expansion of microglia ($p = 10^{-6}$, $p = 10^{-37}$, respectively). Both beta-amyloid and tau were significantly associated with fewer excitatory neurons ($p = 10^{-6}$). Female individuals showed more inhibitory neurons than males ($p = 10^{-6}$). The majority of AD GWAS genes showed altered microglial expression, highlighting the central role of microglia in AD. To recognize genetic variants underlying this association, we used the fraction of each cell type as the phenotype in a GWAS for each cell type, resulting in numerous significant SNPs. For excitatory neurons, we found that *TMEM106B*, a previously reported GWAS locus for frontotemporal dementia, was significantly associated ($p = 10^{-7}$). We also estimated SNP-cell type interaction effects in relevant brain cell types for multiple brain-related diseases using a factorized QTL model. We found 39 independent factors with FDR $< 1\%$, revealing transcription of ~600 protein coding genes are regulated in cell type specific manner. We then used the estimated SNP-cell type interaction effects to find cell-type specific causal genes. We applied multivariate effect sizes of these genes to summary-based transcriptome-wide association studies (TWAS) of multiple brain disorders. We found a significant association between oligodendrocyte expression of *HLA-DRB1* and *HLA-DQB1* and AD ($P = 10^{-7}$) and with maternal history of AD in the UK Biobank ($p = 10^{-6}$). We also discovered that *NP1A2* gene in excitatory neuron is strongly associated in schizophrenia TWAS ($p < 3.2 \times 10^{-7}$) and autism spectrum disorder TWAS ($p = 10^{-6}$).

2408F**Blood total cholesterol polygenic score in Alzheimer's disease.** *N.I.V.*

Nilsson^{1,2}, *J. Poirier*^{1,2}. 1) Department of Psychiatry, McGill University, Montreal, QC, Canada; 2) Douglas Mental Health University Institute, Montreal, QC, Canada.

Sporadic Alzheimer's disease (sAD) is considered a multifactorial disease caused by a combination of risk factors such as old age, genetics, and lifestyle factors. E.g. approximately half of the sAD phenotypic variance is explained by genetics, however most of these variants remain unknown. One of the suggested lifestyle factors in sAD is blood cholesterol. The results are somewhat conflicting, but a consensus is that high blood cholesterol levels specifically in midlife is associated with increased risk of developing sAD. Blood cholesterol levels are also influenced by genetics although not as strongly as for AD. Considering the genetic background of both conditions, and the fact that they are linked in terms of risk, we hypothesized that a polygenic score (pgs) based on blood total cholesterol (TC) levels would associate positively with sAD and its biomarkers. We used a previously published meta-analysis of GWAS for blood total cholesterol levels (Teslovich et al., 2010) to compute a TC pgs in the Alzheimer's Disease Neuroimaging Initiative cohort (ADNI,adni.loni.usc.edu). We found that the TC pgs indeed associates positively with blood TC levels. Logistic and linear regressions with sAD risk and with CSF levels of sAD biomarkers A β and p-tau, respectively, did not reveal any significant associations. However, we did find a significant positive association with CSF t-tau, a biomarker typically associated with neuronal loss in the CNS. These results suggest that there is a very small, if any effect of genetics underlying cholesterol levels in sAD. However, the significant association between TC pgs and CSF t-tau levels suggests that genetics underlying cholesterol levels could have an effect on general neurodegeneration.

2409W

Whole exome sequencing of DNA from a patient with bvFTD (behavioral variant of frontotemporal dementia) reveals a deleterious SNP in MAPT gene. *M. Xi*¹, *R. Zhang*¹, *H. Wang*¹, *Y. Cai*², *A. Baskys*³. 1) Psychiatry, Xijing Hospital, Air Force Medical University, Xi'an, Shaanxi, Peoples Republic of China; 2) Neurobiology, Xuanwu hospital, Changchun street, NO.45, 100053 Beijing, People Republic of China; 3) Graduate College of Biomedical Sciences Western University of Health Sciences Pomona, CA 91766.

FTD is a significant clinical problem affecting younger individuals (ages 40-65) and often masquerading as a treatment resistant psychiatric disorder such as depression or psychosis. It is highly heritable and mutations in microtubule associated protein tau (MAPT), progranulin (PGRN) and chromosome 9 open reading frame 72 (C9orf72) expansion mutations have been found in familial FTD. It has been proposed that clinical presentations and treatment responsiveness of FTD patients vary depending on variant genes present (Young et al., 2017). Thus, MAPT mutation carriers tend to deteriorate when treated with ACh inhibitors and it has been suggested that PGRN mutation effects could be mitigated by small molecule drugs (De Munnick and Van Damme, 2015). To further explore the genotype-phenotype association, we performed whole exome sequencing (WES) of DNA extracted from a blood sample from a proband, a 56 year old Chinese female, with previously well characterized behavioral FTD (bvFTD) syndrome (Pospos et al. 2018), and her symptom-free son, 33. The Ethic's Committee of Xijing Hospital approved the study. SureSelect Human All Exon V6 64M (Agilent) was used to capture target DNA and HiSeq X Ten System (Illumina) was used for sequencing. Paired-end alignment to the 1000 genomes hg19/GRCh37 reference genome was performed using BWA ALN. SAM files were sorted, converted to BAM, and duplicates were marked with Picard. GATK was used for local realignment and base quality score recalibration, variants were called jointly in all samples using the GATK's HaplotypeCaller in the "GENOTYPE_GIVEN_ALLELES" mode. PROVEAN, SIFT, PolyPhen2 HDIV, PolyPhen2 HVAR, MutationTaster, M-CAP and REVEL were used to predict deleteriousness of amino acid changes. MaxEntScan was used for splicing defect prediction. Data showed a rare (MAF=0.014, China Thousand Human Genomes) missense single nucleotide variant c.689(exon 6)A>G in MAPT gene resulting in amino acid change (p.Q230R), which was deleterious by several predictors. Patient's son had no pathogenic MAPT variants. Our findings elucidate genotype-phenotype association in bvFTD and lend support to clinical applications of advanced technologies such as Next Generation Sequencing to improve diagnostic accuracy and minimize unnecessary investigations and treatments.

2410T

Differences in the maternal and paternal contribution to inherited risk for autism. D. Antaki^{1,2,3,5}, W.M. Brandler¹, M. Gujral^{1,2,3}, P. Tandon^{1,2,3}, J. Sebat^{1,2,3}. 1) Beyster Center for Genomics of Psychiatric Diseases, University of California San Diego, La Jolla, CA 92093, USA; 2) Department of Psychiatry, University of California San Diego, La Jolla, CA 92093, USA; 3) Department of Cellular and Molecular Medicine and Pediatrics, University of California San Diego, La Jolla, CA 92093, USA; 4) Human Longevity, Inc., San Diego, CA 92121, USA; 5) Biomedical Sciences Graduate Program, University of California San Diego, La Jolla, CA 92093, USA.

The genetic basis of autism spectrum disorder (ASD) is known to consist of *de novo* and inherited loss of function mutations in haploinsufficient genes. Previous studies have reported that inherited risk is predominantly carried by mothers. However, the distinct contributions of each parent to inherited risk for ASD has not been explored. We investigated paternal and maternal contributions to ASD by analyzing the transmission of private deletions in coding and *cis*-regulatory (CRE-SVs) regions of haploinsufficient genes in whole genomes of 9274 individuals (2600 families). Consistent with previous findings, mothers over-transmitted exonic deletions in functionally constrained genes to affected offspring but not to control siblings (Affected Transmission Rate=63.0%, Control Transmission Rate=48.8%, Affected Binomial P=0.0091). Likewise, fathers transmitted more exonic deletions in the same genes to affected but not to control offspring (Affected Transmission Rate=68.1%, Control Transmission Rate=48.8%, Affected Binomial P=0.0001). Additionally, we found that fathers but not mothers significantly over-transmit CRE-SVs to affected offspring relative to control siblings (Paternal Affected Transmission Rate=63.0%, Paternal Binomial P=0.015, Maternal Affected Transmission Rate=51.1%, Maternal Binomial P=0.43). Taken altogether, paternally inherited structural variants in functionally constrained regions of the genome also contribute risk to ASD with CRE-SVs exhibiting a strong paternal origin effect, suggesting a complex genetic architecture of autism. We are currently expanding our analyses of transmission distortion to incorporate single nucleotide variants and insertions/deletions including the addition of whole genomes from 4712 subjects (1178 families). With over 10,000 whole genomes, we aim to discover distinct modes of inheritance conditioned on the genetic burden of parents. Consistent with our earlier finding of paternally derived exonic deletions, initial findings show an over-transmission of private loss of function single nucleotide variants in functionally constrained genes from fathers to affected children (Paternal Transmission Rate=53.1%, Binomial Pvalue=0.04). Our work demonstrates that the paternal origin effect of risk variants contributes genetic risk for ASD in more complex ways than previously thought.

2411F

Investigating the genetic architecture of different autism spectrum disorder subtypes through deep phenotyping and whole-genome sequencing.

A.J.S. Chan^{1,2}, W. Engchuan¹, M.S. Reuter¹, Z. Wang¹, B. Thiruvahindrapuram¹, S. Lamoureux¹, T. Nalpathamkalam¹, R.V. Patel¹, G. Pellecchia¹, K. Tammimies³, S. Walker¹, B. Trost¹, M. Uddin^{1,4}, M. Zarrei¹, C.R. Marshall¹, D.J. Stavropoulos⁵, D. Merico^{1,6}, R.K.C. Yuen^{1,2}, B.A. Fernandez^{7,8}, S.W. Scherer^{1,2,9}.

1) The Centre for Applied Genomics and Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada; 2) Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada; 3) Center of Neurodevelopmental Disorders (KIND), Department of Women's and Children's Health, Karolinska Institutet, Solna, Sweden; 4) Mohammed Bin Rashid University of Medicine and Health Sciences, College of Medicine, Dubai, United Arab Emirates; 5) Department of Pediatric Laboratory Medicine, Genomic Diagnostics, The Hospital for Sick Children, Toronto, Ontario, Canada; 6) Deep Genomics Inc., Toronto, Ontario, Canada; 7) Provincial Medical Genetics Program, Eastern Health, St John's Newfoundland and Labrador, Canada; 8) Disciplines of Genetics and Medicine, Memorial University of Newfoundland, St John's Newfoundland and Labrador, Canada; 9) McLaughlin Center, University of Toronto, Toronto, Ontario, Canada.

Recent genomic studies have identified hundreds of genes and loci contributing to autism spectrum disorder (ASD) susceptibility, revealing the genetic complexity of ASD. Clinical morphology classification can help stratify ASD heterogeneity into more genetically meaningful subtypes. Using microarray and whole-exome sequencing, we previously determined a molecular diagnostic yield of 15.8% in 95 ASD cases, with the yield higher in the complex (significantly dysmorphic) ASD subtype (Tammimies *et al*, JAMA, 2015). Here, we aimed to examine the molecular diagnostic yield in a larger ASD cohort (n=329), and to dissect the genes and pathways involved in different ASD subtypes using whole-genome sequencing (WGS), with the advantage of detecting a full size spectrum of variants in a single experiment. A total of 329 children with ASD from Newfoundland, Canada were enrolled using a population-based recruitment strategy. Standardized dysmorphology assessments, including scoring for minor physical and major congenital anomalies, were completed and used to stratify the cohort into 193 essential (non-dysmorphic), 57 equivocal, and 79 complex ASD cases. We identified clinically significant single nucleotide variants, insertions/deletions and copy number variants (CNVs) in 15.2% of ASD cases. The yield was higher in the complex subgroup (38.0%; 30/79) compared to the essential subgroup (7.3%; 14/179) (p=1.06×10⁻⁶). Clinically significant variants included 16p11.2 microdeletion and microduplication, and variants impacting *SHANK3*, *CHD8*, *NRXN1*, and *ASTN2/TRIM32*. Three ASD cases (~1%) had two clinically significant and/or ASD-risk variants. We compared global burden and gene-set enrichment of rare CNVs among ASD subtypes. We found that ASD cases with more dysmorphologies were associated with increased number of genes impacted by deletions (p=1.52×10⁻⁶). These cases had a significant enrichment for rare deletions in neuronal genes, with the highest enrichment in genes responsible for neuron projection (p=0.0047). Our findings show that WGS captures clinically significant variants with different size spectra for molecular diagnosis of ASD, and that differences in the genetic architecture among ASD subtypes can be characterized using a combination of rich genome-wide data and extensive clinical data.

2412W

Replication of a risk haplotype on 1p36.33 for autism spectrum disorder.

N.H. Chapman¹, R.A. Bernier^{2,3}, S.J. Webb^{2,3}, J. Munson^{2,3}, E.M. Blue¹, D.H. Chen⁴, E. Heigham⁴, W.H. Raskind^{1,2,5}, E.M. Wijsman^{1,5,6}. 1) Division of Medical Genetics, University of Washington, Seattle, WA; 2) Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA; 3) Center on Human Development and Disability, University of Washington, Seattle, WA; 4) Department of Neurology, University of Washington, Seattle, WA; 5) Department of Genome Sciences, University of Washington, Seattle, WA; 6) Department of Biostatistics, University of Washington, Seattle, WA.

Hundreds of genes have been implicated in the genetics of autism spectrum disorders (ASDs). In genetically heterogeneous conditions, large families with multiple affected individuals can provide strong evidence implicating a rare variant, and replication of the same variant in multiple families is very unusual. We previously published linkage analyses and follow-up exome sequencing in 7 large families with ASDs, implicating 14 candidate exome variants. These included rs200195897, which was transmitted to 4 affected individuals in 1 family. In this paper, we attempted replication of those variants in the MSSNG database. MSSNG is a unique resource for replication of ASD risk loci, containing whole genome sequence (WGS) on thousands of individuals diagnosed with ASDs and their immediate family members. For each of the exome variants previously identified, we obtained all carriers and their relatives in MSSNG, using a TDT test to quantify evidence for transmission and association. We replicated the transmission of rs200195897 to 4 affected individuals in 3 additional families. rs200195897 was also present in 3 singleton affected individuals, and no unaffected individuals other than transmitting parents. The availability of WGS in parents of 4 affected individuals allowed us to identify 2 additional rare variants (rs566472488 and rs185038034) transmitted on the same background haplotype on 1p36.33. We sequenced these variants in the family in which rs200195897 was initially ascertained, confirming their presence. We have identified a haplotype carrying 3 rare variants (alternate allele frequency 0.0007 in Non-Finnish Europeans) transmitted to 8 individuals with autism in 4 families, and present in 3 other affected individuals. To our knowledge, this is the first example of a very rare haplotype being transmitted with ASD in multiple families. The candidate risk variants include a missense mutation in *SAMD11*, an intronic variant in *NOC2L*, and a regulatory region variant close to both genes. *NOC2L* is a transcription repressor, and several genes involved in transcription regulation have been previously associated with ASDs.

2413T

Both rare and common genetic variants contribute to autism in the Faroe Islands.

F. Cliquet^{1,2,3,4}, C. Leblond^{1,2,3,4}, C. Carton^{1,2,3,4}, G. Huguet^{1,2,3,4}, A. Mathieu^{1,2,3,4}, T. Kergrohen^{1,2,3,4}, J. Buratti^{1,2,3,4}, N. Lemi re^{1,2,3,4}, A. Bolland⁵, J.F. Deleuze⁶, T. Stora⁶, R. Biskupstoe⁷, J. Halling⁷, G. Andorsd ttir⁸, E. Billstedt⁹, C. Gillberg⁹, T. Bourgeron^{1,2,3,4}, GenMed consortium. 1) Institut Pasteur, Human Genetics and Cognitive Functions Unit, Paris, France; 2) CNRS UMR 3571 Genes, Synapses and Cognition, Institut Pasteur, Paris, France; 3) University Paris Diderot, Sorbonne Paris Cit , Paris, France; 4) Centre de Bioinformatique, Biostatistique et Biologie Int grative, Paris, France; 5) Centre National de Recherche en G n mique Humaine (CNRGH), CEA, Evry France; 6) Psychiatric Centre at Landssjukrahusid, Faeroe Islands; 7) Ministry of Health Genetic Biobank of the Faroes T rshavn Faeroe Islands; 8) Institute of Neuroscience and Physiology, Sahlgrenska Academy, University of Gothenburg, Sweden; 9) Gillberg Neuropsychiatry Centre G teborg, Sweden.

The number of genes associated with autism spectrum disorder (ASD) is increasing, but few studies have been performed on epidemiological cohorts and in isolated populations. Here, we investigated 357 individuals from the Faroe Islands including 36 patients with ASD, 136 of their relatives and 185 non-ASD controls. Data from SNP array and whole exome sequencing revealed that patients had a higher burden of copy-number variants ($P < 0.05$), higher inbreeding status ($P < 0.005$), higher load of homozygous deleterious mutations ($P < 0.0005$), and a higher ASD polygenic risk score ($P < 0.01$) compared to controls. Overall, for 11% of individuals with ASD, a known genetic cause was identified, for 39% at least one strongly deleterious mutation was identified in a compelling candidate gene and for 50% no obvious genetic cause was detected. Our analysis supports the role of several genes/loci associated with ASD (e.g. *NRXN1*, *ADNP*, 22q11 deletion) and identified new truncating (e.g. *GRIK2*, *ROBO1*, *NINL* and *IMMP2L*) or recessive deleterious mutations (e.g. *KIRELL3* and *CNTNAP2*) affecting ASD-risk genes. It also revealed three genes involved in synaptic plasticity, *RIMS4*, *KALRN* and *PLA2G4A*, carrying *de novo* deleterious mutations in patients without intellectual disability. In summary, our analysis provides a better understanding of the genetic architecture of ASD in isolated populations by highlighting the role of both common and rare gene variants and pointing at new ASD-risk genes. It also indicates that more knowledge about how multiple genetic hits affect neuronal function will be necessary to fully understand the genetic architecture of ASD.

2414F

Genetic investigation of insistence on sameness in autism. M.L. Cuccaro^{1,2}, E.R. Martin^{1,2}, S. Luz², H.N. Cukier^{2,3}, A.J. Griswold^{2,4}, D. Dykxhoorn^{1,2}, J.M. Vance^{1,2}, M.A. Pericak-Vance^{1,2}. 1) Human Genetics, University of Miami School of Medicine, Miami, FL., USA; 2) Dr. John P. Hussman Institute of Human Genetics, University of Miami, Miami, FL, USA; 3) Neurology, University of Miami School of Medicine, University of Miami, FL, USA; 4) Pathology, University of Miami School of Medicine, University of Miami, FL, USA.

Background: Restricted and repetitive behaviors (RRBs) are a defining feature of autism and fall into two groups: insistence on sameness (IS) and repetitive sensory motor behaviors (RSMB). IS is the more robust of the two and appears to have reduced overlap with intellectual disability phenotypes. To date, there has been limited study of the genetic basis of IS. **Objective:** We tested for genetic association with IS in a sample of individuals with autism with targeted sequence data. We hypothesize that multiple genetic variants will be associated with the IS score. **Methods:** Using Autism Diagnostic Interview-Revised (ADI-R) items (difficulties with minor changes, resistance to trivial changes, abnormal response to sensory stimuli, circumscribed interests, compulsions/rituals, sensitivity to noise), we calculated IS scores for our dataset of 1931 participants from the Hussman Institute for Human Genomics and Simons Simplex Collection. All individuals were sequenced with a 17Mb custom capture covering 681 genes within regions identified by GWAS of ASD. We conducted the SKAT-O gene-based test and single variant tests using IS scores as continuous outcomes. A Bonferroni correction for the number of genes tested was used as a significance threshold. **Results:** For the gene-based test our top four results were: *ATXN1*, *ARHGAP40*, *BICD1* and *GCOM1* with p-values of 5.6E-03, 5.8E-03, 6.2E-03 and 7E-03, respectively. When only missense variants were analyzed, the top three genes were *PANK1*, *TIMM22* and *ZNF397* with p-values of 1E-02, 1.2E-02 and 1.3E-02 respectively. *ATXN1*, located on the short arm of chromosome 6, is a protein coding gene and is associated with spinocerebellar ataxia type 1 (SCA 1). *ZNF397* is located on the long arm of chromosome 18 and is part of the zing finger family. This gene is well established for its role in and has shown prior association with autism. Overall, we identified 32 genes with p-values of less than 0.05. **Conclusions:** We did not find genome-wide association to the IS trait. Our top nominally significant results suggest several genes that may be associated with the IS trait in autism but need further study. Our findings serve as a first step in dissecting the genetic basis of the IS trait in autism. We propose follow up studies use a refined IS trait. Generally, our results suggest the value of identifying genes associated with phenotypic traits that are a part of autism as an alternative to the broader autism diagnostic phenotype.

2415W

Analysis of convergent molecular pathways during neural development using ASD-specific induced pluripotent stem cells. D.M. Dykxhoorn¹, B.A. DeRosa¹, D. Van Booven¹, C. Garcia-Serje¹, M.L. Cuccaro¹, J.M. Vance¹, M. Nestor², H.N. Cukier¹, M.A. Pericak-Vance¹. 1) Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami FL; 2) Hussman Institute for Autism, Baltimore, MD.

Autism spectrum disorders (ASDs) are a phenotypically and etiologically complex group of neurodevelopmental conditions that are behaviorally defined based on the presence of social-communicative problems and restricted and repetitive behaviors. Recent studies have shown that ASD risk loci are highly enriched in genes expressed during early neocortical development and genes which encode proteins that function in biological pathways involved in regulating transcription, chromatin remodeling, cell adhesion, signaling complexes, and synapse function. The aim of this study is to use induced pluripotent stem cell (iPSC)-based approaches to identify key differences in important neurobiological processes by evaluating the expression of gene networks that regulate these processes, both in a temporal and neural-specific manner. This RNA-seq analysis was complemented by functional analysis of these neurons using electrophysiological, morphological, and biochemical analyses. Transcriptional analyses of ASD and control neurons at culture days 35 and 135 of their *in vitro* development showed ASD-specific transcriptional signatures mainly affecting pathways/networks involved in neuronal differentiation, the cytoskeletal matrix structure formation (i.e. axon guidance and cell migration), regionalization, patterning, DNA and RNA metabolism. Additionally, developing networks of neurons were analyzed using multi-electrode array (MEA) recordings, measurements of calcium transients, immunocytochemical analyses for important morphological markers, and cell migration assays at time-points aligning with the transcriptional analyses. Neurons from ASD individuals demonstrated significantly decreased network spiking activity from MEA recordings (40-80% decrease, $p < 0.05$) as well as decreased numbers of calcium transients (30-60% decrease, $p < 0.01$). Additionally, ASD lines showed significant differences in neurite morphology and decreased cell migration ($p < 0.05$) at early neuronal differentiation times. The results of this study suggest that iPSC-derived neurons from individuals with ASD may have early deficits in network activity and morphology based on a combination of cell based assays, including MEA, calcium transients, and quantification of neurite outgrowth that complement with transcriptomic profiles. Taken together, these data suggests a convergence of pathophysiological processes that effect neuronal functionality in ASD.

2416T

Pilot study results from SPARK: A scalable approach for genomic research in autism spectrum disorder. P. Feliciano¹, X. Zhou², I. Astrovskaya¹, T. Turner³, T. Wang³, S. Murali³, R. Barnard⁴, J. Wang², A. Hsieh², E. Ruzzo⁵, A. Sabo⁶, E. M. Berry-Kravis⁷, C. Eng⁸, D. Muzny⁹, L. Grosvenor¹, M. Geisheker³, L. Brueggeman⁴, T. Pramparo³, M. Dennis¹⁰, J. S. Sutcliffe¹¹, T. W. Yu¹², J. J. Michaelson⁸, C. Lese Martin¹³, R. Gibbs⁶, E. E. Eichler³, B. J. O'Roak⁴, N. Volfovsky¹, Y. Shen², W. K. Chung^{2,11}. 1) Simons Foundation, New York, NY; 2) Columbia University, New York, NY; 3) University of Washington, Seattle, WA; 4) Oregon Health Sciences University, Portland, OR; 5) University of California, Los Angeles, CA; 6) Baylor College of Medicine, Houston, TX; 7) Rush University Medical Center, Chicago, IL; 8) The University of Iowa, Iowa City, IA; 9) University of California, San Diego, CA; 10) University of California, Davis, CA; 11) Vanderbilt University, Nashville, TN; 12) Children's Hospital Boston, Boston, MA; 13) Geisinger, Danville, PA.

Autism spectrum disorder (ASD) is a phenotypically heterogeneous condition, and its genetic architecture involves a combination of rare de novo and inherited variants as well as common variants in at least several hundred genes. The heterogeneity in ASD increases the complexity of gene discovery and has led to challenges with reproducibility of clinical research findings. Although the largest genetic studies of ASD have included more than 16,000 individuals with ASD, significantly larger sample sizes are needed to identify genetic risk factors. SPARK (SPARKforAutism.org) is a new paradigm for scalable clinical and genomic research that combines convenient online access to participants, longitudinal collection of phenotypic data, return of meaningful individual-level genetic and phenotypic results and data sharing. All genomic and phenotypic SPARK data are available at sfaribase.org, and all SPARK participants are available for recontact and additional research studies. Since launching in 2016, SPARK has consented 42,848 individuals with ASD to participate in longitudinal research. Saliva for genetic studies has been collected from 22,283 individuals with ASD and 37,499 of their first-degree family members. In a pilot study of 462 families with ASD (12 of whom self-reported a known genetic diagnosis), we identified both pathogenic (14%) and probable (8%) genetic causes of ASD in 99 (22%) out of 450 families without a reported genetic diagnosis. Genetic causes consisted of de novo likely gene-damaging (dnLGD) variants in constrained genes ($pLI \geq 0.5$) in 44 families, de novo predicted damaging missense ($pLI \geq 0.9$, REVEL Score ≥ 0.75) variants in 16 families, deleterious CNVs in 37 families and X-linked mutations in 2 families. Our analyses identified deleterious variants in genes known to be high-confidence risk genes for ASD, as well as in genes previously implicated in neurodevelopmental disorders, including *DPP6*, *MEIS2*, *QRICH1*, and *SETD1A*. We also identified two dnLGD variants in *BRSK2* (TADA q -value=0.002), supporting *BRSK2* as a new candidate gene for ASD. In one proband with a known 2 Mb deletion at 4q21, we also detected 2 de novo, predicted damaging missense variants in *MED13L* and *CLCN4*, providing support for multiple genetic contributions in some cases and demonstrating the value of both CNVs and sequence variants. Exome/genome sequencing and SNP-genotyping are in progress for 4,842 additional probands and their available first-degree family members.

2417F

Disruption of CSDE1 leads to autism and interferes with neuronal development and synapse plasticity. H. Guo^{1,2}, Y. Li¹, L. Shen¹, H. Wu¹, L. Liu³, T. Wang^{1,2}, K. Hoekzema², N. Pang⁴, J. Zhao⁵, C. Romano⁶, M. Nordenskjöld⁷, J. Gecz⁸, H. Li⁹, R. Bernier¹⁰, W. Xie³, E. Eichler^{2,11}, K. Xia¹. 1) Center of Medical Genetics, School of Life Sciences, Central South University, Changsha, Hunan, China; 2) Department of Genome Sciences, University of Washington School of Medicine, Seattle, Washington, USA; 3) Institute of Life Sciences, the Key Laboratory of Developmental Genes and Human Disease, Southeast University, Nanjing, Jiangsu, China; 4) Department of Pediatrics, the Xiangya Hospital, Central South University, Changsha, China; 5) Mental Health Institute of the Second Xiangya Hospital, Central South University, Changsha, Hunan, China; 6) Unit of Pediatrics & Medical Genetics, Oasi Research Institute-IRCCS, Troina, Italy; 7) Department of Molecular Medicine and Surgery, Center for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden; 8) School of Medicine and the Robinson Research Institute, the University of Adelaide at the Women's and Children's Hospital, Adelaide, South Australia, Australia; 9) Key Laboratory of Developmental Disorders in Children, Liuzhou Maternity and Child Healthcare Hospital, Liuzhou, Guangxi, China; 10) Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, Washington, USA; 11) Howard Hughes Medical Institute, University of Washington, Seattle, Washington, USA.

RNA-binding proteins are key players in the regulation of gene expression and have been implicated in disease susceptibility, especially in neurological disorders. Here, we present evidence that disruptive mutations of *CSDE1*, which encodes a highly constrained RNA-binding protein involved in neural differentiation, associate with autism spectrum disorders (ASD) and related neurodevelopmental disorders. We observe an excess of *CSDE1* de novo truncating mutations in ASD ($P < 8 \times 10^{-7}$, $P_{adj} < 0.01$) and a significant burden of putative disruptive mutations in a broader neurodevelopmental disorder cohort ($P = 7 \times 10^{-6}$). Patients display consistent phenotypes, including ASD, intellectual disability, and motor and language developmental delay. Using HITS-CLIP, we show that *CSDE1* binding targets are enriched in the targets of two ASD-associated RNA-binding proteins, FMRP and RBFOX. Pathway enrichment analysis reveals that *CSDE1* targets are mainly involved in neuronal development and synaptic plasticity. *CSDE1* knockdown experiments in primary mouse cortical neurons show an increase in neurite length and a decrease in dendritic spine density, while knockout and knockdown experiments in *Drosophila* show defects in synaptic plasticity and development, including an increase in neuromuscular junction bouton number. Our study defines a subtype of autism caused by *CSDE1* putative disruptive mutations and highlights the functional role of *CSDE1* in neuronal development and synaptic plasticity.

2418W

Investigation of novel rare variants in *ASTN2* in Japanese patients with autism spectrum disorders and schizophrenia. K. Ishizuka, M. Toyama, B. Aleksic, N. Ozaki. Department of Psychiatry, Nagoya University Graduate School of Medicine, Nagoya, Aichi, Japan.

ASTN2 is a protein that is expressed in the brain and may function in neuronal migration. A deletion at this locus has been associated with schizophrenia. Impairments caused by genetic variants in *ASTN2* might be important for the pathomechanisms in neuropsychiatric disorders including autism spectrum disorders (ASD) and schizophrenia. In the current study, we focused on the investigation of rare coding variants in *ASTN2*, which play major roles in disease causation and might contribute to the missing heritability from genome-wide association studies. First, we conducted exon-targeted resequencing of *ASTN2* with next-generation sequencing technology in 562 Japanese ASD and schizophrenia individuals and detected 13 heterozygous missense variants with allele frequencies of $\leq 1\%$. Analyzing the burden of rare coding variants detected in our patients and in Japanese controls reported by iJGVD (Integrative Japanese Genome Variation; a public database), statistically significant association between carriers with allele frequency of $\leq 0.1\%$ and neuropsychiatric disorders were found. We performed *de novo* analysis when DNA samples from parents were available. All inheritance statuses were either from apparently unaffected parents or of unknown origin, suggesting variable penetrance. While none of the 13 variants scored more than 0.50 based on the *in silico* prediction tool REVEL (rare exome variant ensemble learner; an ensemble method for predicting the pathogenicity of missense variants), 62% of the rare variants (8/13) identified in this study were predicted to be protein-disrupting using SIFT and PolyPhen-2. Taken together, we regarded two missense variants (p.T1067I and p.G539R) as novel ones since each was not registered in either ExAC or iJGVD, the public database; however, we could not distinguish the clinical features of patients with p.T1067I or p.G539R and patients without the variants. At this stage, we could neither confirm nor dismiss the significance of rare *ASTN2* variants in either neuropsychiatric disorder. To assess the impact of the variants discovered here, further investigation by sample size expansion, case-control analysis and biological analysis will be performed. All procedures performed in this study involving human participants were approved by the Ethics Committee of the Nagoya University Graduate School of Medicine. Written informed consent was obtained from the participants and from the parents of the patients under 20 years old.

2419T

Shared rare recurrent copy number variations between autism spectrum disorder and attention deficit hyperactivity disorder. M.E. Khan^{1,2}, J.T. Glessner^{1,2}, X. Chang^{1,2}, Y. Liu^{1,2}, N. Williams^{1,2}, F.G. Otieno^{1,2}, M. Lemma^{1,2}, J. Li^{1,2}, C. Kao^{1,2}, P.M.A. Sleiman^{1,2}, H. Hakonarson^{1,2}. 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

Neurodevelopmental disorders (NDDs) such as autism spectrum disorder (ASD) and attention deficit hyperactivity disorder (ADHD) are representative medical cases often exhibiting elusive phenotypes that require rigorous examination prior to diagnosis. Structural variants have been implicated in NDDs with complex genetic associations exemplified by recurrent copy number variations (CNVs). Diseases broadly associated with similar cell type impairments are biologically hypothesized to share similar etiologies. Investigative platforms aimed to address genetic-based inquiries, such as high-density microarray technologies, have unraveled principles of disease biology. This development has allowed for the application of groundbreaking techniques in fields aimed at NDDs and other psychiatric diseases. Previous studies investigating NDD have uncovered CNVs associated with genes within shared genomic networks. In this study, we investigated 3,082 ASD and 4,243 ADHD cases matched to 11,134 controls with the aim of identifying CNVs conferring disease susceptibility and further elucidating shared disease biology across NDDs. Rigorous quality control through visual inspection of genotype and intensity was undertaken to ensure CNV confidence prior to association studies. To assess the association of metabotropic glutamate receptor (mGluR) network functions underlying NDD pathology, we exhaustively queried patients for CNVs associated with 273 genomic regions of interest defining the mGluR gene network (Elia et al, 2011). Critical CNV boundaries were detected with PennCNV and associated to disease status using ParseCNV software. We observed DPP6 deletions and duplications in 13 cases and no controls. Additionally we saw significant enrichment of CNTN4, GRM5 and GRM8 deletions in cases versus controls. Together, these results suggest disruption of neuronal cell-adhesion and transmission pathways confers substantial risk to multiple NDDs. We identified rare recurrent CNVs enriched in both ASD and ADHD cases showcasing an overrepresentation of these events in NDD cohorts.

2420F

Polygenic risk scores reveal subtypes of autism that differ in coding *de novo* mutational load. B. Lohman^{1,4}, H. Coon^{2,3}, G. Marth^{1,4}, A. Quinlan^{1,4}, A. Docherty², *Autism Sequencing Consortium*. 1) Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, UT; 2) Department of Psychiatry, University of Utah School of Medicine, Salt Lake City, UT; 3) Department of Biomedical Informatics, University of Utah School of Medicine, Salt Lake City, UT; 4) USTAR Center for Genetic Discovery, University of Utah School of Medicine, Salt Lake City, UT.

The rapid advance of DNA sequencing is poised to revolutionize research and treatment of human disease, both at the population and individual levels. To this end, the Simons Foundation Autism Research Initiative (SFARI) has made a major contribution by generating whole genome sequence for 2076 individuals in 4-member family units; each family containing parents and both a child diagnosed with Autism Spectrum Disorder (ASD; proband) and an unaffected sibling. These data have provided a powerful opportunity to map the genetic basis of ASD and discover genome-wide *de novo* mutations. However, very large GWAS have confirmed a complex and highly polygenic architecture for all major psychiatric disorders. Such complex traits are the product of many common loci of small effect, and few rare variants. Polygenic risk scores (PRS), aggregating the effects of thousands of loci, have been used to successfully subtype individuals into latent genetic classes, and to predict clinical symptoms and prognosis, not only for many cancers but also for psychiatric disorders, including schizophrenia and major depression. Here, using polygenic risk scores for 35 physical and psychological phenotypes (e.g. coronary artery disease, HDL/LDL, schizophrenia and major depression), we trained a decision tree model on Phase1 of SFARI (n = 519 families) to predict case status and to classify probands into distinct subtypes. Our model identified two distinct subtypes of probands, driven by high- and low-ASD PRS. We then tested for differences in the number of *de novo* mutations between subtypes, identifying significant enrichment for coding mutations in the low-ASD PRS subtype. This result is consistent with the hypothesis that larger effect, more penetrant *de novo* variants may require less burden of polygenic risk to result in expression of the ASD phenotype. Conversely, in cases with higher polygenic risk for ASD, such variants may be more elusive. Our work demonstrates that successful genetic subtyping of individuals, while accounting for critical covariates like sex and ancestry, produces distinct groups with more homogenous genetic backgrounds, simplifying downstream analysis and revealing previously unseen patterns that would be missed by traditional GWAS. We will present our efforts to replicate both our subtyping method and enrichment of *de novo* mutations in the low ASD PRS subtype in SFARI Phases 2, 3, and 4 (n = 1,868 families).

2421W

Cadherins matter to autism spectrum disorders: Which ones? M.R.

Passos-Bueno, C.I.S Costa, E.M. Montenegro, E.S. Moreira, S.S. Costa, N. Lourenço, C. Rosenberg, A. Krepischi, I.M.W. Silva. Centro de Pesquisa sobre o Genoma Humano, Instituto de Biociências, Universidade de São Paulo, São Paulo, São Paulo, Brazil.

Autism spectrum disorder (ASD) is a neurodevelopmental condition with a known strong genetic component. Genomic analyses have shown that *de novo* and rare loss-of function mutations are important for the etiology of ASD and oligogenic/multifactorial models are expected to explain most of the cases. It has also been shown that disruption of genes functionally important for neuronal cytoskeleton and synaptogenesis contributes to the phenotype, but one currently challenge is to identify which are the critical pathogenic genes. CNV analysis in a Brazilian cohort of 121 patients by chromosomal microarray analysis (CMA) revealed 31 potentially pathogenic CNVs. Of these, 26 contained 27 ASD candidate genes and 15 CNVs contained genes not yet implied in ASD patients. Two out of these 31 CNVs embraces genes of the cadherin family (*CDH11*, *CDH8* and *CDH13*). Screening of a Brazilian cohort composed by 1,029 patients with neurodevelopmental disorders revealed one CNV embracing *CDH13*, *TRIM16* and *PTPNR2* genes, thus reinforcing the association of *CDH13* with neurodevelopmental disruption. *CDH11* and *CDH8* have been pointed out as possible ASD candidate genes and are listed in SFARI database, but with minimum evidence (score 4; <http://sfari.org>), however, *CDH13*, which plays an important role in axonal pathfinding and synaptogenesis is not yet included in this database neither considered as a good ASD candidate gene. Our findings suggest that disruption of these cadherin genes may represent a predisposing hit in ASD and *CDH13* should be included as a novel candidate gene for this condition. FAPESP/CEPID, CNPq.

2422T

De novo duplication on chromosome 19 observed in nuclear family displaying several neurodevelopmental disorders. C. Sjaarda^{1,2}, B. Kaiser^{1,2}, A. McNaughton^{1,2}, M. Hudson^{1,2}, M. Ayub¹, A. Guerin³, X. Liu^{1,2}. 1) Psychiatry, Queens University, Kingston, Ontario, Canada; 2) Queen's Genomics Lab at Ongwanada (QGLO), Ongwanada Resource Center, Kingston, Ontario, Canada; 3) Pediatrics, Queen's University, Kingston, Ontario, Canada.

Pleiotropy and variable expressivity have been cited for the occurrence of seemingly distinct neurodevelopmental disorders due to a common genetic etiology within the same family. Here we present a family with a *de novo* 1 Mb duplication involving 18 genes on chromosome 19. Within the family there are multiple cases of neurodevelopmental disorders including Autism Spectrum Disorder, Attention Deficit and Hyperactivity Disorder, Intellectual Disability, and psychiatric disease in individuals carrying this Copy Number Variant (CNV). Quantitative PCR confirmed the CNV was *de novo* in the mother and inherited by both sons. Whole transcriptome analysis demonstrated ~1.5-fold increase in RNA transcript abundance in 14 of the 15 genes that are expressed in the blood of individuals carrying the CNV compared to relatives without the CNV. Examination of transcript abundance across the rest of the transcriptome identified 515 differential expressed genes (p value < 0.05) mapping to ribosome, immune response and cell adhesion pathways. These results suggest the observed CNV may result in varied phenotypic characteristics in family members through altered gene expression, and more broadly that different psychiatric disorders can share genetic variants which lay a framework for later psychiatric dysfunction.

2423F

Characterization of loss of Y chromosome in Autism spectrum disorder. S.P. Smieszek, Y. Song, W.S. Bush, J.L. Haines. Case Western Reserve University, Cleveland, OH.

Substantial evidence suggests that autism spectrum disorder (ASD) is caused by rare inherited or spontaneous genetic mutations, such as copy number changes and single nucleotide alterations. Nevertheless, the genetic underpinnings detected thus far account for small proportion of ASD. One striking characteristics of ASD is a strong gender bias, with males four times more likely to be on the spectrum than females. Typically, the Y chromosome has not been examined given its relative dearth of coding genes and its complex structure. In addition, the Loss of the Y chromosome (LOY) is the most frequent somatic alteration in men, and its loss and/or duplication could relevant to ASD. LOY already has been associated with shorter lifespan and multiple cancer types. We examined the structure of the Y chromosome (loss and gain) and the role it plays in ASD, focusing on the possibility that it could be a marker of genomic instability in the fathers of ASD cases. We examined full loss of the Y chromosome (LOY) and the mosaic LOY (mLOY) in 2,020 males from the ASD MSSNG whole genome sequencing dataset. We extracted DNA sequence depth information using equally spaced 10kb intervals across the Y chromosome and calculated individual log R Ratio (LRR) for every interval and median LRR for every subject. We detected significant LOY in the affected individuals when compared to controls. We additionally detect familial cases of 100% clonal duplication of the Y chromosome in several affected individuals including one father-son dyad. Furthermore we observe a correlation between *de novo* mutations in the sons and the mLOY in the fathers ($P < 1 \times 10^{-8}$). Moreover we investigated if the LOY phenomena could be a marker of genomic instability in the fathers of ASD individuals. We compared the mLOY distribution of fathers with a control set of fathers and found a significantly higher incidence of mLOY in fathers of ASD (adjusting for age) ($P < 0.003$). We also find greater mLOY incidence when comparing affected and unaffected brothers. We propose that mLOY in fathers is a marker of risk for ASD in their sons. This is the first such attempt to characterize the DNA sequence and structure of the whole Y chromosome, which provides a foundation for further investigation of this poorly characterized genomic region.

2424W

Association of RXFP3 gene polymorphism with co-occurring autism and obesity. Z. Talebizadeh^{1,2}, A. Shah¹, J. Noel-MacDonnell¹. 1) Children's Mercy Hosp, Kansas City, MO; 2) University of Missouri-Kansas City School of Medicine, Kansas City, MO.

Concern has been raised over the prevalence of obesity in children with ASD, but this potential co-occurring condition has not been fully studied in this population. The majority of patients with autism take antipsychotic medications, one of the main side effects of which is weight gain (i.e., antipsychotic-induced weight gain-AIWG). A general assumption is that the observed elevated rate of obesity in autism is caused by AIWG. The purpose of this study was to delineate the relationship between autism and obesity. We assessed the prevalence of known AIWG associated alleles in obese and non-obese autistic subjects in order to show that AIWG may not be the only reason for the observed higher rate of obesity. We curated a list of about n=150 SNPs with at least one report of association with AIWG. We assessed phenotypic and genotyping data from Simons Simplex Collection (SSC). We analyzed genotyping data for n=115 AIWG SNPs from probands from SSC. BMI data on probands was used to identify obese and non-obese subjects (> 1000 probands met the applied BMI cut off). Similarly, we evaluated unaffected SSC siblings stratified to obese and non-obese groups. Only one out of 115 AIWG SNPs (rs7702361) achieved an FDR corrected significant p-value, comparing obese versus non-obese autistic subjects. No difference was found for rs7702361, comparing obese versus non-obese unaffected siblings. Comparing obese probands stratified by Mood Stabilizers use did not detect any differences for rs7702361, suggesting that association of this SNP with obesity was independent of drug exposure. Similarly, we evaluated genotyping data for n=74 BMI SNPs. This analysis showed both shared (i.e., familial BMI/obesity risk factors) and distinct BMI associated SNPs in probands and siblings, which may indicate potential differences in the underlying mechanism of obesity in autistic and non-autistic family members. The results from our study shows, for the first time, that obesity in autism may be associated with implications in metabolic pathways including relaxin. rs7702361 is located in the RXFP3 gene, a member of the insulin/relaxin superfamily that regulates appetite and body weight. This gene is selectively activated by a neuropeptide, relaxin-3, modulating emotional-behavioral functions, including response to stress, appetite regulation, and anxiety. Further investigations are warranted on the potential role of the RXFP3 gene in association with co-occurring autism and obesity.

2425T

Insights into the genetic architecture of autism by whole genome sequencing analyses of >3000 families. T.N. Turner¹, B.P. Coe¹, K. Hoekzema¹, D.E. Dickel¹, B.J. Nelson¹, L.A. Pennacchio^{2,3}, T. Maniatis⁴, M.C. Zody⁴, E.E. Eichler^{1,5}. 1) Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA; 2) Functional Genomics Department, Lawrence Berkeley National Laboratory, Berkeley, CA; 3) U.S. Department of Energy Joint Genome Institute, Walnut Creek, CA; 4) New York Genome Center, New York, NY; 5) Howard Hughes Medical Institute, University of Washington, Seattle, WA.

Autism has a high heritability indicating a genetic component. Whole-exome sequencing (WES) and microarray analyses have begun to identify the pathogenic variant in 21% of cases. We have expanded our recent whole-genome sequencing analysis of 516 families to include an additional 1,271 quad (5,084 genomes) and 1,625 trio (4,875 genomes) families from the Simons Simplex Collection and MSSNG projects, respectively. We replicate differential *de novo* mutation (DNM) rates in unique (1.3×10^{-8}) versus ancient repeat (1.5×10^{-8}) portions of the genome (chi-squared test $p = 3.7 \times 10^{-189}$) irrespective of disease status. We identify a likely gene-disrupting (LGD) DNM false negative rate of ~3.9% in comparison to older exome data, often in the first exon ($p = 5.0 \times 10^{-10}$, OR = 6.1), including genes known to be associated with autism (SCN7A, CACNA1A, SYNGAP1, SHANK2, BIRC6, ANKRD11, GIGYF1, ZWILCH, and ADCY5). Combining these data with published WES (<http://denovo-db.gs.washington.edu>) assessed for 12,248 neurodevelopmental disorder trios, we identified eight additional genes enriched for DNM. Restricting our analysis to families that do not have any known LGD or large copy number variants (CNV), we focused on three functional categories of DNM: missense, untranslated region (UTR) and putative, noncoding regulatory regions that are also sites for transcription factor binding in fetal brain (pNCR TFBS). We observe DNM enrichment for severe missense ($p = 1.6 \times 10^{-3}$, OR = 1.7) and in genic untranslated regions (UTRs) ($p = 0.02$, OR = 1.5). We also observe significant enrichment for *de novo* events in putative regulatory regions of the fetal brain (pNCR TFBS; $p = 1.5 \times 10^{-3}$, OR = 1.2). If we restrict this analysis to genes previously implicated in autism (SFARI n = 845 genes), these three functional categories show a combined significance ($p = 2.9 \times 10^{-3}$, OR = 1.7). We also observe an excess of patients with multiple DNMs in different genes ($p = 0.03$, OR = 1.3) providing further support for a multifactorial model of disease. We are currently integrating these data with a high-resolution CNV map for each genome and working on a second cohort of 8,808 genomes from an additional 2,936 families in order to test for replication of these findings.

2426F

Large-scale targeted sequencing identifies risk genes for autism and developmental delay. T. Wang¹, K. Hoekzema¹, B.P. Coe¹, M.R. Geisheker¹, M.A. Gillentine¹, T.N. Turner¹, I.E. Scheffer^{2,3,4}, D.G. Amaral⁵, G. Santen⁶, E. Trabetti⁷, Z. Sedláček⁸, J.J. Michaelson⁹, K. Xia¹⁰, E. Courchesne¹¹, R.F. Kooy¹², A. Lindstrand^{13,14}, C. Romano¹⁵, H. Peeters¹⁶, H. Hakonarson^{17,18,19}, J. Gecz^{20,21,22}, R.A. Bernier²³, E.E. Eichler²⁴. 1) Department of Genome Sciences, University of Washington, Seattle, WA, USA; 2) Department of Paediatrics, University of Melbourne, Royal Children's Hospital, Melbourne, Victoria, Australia; 3) Department of Medicine, University of Melbourne, Austin Health, Melbourne, Australia; 4) The Florey Institute of Neuroscience and Mental Health, Parkville, Victoria, Australia; 5) MIND Institute and the University of California Davis School of Medicine, Sacramento, California, USA; 6) Department of Clinical Genetics, Leiden University Medical Center (LUMC), Leiden, the Netherlands; 7) Department of Neurosciences, Biomedicine and Movement Sciences, University of Verona, Italy; 8) Department of Biology and Medical Genetics, Charles University 2nd Faculty of Medicine and University Hospital Motol, Prague, Czech Republic; 9) Department of Psychiatry, University of Iowa Carver College of Medicine, Iowa City, IA, USA; 10) Center of Medical Genetics, School of Life Sciences, Central South University, Changsha, Hunan, China; 11) Department of Neurosciences, UC San Diego Autism Center, School of Medicine, University of California San Diego, La Jolla, California, USA; 12) Department of Medical Genetics, University of Antwerp, Antwerp, Belgium; 13) Department of Molecular Medicine and Surgery, Center for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden; 14) Department of Clinical Genetics, Karolinska University Hospital, Stockholm, Sweden; 15) Unit of Pediatrics & Medical Genetics, Oasi Research Institute-IRCCS, Troina, Italy; 16) Centre for Human Genetics, KU Leuven and Leuven Autism Research (LAuRes), Leuven, Belgium; 17) The Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA; 18) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA; 19) Divisions of Human Genetics and Pulmonary Medicine, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA; 20) School of Medicine and the Robinson Research Institute, the University of Adelaide at the Women's and Children's Hospital, Adelaide, South Australia, Australia; 21) Genetics and Molecular Pathology, SA Pathology, Adelaide, South Australia, Australia; 22) South Australian Health and Medical Research Institute, Adelaide, South Australia, Australia; 23) Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, USA; 24) Howard Hughes Medical Institute, University of Washington, Seattle, WA, USA.

Neurodevelopmental disorders (NDDs) are genetically and phenotypically heterogeneous. Ultra-rare and *de novo* gene-disruptive mutations have been shown to play an important role in disease etiology but most of the underlying genes have not yet been discovered. Based on a meta-analysis of *de novo* mutations from 10,927 published exomes of children with developmental delay (DD) and autism spectrum disorder (ASD) in denovo-db (<http://denovo-db.gs.washington.edu/denovo-db/>), we selected 65 candidate genes for targeted sequencing. These candidate genes include 43 with excess *de novo* mutations ($p < 2.5 \times 10^{-6}$), 14 genes with autism sex bias, 6 genes from a network analysis of high-functioning autism, and 2 candidate genes mapping to copy number variant genomic disorders. We designed 2,400 molecular inversion probes and targeted the protein-coding portion of these genes for sequencing in a cohort of 18,836 NDD cases (9,710 with ASD and 9,126 with DD), obtained from 17 centers across the world. After QC, we validated a total of 173 likely gene-disruptive mutations and identified 218 severe missense mutations (CADD score > 30) that were not present in a non-psychiatric set of control exomes from ExAC ($n = 45,376$). Controlling for sequence coverage differences, several genes (*CTCF*, *TCF12*, *SPEN*, *ZBTB18*, *HNRNPU*, *KAT6A*, *KIAA2022*, *CHAMP1*, *SATB2*, *AHDC1*, *KCNQ3* and *BRAF*) show evidence of increased mutational burden in cases compared to controls. We are now evaluating transmission of these mutations within the context of families and assessing phenotypic features to determine their association with high-functioning autism, genomic disorders, and gender biases with DD and ASD.

2427W

Exploring polygenic risk scores for depression in a transgender population. A. Pandit, G.R. Abecasis. University of Michigan, Ann Arbor, MI.

Transgender (trans) individuals are part of a highly stigmatized community for whom prejudice is commonplace. The 2011 National Transgender Discrimination Survey revealed that an estimated 70% of trans individuals have attempted to hide their identity to avoid discrimination in personal and professional settings. To help facilitate medical care for these individuals, the DSM added "gender dysphoria," a controversial term that references the discomfort and stress caused when one's birth gender and perceived gender are incongruent. Studies have found significantly higher rates of depression and anxiety in trans individuals, which is thought to stem from the overwhelming amounts of stigma and prejudice these people experience throughout their lives, oftentimes starting from a very young age. In the same national survey, 41% of transgender respondents answered that they had attempted suicide, compared to 1.6% of U.S. Americans. Known risk factors for depression in the trans community include discrimination, obstacles surrounding hormone therapy and sex-reassignment surgeries, socioeconomic factors, and access to health and social services. However, to date no research has explored genetic risk factors for depression in this population; we sought to examine these patterns using data collected from the Genes for Good study. Using demographics survey responses, we identified 103 presumptive trans subjects and 8,391 cisgender subjects. We then matched trans and cisgender subjects based on age, birth sex, education and income, matching 3 - 6 cisgender controls to each trans participant, resulting in 577 controls. As expected, trans subjects had increased odds of depression compared to cisgender controls: odds of answering "yes" to ever being depressed were significantly higher (95.5% of trans individuals, OR=4.51, $p=0.005$). However, when we calculated polygenic risk scores (PRS) of depression using GWAS results from the Psychiatric GWAS Consortium, we found no significant association between depression PRS and trans status. Our results support the idea that depression in the transgender community is influenced by societal factors and that it is a consequence, rather than a cause, of gender dysphoria and transgender identity. Further, our data illustrates how collecting information on both gender and birth sex separately allowed us to explore the interplay of health and genetics in a marginalized community.

2428T

The interaction between genetic variability and oral microbiome in alcohol use disorder and nicotine dependence. R. Polimanti^{1,2}, J.L. Montalvo-Ortiz^{1,2}, Y.Z. Nunez^{1,2}, S. Southwick^{1,2,3}, J.H. Krystal^{1,2,3}, H.R. Kranzler⁴, R.H. Pietrzak^{1,2,3}, J. Gelernter^{1,2}. 1) Department of Psychiatry, Yale University School of Medicine, West Haven, CT; 2) Veterans Affairs Connecticut Healthcare Center, West Haven, CT; 3) Clinical Neurosciences Division, U.S. Department of Veterans Affairs National Center for PTSD, West Haven, CT; 4) Department of Psychiatry, University of Pennsylvania School of Medicine and VISN 4 MIRECC, Crescenz VAMC, Philadelphia, PA.

Alcohol drinking and tobacco smoking have substantive effects on the microbiome. To understand whether microbiome variability may be associated not only with drinking and smoking states (i.e., present behaviors) but also with alcohol and tobacco traits (i.e., characteristic behaviors), we investigated the oral microbiome in individuals with alcohol use disorder (AUD) and nicotine dependence (ND) in 739 individuals from the Yale-Penn and NHRV (National Health and Resilience in Veterans) cohorts. Microbiome screening was performed via targeted sequencing of the microbial 16S rRNA gene in DNA samples extracted from saliva. Taxonomic annotation was completed using the SILVA database as reference. We conducted the association analysis using species-level information and applied two analytic approaches. We performed a single-species association analysis of relative quantity considering 14 species that were identified in >90% of the sample. We observed a negative association between AUD and the relative quantity of *Streptococcus sp. 11aTha1* (Beta = -0.26, P = 0.006). The second analytic approach consisted of investigating the overall profile of the microbiome species identified (N = 1,128) using principal component analysis. In this analysis, we considered the top 13 microbiome-based principal components. The association analysis showed a significant positive association between ND and the microbiome-based PC11 (Beta = 0.56, P = 4.08×10⁻⁴). To investigate further how AUD and ND interact with the inter-individual variability of the oral microbiome, we conducted a gene-by-environment genome-wide interaction study. We observed a genome-wide significant interaction between rs79424066 and ND with respect to microbiome-based PC11 (P = 2.25×10⁻⁴). Rs79424066 is associated with *NRXN1* expression in brain and non-brain tissues. *NRXN1* gene encodes cell-surface receptors that bind neuroligins to form a calcium-dependent neuroligin/neurexin complex at synapses in the central nervous system. This trans-synaptic complex is required for efficient neurotransmission and is involved in the formation of synaptic contacts. In summary, we report the first evidence of an association of AUD and ND with oral microbiome. Additionally, our genome-wide investigation indicated that ND may interact with variation in brain-relevant genes with respect to the oral microbiome.

2429F

Intersection of impulsive personality traits and drug experimentation: A genome-wide association approach. S. Sanchez-Roige¹, P. Fontanillas², S. Elson², The. 23andMe Research Team², J. Gray^{3,4,5}, H. deWit⁶, J. MacKillop⁷, A. Palmer^{1,8}. 1) University of California San Diego, San Diego, CA; 2) 23andMe, Inc., Mountain View, CA, 94041, USA; 3) Department of Psychology, University of Georgia, USA; 4) Center for Alcohol and Addiction Studies, Brown University School of Public Health, Providence, RI 02903, USA; 5) Center for Deployment Psychology, Uniformed Services University, Bethesda, MD 20814, USA; 6) Department of Psychiatry and Behavioral Neuroscience, University of Chicago, Chicago, IL 60637, USA; 7) Peter Boris Centre for Addictions Research, McMaster University/St. Joseph's Healthcare Hamilton, Hamilton, ON L8N 3K7, Canada; Homewood Research Institute, Guelph, ON N1E 6K9, Canada; 8) Institute for Genomic Medicine, University of California San Diego, La Jolla, CA, USA.

Impulsive personality traits are complex heritable traits associated with numerous neuropsychiatric disorders, particularly substance use disorders. In collaboration with the personal genetics company 23andMe, Inc., we performed several genome-wide association studies (GWAS) on measures of impulsive personality traits (Barratt Impulsiveness Scales, BIS; and Impulsive Behavior Scales, UPPSP) and drug experimentation (number of different classes of drugs an individual has used) among ~23,000 research participants of European ancestry. The impulsivity and drug experimentation traits showed SNP-heritabilities that ranged from 5 to 11%. Genetic variants of the gene *CADM2* were significantly associated with the UPPSP sensation seeking subscale (P = 8.3 × 10⁻⁹, rs139528938) and showed a suggestive association with drug experimentation (P = 3.0 × 10⁻⁷, rs2163971; r² = 0.68 with rs139528938). Both are highly correlated (r² = 0.69 and 0.57, respectively) with an intronic (rs57401290) SNP in *CADM2* that showed a genome-wide significant association with self-reported risk-taking propensity (rs57401290: p = 5.3 × 10⁻⁹), and nominal associations with the number of sexual partners (rs57401290: p = 6.0 × 10⁻⁷) and the number of children (rs57401290: p = 6.2 × 10⁻⁷) in the UKBiobank cohort. Multiple subscales from both BIS and UPPSP showed strong genetic correlations (>0.5) with substance use traits, including drug experimentation, smoking initiation, and lifetime cannabis use. Several BIS and UPPSP subscales were genetically correlated with attention-deficit/hyperactivity disorder (ADHD), supporting their use as endophenotypes for ADHD. The UPPSP sensation seeking subscale was genetically correlated with the personality trait extraversion. Several impulsivity traits from BIS and UPPSP were moderately correlated but not collinear, suggesting both common etiology and unique genetic variance among the impulsivity traits. Our findings demonstrate a role for common genetic contributions to individual differences in impulsivity. Furthermore, this study is the first to provide a genetic dissection of the relationship between types of impulsivity and various psychiatric disorders.

2430W

Genome-wide association study of posttraumatic stress disorder in female rape victims. C.M.F. Carvalho^{1,2,3}, M.L. Santoro^{1,2,3}, F. Talarico^{2,3}, B.M. Coimbra¹, G. Xavier^{2,3}, S. Nasser⁴, E.A. Zanardo⁴, L.D. Kulikowski⁴, A.F. Mello¹, V.K. Ota^{2,3}, M.F. Mello¹, S.I.N. Belangero^{1,2,3}. 1) Department of Psychiatry, Universidade Federal de São Paulo, São Paulo, Brazil; 2) Genetics Division of Department of Morphology and Genetics of Universidade Federal de São Paulo (UNIFESP), Brazil; 3) LiNC - Interdisciplinary Laboratory of Clinical Neurosciences of Universidade Federal de São Paulo (UNIFESP), Brazil; 4) Hospital das Clínicas, Faculty of Medicine, Universidade de São Paulo (FMUSP), Citogenomic Laboratory, Brazil.

Posttraumatic stress disorder (PTSD) is a clinical condition that can develop in people who have experienced a traumatic situation, such as sexual assault. Current results of PTSD GWAS (genome wide association study) have suggested that genetic variants may be associated with predisposition to PTSD. However, few studies have been able to identify a consistent genetic marker for PTSD. Therefore, we examined the cross-sectional association of PTSD diagnosis in female rape victim, using the GWA approach. This is the first study assessing GWAS with just rape victim women. The study population consisted of 51 rape victim women with PTSD diagnosis and 45 female healthy controls from Brazil. Assessments included MINI International Neuropsychiatric Interview compatible with Diagnostic and Statistical Manual of Mental Disorder, Clinician-Administered PTSD Scale that evaluate the symptoms of PTSD and Life Events Checklist that appraise the trauma exposure. DNA was extracted from blood and all individuals were genotyped using the Infinium Global Screening Array BeadChip (Illumina) that assesses around 600,000 markers. For association analysis we used logistic regression models adjusted to five principal-component and Bonferroni correction. After quality control of samples, two participants of the study were excluded due to identity-by-descent ($\text{pi}_{\text{hat}}=0.145$) and none individual was excluded due to missing genotype (<0.1). Furthermore 316,320 variants were removed due to missing genotype (<0.1), Hardy-Weinberg ($p>0.00001$), minor allele threshold (<0.05). Thus, remained for the analysis 50 cases (PTSD group) and 44 controls (no-PTSD group). No single nucleotide variants (SNVs) were associated with PTSD diagnosis, after Bonferroni correction ($n=324,912$) and genomic inflation lambda (based on median chi-square) was of 1.057. We did not find association between SNVs and PTSD. Previous studies found associations between SNV and PTSD but predominantly in military, however, few studies investigated genetic variants in women with PTSD even more victims of sexual abuse. Furthermore, it is difficult to detect genetic variants associated to mental disorders in small samples because of the limited power to detect discrete differences. We found negative results when assessing effects of SNV on the risk to develop PTSD probably due to our limited size sample. Therefore, our next goal is to increase the sample size and consequently the statistical power necessary for genome-wide approaches.

2431T

Exploratory genome-wide association analysis of response to ketamine and dissociative side effects in depression. Y. Yao¹, W. Guo¹, R. Machado-Vieira², S. Mathew³, J. Murrough⁴, D. Charney⁴, M. Gruenbaum⁵, M. Oquendo⁶, B. Kadriu⁷, N. Akula⁸, I. Henter⁹, P. Yuan⁹, K. Merikangas¹⁰, W. Drevets¹¹, M. Furey¹¹, J. Mann¹², F. McMahon⁸, C. Zarate⁷. 1) National Institute of Mental Health, Bethesda, MD; 2) Department of Psychiatry and Behavioral Sciences, University of Texas Health Science Center at Houston, Houston, TX; 3) Department of Psychiatry and Behavioral Sciences, Menninger; 4) Department of Psychiatry and Behavioral Sciences, Baylor College of Medicine, Houston, Texas; 5) Departments of Psychiatry and Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY; 6) Columbia University Medical Center/New York State Psychiatric Institute, New York, NY; 7) Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania; 8) Experimental Therapeutics and Pathophysiology Branch, National Institute of Mental Health, National Institutes of Health, Bethesda, MD; 9) Human Genetics Branch, National Institute of Mental Health, National Institutes of Health, Bethesda, MD; 10) Section on PET Neuroimaging Sciences, National Institute of Mental Health, National Institutes of Health, Bethesda, MD; 11) Genetic Epidemiology Branch, National Institute of Mental Health, National Institutes of Health, Bethesda, MD; 12) Janssen Pharmaceuticals, Neuroscience Research and Development, La Jolla, CA, USA; 13) Departments of Psychiatry and Radiology, College of Physicians and Surgeons, Columbia University, New York State Psychiatric Institute, New York, NY.

Growing evidence suggests that the glutamatergic modulator ketamine has rapid antidepressant effects in treatment-resistant depressed subjects. The anticholinergic agent scopolamine has also shown promise as a rapid-acting antidepressant. This study applied genome-wide markers to investigate the role of genetic variants in predicting acute antidepressant response to both agents. The ketamine-treated sample included 157 unrelated European subjects with major depressive disorder (MDD) or bipolar disorder (BD). The scopolamine-treated sample comprised 37 unrelated European subjects diagnosed with either MDD or BD who had a current Major Depressive Episode (MDE), and had failed at least two adequate treatment trials for depression. Change in Montgomery-Asberg Depression Rating Scale (MADRS) or the 17-item Hamilton Depression Rating Scale (HAM-D) scale scores at Day 1 (24 hours post-treatment) was considered the primary outcome. Here, we conduct pilot genome-wide association study (GWAS) analyses to identify potential markers of ketamine response and dissociative side effects. Polygenic risk score analysis of SNPs ranked by the strength of their association with ketamine response was then calculated in order to assess whether common genetic markers from the ketamine study could predict response to scopolamine. Findings revealed that genetic variants associated with ketamine response accounted for approximately 6% of the variance in scopolamine response, suggesting modest potential genetic overlap in predictors of response to these agents. These suggestive findings require replication in larger samples in light of low power of analyses of these small samples. Nevertheless, these data provide a promising illustration of our future potential to identify genetic variants underlying rapid treatment response in mood disorders and may ultimately guide individual patient treatment selection in the future.

2432F

Towards an understanding of the genetic architecture of the human cortex: Examining the effects of common variants on cortical thickness and surface area. S.E. Medland, *The Enhancing Neuro Imaging Genetics through Meta-Analysis (ENIGMA) Consortium*. Psychiatric Genetics, QIMR Berghofer, Herston, Brisbane, Queensland, Australia.

The cerebral cortex underlies our complex cognitive capabilities. Its structure shows variation in surface area and thickness within the population, which can be captured *in vivo* by magnetic resonance imaging. To improve understanding of neurobiological mechanisms underlying adaptive and pathological functioning, we undertook genome-wide association meta-analyses using MRI data from 51,246 individuals from 58 cohorts. To study the genetic architecture of the human cortex we analysed 70 MRI-derived cortical measures: surface area and thickness across the entire brain, and in 34 functionally specific regions averaged across the hemispheres. Across cohorts, structural T1-weighted MRI brain scans were analysed locally using harmonized analysis and quality-control protocols. Cortical parcellations were performed with freely available and validated segmentation software. Corrections for the summary measures were included in the regional measures to account for the omnibus effects of brain size. We identified 250 loci that met nominal genome-wide significance ($P < 5 \times 10^{-8}$; 218 influencing surface area and 32 influencing thickness) within 107 surviving multiple testing correction ($P < 8.3 \times 10^{-10}$). Across loci the strongest evidence of association was seen between variants within 15q14 and the surface area of the pre-central gyri (rs1080066, $P = 5.3 \times 10^{-122}$; variance explained = 0.87%). To localize this finding we examined the association of this loci with sulci depth in the functionally defined sensory-motor regions, finding that the association was strongest with a region of the central sulcus which shows highly variable patterns in MRI data from non-human primate species (analysed using the same protocols). We found significant enrichment for loci influencing total surface area within the germinal zone, present during foetal development. Loci impacting regional surface area clustered near genes involved in Wnt signalling, a pathway influencing progenitor expansion and areal identity. We observed significant positive genetic correlations between global Surface Area and IQ, educational attainment, and Parkinson's disease. Negative genetic correlations were evident between global SA and insomnia, ADHD, depressive symptoms, Major Depressive Disorder and Neuroticism. No significant genetic correlations were observed between global TH and any of the cognitive or neuropsychiatric traits assessed.

2433W

Integration of gene expression and chromatin accessibility data to improve understanding of the genetic architecture of bipolar disorder.

T. Schwarz¹, N. Mancuso², R. Ophoff^{3,4}, B. Pasaniuc^{2,3,5}. 1) Bioinformatics Interdepartmental Program, University of California Los Angeles, Los Angeles, CA USA; 2) Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA USA; 3) Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA USA; 4) Center for Neurobehavioral Genetics, University of California Los Angeles, Los Angeles, CA, USA; 5) Department of Biomathematics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA USA.

Describing the genetic architecture that contributes to highly polygenic disease traits in the brain such as Bipolar Disorder (BD) presents a major challenge. Because of the complexity of such traits, we have yet to develop a strong understanding of the causal mechanisms behind them. In this work we estimate the genetic basis of gene expression in of 200 cell lines (100 cases, 100 control for BD) and contrast them with risk of BD. We overlap eQTLs with GWAS for BD as well as ATAC sequencing analysis, among other techniques to study the genetic landscape. The combination of data from gene expression along with chromatin accessibility allows us to effectively outline the dynamic interaction between chromatin and transcripts, and gain new insight into the characterization of the disease-specific genetic risk of BD. Additional statistical analysis has helped us identify signature genes with promise to have significant implications in BD. Furthermore, our results provide evidence of potentially crucial pathways in the context of BD.

2434T

GWASs of alcohol consumption and alcohol use disorder in a multi-ancestry U.S. veteran population (N=274,424) yield both overlapping and differing risk loci. H. Zhou^{1,2,11}, H.R. Kranzler^{3,4,10}, R.L. Kember^{3,4}, R.V. Smith^{4,5}, A.C. Justice^{1,2,6}, S. Damrauer^{3,4}, D.M. Klarin^{7,8,9}, R. Polimanti^{1,2}, Z. Cheng^{1,2}, J. Reid¹⁰, J. Overton¹⁰, J.P. Tate², W.C. Becker^{1,2}, J. Concato^{1,2}, K. Xu^{1,2}, H. Zhao^{1,6}, J. Gelernter^{1,2}, VA Million Veteran Program. 1) Yale School of Medicine, New Haven, CT; 2) Veterans Affairs Connecticut Healthcare System, West Haven, CT; 3) University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; 4) Crescenz Veterans Affairs Medical Center, Philadelphia, PA; 5) University of Louisville School of Nursing, Louisville, KY; 6) Yale School of Public Health, New Haven, CT; 7) Boston VA Healthcare System, Boston, MA; 8) Massachusetts General Hospital, Boston, MA; 9) Broad Institute of MIT and Harvard, Boston, MA; 10) Regeneron Pharmaceuticals, Tarrytown, NY; 11) These authors contributed equally.

Although both alcohol consumption and alcohol use disorder (AUD) are heritable, little is known about their polygenic architecture or genetic overlap. We conducted genome-wide association studies (GWASs) of electronic health record (EHR) AUDIT-C scores (measuring consumption) and ICD-9/10 AUD diagnoses in a multi-ancestry Million Veteran Program (MVP) sample of 274,424 individuals. Meta-analysis across the 5 population groups identified 13 independent loci (9 novel) for AUDIT-C and 8 independent loci (5 novel) for AUD. Although 5 loci contributed to both traits (*ADH1B*, *ADH1C*, *SLC39A8*, *GCKR* and *FTO*), there were 8 distinct associations for AUDIT-C (e.g., *IGF2BP1*: insulin like growth factor 2 mRNA binding protein 1) and 3 for AUD (e.g., *DRD2*: D2 dopamine receptor), indicating both genetic overlap and differences between the traits. Partitioning heritability using LD score regression (LDSC) showed significant enrichment in central nervous system and in epigenetic features in brain tissues for both traits, while enrichments for AUDIT-C were also detected in cardiovascular and other cell types. The genetic correlation (r_g) between the traits was 0.522 (se=0.0383) in EAs and 0.9304 (se=0.1215) in AAs. Genetic correlations with some other traits were in opposite directions for the two alcohol traits: AUDIT-C was negatively correlated with anthropometric (e.g., BMI, r_g =−0.35, p =3.25×10^{−10}), cardiometabolic and glycemic traits and major depressive disorder (MDD, r_g =−0.22, p =7.72×10^{−9}), and positively correlated with reproductive traits and educational attainment (r_g =0.25, p =1.40×10^{−16}); AUD was negatively correlated with cognitive traits and educational attainment (r_g =−0.27, p =2.88×10^{−15}), and positively correlated with psychiatric disorders (e.g., MDD, r_g =0.41, p =2.19×10^{−20}). Phenome-wide association studies of phenotypes using MVP EHR data for the top variants showed many to be highly pleiotropic. The polygenic risk scores (PRSs) for AUDIT-C and AUD were significantly associated with alcoholism (103 cases and 6,132 controls) and alcohol-related disorders (137 cases and 6,132 controls) in an independent sample (Penn Medicine BioBank). Based on a comparison of two alcohol-related phenotypes in a large, relatively homogeneously ascertained cohort, including groups representing 5 different ancestries, we conclude that heavy drinking, while a necessary precondition for developing AUD, is not sufficient to develop the disorder.

2435F

Excess of missense *de novo* mutations in voltage-gated ion channels in neurodevelopmental disorders. M.R. Geisheker¹, T.N. Turner¹, K. Hoekzeema¹, B.P. Coe¹, T. Wang¹, B. Juarez², I. Scheffer^{3,4,5}, D. Amaral⁶, G.W.E. Santen⁷, Z. Sedlacek⁸, J.J. Michaelson⁹, K. Xia¹⁰, E. Courchesne¹¹, R.F. Kooy¹², M. Nordenskjöld^{13,14}, C. Romano¹⁵, H. Peeters¹⁶, H. Hakonarson^{17,18,19}, J. Gecz^{20,21,22}, R.A. Bernier²³, L.S. Zweifler^{2,23}, E.E. Eichler^{1,24}. 1) University of Washington School of Medicine, Seattle, WA; 2) Department of Pharmacology, University of Washington, Seattle, WA, USA; 3) Department of Paediatrics, University of Melbourne, Royal Children's Hospital, Melbourne, Victoria, Australia; 4) Department of Medicine, University of Melbourne, Austin Health, Melbourne, Australia; 5) The Florey Institute of Neuroscience and Mental Health, Parkville, Victoria, Australia; 6) MIND Institute and the University of California Davis School of Medicine, Sacramento, California, USA; 7) Department of Clinical Genetics, Leiden University Medical Center (LUMC), Leiden, the Netherlands; 8) Department of Biology and Medical Genetics, Charles University 2nd Faculty of Medicine and University Hospital Motol, Prague, Czech Republic; 9) Department of Psychiatry, University of Iowa Carver College of Medicine, Iowa City, IA, USA; 10) Center of Medical Genetics, School of Life Sciences, Central South University, Changsha, Hunan, China; 11) Department of Neurosciences, UC San Diego Autism Center of Excellence, School of Medicine, University of California San Diego, La Jolla, California, USA; 12) Department of Medical Genetics, University of Antwerp, Antwerp, Belgium; 13) Department of Molecular Medicine and Surgery, Center for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden; 14) Department of Clinical Genetics, Karolinska University Hospital, Stockholm, Sweden; 15) Unit of Pediatrics & Medical Genetics, Oasi Research Institute-IRCCS, Troina, Italy; 16) Centre for Human Genetics, KU Leuven and Leuven Autism Research (LAuRes), Leuven, Belgium; 17) The Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA; 18) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA; 19) Division of Pulmonary Medicine, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA; 20) School of Medicine and the Robinson Research Institute, the University of Adelaide at the Women's and Children's Hospital, Adelaide, South Australia, Australia; 21) Genetics and Molecular Pathology, SA Pathology, Adelaide, South Australia, Australia; 22) South Australian Health and Medical Research Institute, Adelaide, South Australia, Australia; 23) Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, USA; 24) Howard Hughes Medical Institute, Seattle, WA, USA.

Missense *de novo* mutations (DNMs) account for more cases of neurodevelopmental disorders (NDDs) than likely gene-disruptive (LGD) mutations but have been less studied because of their high incidental rate. Missense-specific genes likely account for much of the missing heritability of NDDs and some may be genes with such critical functions that LGD mutations are incompatible with life. Missense mutations also contribute to pathogenicity via gain-of-function effects and are more likely to occur at functionally important locations. To gain statistical power in gene identification, we studied 9,712 missense DNMs in 11,459 individuals with autism spectrum disorder, developmental delay, intellectual disability, or epilepsy, as these NDDs share genotypic and phenotypic components. We identified 220 genes with missense DNMs that are more clustered ($p < 0.05$) in NDD cases than population controls. This set of genes is enriched for roles in the neuronal action potential ($p = 2.92 \times 10^{-2}$, FDR corrected). We developed molecular inversion probes targeting clustered regions of these genes and 24 genes with recurrent mutations for sequencing in 18,836 additional NDD cases. We have identified 415 missense events in clustered regions in 8,960 cases, including 17 events at 13 targeted recurrent sites and 60 events at amino acids that bear mutations in other NDD cases. Of the newly identified events, 23 fall in seven voltage-gated ion channels; three of these are in functionally important transmembrane domains. Out of 176 missense DNMs previously seen in 51 of these channels in NDD cases, 85 (48%) occur in these domains or the channel pore, and there is a significant enrichment for mutations in the S4 segment ($p = 1.85 \times 10^{-13}$). Notably, 24/29 (83%) S4 events occur at arginine or lysine residues that provide the molecular mechanism for voltage sensing and channel gating. None of these missense DNMs have been seen in population controls in ExAC (N = 45,376). Eight of the observed substitutions are known to be pathogenic, and while the remaining 16 are novel, homology suggests that they are similarly damaging. The specific effect of these mutations makes them potential targets for therapeutic intervention, and these findings demonstrate the importance of identifying new classes of NDD risk genes by investigating the pattern of missense DNMs.

2436W

Molecular effects of the neuropsychiatric 15q13.3 microdeletion in an induced pluripotent stem cell model of neuronal development. S. Zhang^{1,2}, C. Purmann^{1,2}, X. Zhang^{1,2}, S. Ma³, K. Davis^{1,2}, J. Bernstein⁴, W.H. Wong³, J. Hallmayer², A.E. Urban^{1,2}. 1) Department of Genetics, Stanford University, Stanford, CA; 2) Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA; 3) Department of Statistics, Stanford University, Stanford, CA; 4) Department of Pediatrics, Stanford University, Stanford, CA.

The 1.5 Mb microdeletion on chromosome 15q13.3 is strongly associated with several neurodevelopmental and neuropsychiatric disorders, including developmental disability, seizures, schizophrenia and autism spectrum disorder. There are at least eight genes affected by the microdeletion, and quite possibly in a manner specific to tissue and developmental time point. Novel approaches are needed to analyze the deletion's effects on the molecular and cellular level. We created a cohort of induced pluripotent stem cells (iPSCs), from fibroblasts from three patients with the 15q13.3 deletion, and from matching controls. The iPSCs' genomes have been characterized with digital droplet PCR and whole-genome sequencing, and the iPSCs were characterized by staining for the canonical pluripotency markers. Furthermore, we induced the iPSCs into an early neuronal cell type (induced Neurons, iNs). The iNs stained positive for neuronal identity markers and also show neuronal gene expression patterns. RNA-Seq, ATAC-Seq, and genome-wide capture bisulfite sequencing was carried out for the iPSC and iN stages. We saw mostly decreased gene expression levels, for the genes within the deletion boundaries, while the patterns of DNA methylation and chromatin accessibility within the deletion remain mostly unchanged. Outside the CNV locus, transcriptome wide, there are differentially expressed genes (DEGs) on all chromosomes in the 15q13.3 cells. Specifically, there are 178 DEGs in the iPSCs and 369 DEGs in the iNs. Analyses of these DEGs using Ingenuity Pathway Analysis showed alterations in neurodevelopmental pathways, most notably the Wnt pathway. DNA methylation analysis revealed differentially methylated regions (DMRs) spread out genome-wide and also implicated neurodevelopmental genes and gene families, in particular the protocadherins. Analysis of chromatin state showed regions of differential chromatin openness genome-wide as well. In summary, we have created a cohort of iPSCs and iPSC-derived induced neurons from patients with the neuropsychiatric 15q13.3 microdeletion. We can detect gene expression changes for both the genes within the CNV locus as well as transcriptome-wide. Epigenomic marks are generally unchanged within the CNV locus, which correlates with the gene dosage dependent changes in gene expression there, while genome-wide there are marked effects on the levels of epigenomic regulation in the 15q13.3 patient lines.

2437T

Genomics of sleep disordered breathing associated with cognitive behavioral development in the Canadian Healthy Infant Longitudinal Development (CHILD) study. A. Ambalavanan¹, J. Choi¹, A. Lamri², S.K. Tamana³, D. Lefebvre², M. Sears², M. Azad⁴, A. Becker⁵, S. Turvey⁶, T. Moraes⁷, P. Subbarao^{6,7}, P. Mandhane³, Q.L. Duan^{1,8}. 1) Department of Biomedical and Molecular Sciences, Queens University, Kingston, ON, Canada; 2) Dept of Clinical Epidemiology & Biostatistics, McMaster University, Hamilton, ON, Canada; 3) Department of Pediatrics, University of Alberta, Edmonton, AB, Canada; 4) Department of Pediatrics & Child Health and Community Health Sciences, University of Manitoba, Winnipeg, MB, Canada; 5) Director, Clinical Research, BC Children's Hospital, Vancouver, BC, Canada; 6) Department of Paediatrics, University of Toronto, Toronto, ON, Canada; 7) The Hospital for Sick Children, Toronto, ON, Canada; 8) School of Computing, Queen's University, Kingston, ON, Canada.

Sleep disordered breathing (SDB), a collective term for chronic conditions including habitual snoring and obstructive sleep apnea, affects up to 10% of children between 2 and 8 years old. Studies have shown that SDB in children is associated with neurobehavioral functions related to executive functioning and attention deficit hyperactivity disorder (ADHD). We hypothesize that pre-school SDB may interact with genetic polymorphisms to determine predisposition to cognitive function outcomes during early childhood. In the Canadian Healthy Infant Longitudinal Development (CHILD) study, we selected a total of 513 Caucasian subjects (between 3 months and 5 years) with available parent reported Sleep Related Breathing Disorder (SRBD) subscale preschool SDB in addition to cognitive behavioural externalizing & internalizing checklist (CBCL) scores. In this study, we conducted a genetic analysis using genomics data (Illumina HumanCore Exome BeadChip) of CHILD subjects, which targeted top-ranked 40 autosomal genes previously correlated with Autism Spectrum Disorder (ASD) risk (Krishnan et al 2016, Nat Neurosci). Using a linear regression analysis model, we tested 46,615 SNPs for correlation with CBCL scores of children at age 5 years by including SDB as a covariate in the additive model. Single variant analysis (minor allele frequency over 5%) of CBCL scores resulted in two significant associations with additive effects with SDB: rs62009932 in *RBFOX1* ($\beta=2.8$, $p=2.47 \times 10^{-5}$) and rs2786239 in *GRIK2* ($\beta=4.06$, $p=2.6 \times 10^{-5}$), which are intronic SNPs correlated with externalizing and internalizing CBCL scores, respectively. On-going analysis include a multivariate test of the ASD genes with multiple cognitive behavioural traits to investigate pleiotropic effects. Furthermore, we will further explore gene-environmental interactions using assessments such as parental SDB, sleep duration, apnea-hypopnea index, sleep habits, and physical activity as exposures. Our study has identified that genes correlated with ASD may also contribute to cognitive behavioural outcomes in children during early childhood and that SDB could modulate the genetic predisposition to cognitive behavioural development in children.

2438F

A genome-wide association study of suicidality in the UK Biobank cohort reveals two genome-wide significant loci and substantial genetic correlation with depression-related traits. M.E.S. Bailey¹, J. Ward², R. Shaw², N. Graham², R.J. Strawbridge^{2,3}, A. Ferguson², L.M. Lyall², B. Cullen², D.M. Lyall², R. Pearsall², J.P. Pell², D.J. Smith². 1) School of Life Sciences, CMVLS, University of Glasgow, Glasgow, United Kingdom; 2) Institute of Health and Wellbeing, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK; 3) Department of Medicine Solna, Karolinska Institute, Stockholm, Sweden.

Suicidal acts and ideation are a global public health priority, with up to 1 million individuals worldwide dying by suicide each year. Almost all individuals who attempt suicide also have major depressive disorder (MDD), but not all individuals with MDD experience suicidal ideation. Heritability estimates for suicidal behaviour are mostly in the range 30-55%. We investigated the genetic contribution to suicidality by carrying out the largest genome-wide association study (GWAS) to date in 122,822 White British individuals from the UK Biobank (UKB) population cohort ($n = 39,265$ indivs. with suicidal/self-harming ideation/behaviour). An ordinal phenotype variable was created comprising 5 increasing levels of suicidal and self-harming ideation and behaviour - 0: no self-harm ideation or action; 1: thoughts that life is not worth living; 2: contemplation of self-harm or suicide; 3: actual self-harm; 4: attempted suicide. A model was run using the *clm* function in the 'ordinal' package in R, assuming additive allelic effects at each SNP and adjusting for 8 genetic PCs, age, sex and genotyping chip. LDSR analysis of the resulting summary statistics suggested a significant polygenic contribution and SNP heritability was estimated as 7.6%. Two loci, including a total of 36 SNPs, were associated with the outcome phenotype at genome-wide significance. One locus, on chr.11 (lead SNP: rs598046; $p = 1.1 \times 10^{-8}$), was located within *CNTN5*, a brain-expressed gene encoding an immunoglobulin family protein involved in neuronal cell interactions. Suggestive evidence has been reported for association between *CNTN5* SNPs and both loneliness and response to lithium treatment in bipolar disorder. The other locus, on chr.9 (lead SNP: rs62535711; $p = 1.3 \times 10^{-8}$), was located within the *ZCCHC7* gene. Significant genetic correlations were found with MDD ($r_g=0.63$), neuroticism ($r_g=0.63$), mood instability ($r_g=0.50$) and trait anxiety ($r_g=0.75$). These findings support a biological basis for the known correlation between suicidal behaviour/ideation and neuroticism/MDD, but cannot distinguish shared pleiotropic variants from mediated effects of neuroticism/depression-related variants, nor genes for suicidality *per se* from genes predisposing to other forms of self-harm with effects captured by the UKB questionnaire. Further understanding of the biology of suicidality will inform the development of more effective preventative approaches aimed at reducing these potentially devastating outcomes.

2439W

Significant overlap of genomic regions from extended high-risk autism families with evolutionarily accelerated regions in a species with highly unusual social behavior. H. Coon¹, E. Ferris², T.M. Darlington¹, J. Morgan¹, R. Sargent³, A. Fraser⁴, K.R. Smith⁴, N.J. Camp³, C. Gregg². 1) Psychiatry, University of Utah School of Medicine, Salt Lake City, UT; 2) Neurobiology and Anatomy, University of Utah School of Medicine, Salt Lake City, UT; 3) Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City, UT; 4) Population Sciences, Huntsman Cancer Center, Salt Lake City, UT.

Introduction. Genetic discoveries for Autism Spectrum Disorder (ASD) have been numerous in the past decade, with evidence for over 100 common and rare variants. However, these discoveries account for only a portion of the genetic risk of ASD. Much of this remaining risk may reside in notoriously elusive variants controlling gene regulation. To address this substantial knowledge gap, we compared genomic results obtained from extended families at high risk for ASD with genomic regions exhibiting accelerated evolution in mammalian species with specific evolved characteristics, including a species with highly unusual social behaviors. **Methods.** Using a novel statistical analysis strategy (Shared Genomic Segments), we discovered 10 genomic regions with significant evidence of familial sharing in large, high-risk ASD families. ASD was diagnosed using gold standard clinical interviews and observation. Genotypes from the Illumina OmniExpress array were used to identify significant genomic segments shared by ASD cases in the high-risk families. Resulting significant shared segments were then interrogated for harboring regions of accelerated evolution in mammalian species that evolved various distinctive traits, including elephant, microbat, orca, dolphin, naked mole rat (NMR), ground squirrel, hamster, orangutan, and prairie vole. Evolved traits can map onto phenotypes of interest for human disease; for example, cancer resistance in elephants. **Results.** Shared genomic segments that segregate in high-risk ASD families were found to harbor significantly more accelerated regions (ARs) in the NMR than expected by chance ($p=1.84E-14$). The NMR is one of two known eusocial mammals. The result was robust to simulations accounting for nonrandom genomic distribution of ARs. **Conclusion.** We show that genomic regions linked to ASD risk are significantly enriched for ARs specifically in NMR, a species that evolved highly unusual social behaviors. Results reveal novel candidate genomic regions for shaping mammalian sociability. Our results establish a combined phylogenomics and human genetics approach to identify new elements of genomic control for an aspect of human behavior strongly associated with autism and other psychiatric disorders.

2440T

Assessment of shared genetic sex-differences across neuropsychiatric and behavioral traits. E.A. Khrantsova^{1,2}, J. Martin³, L.K. Davis^{4,5,6}, B.E. Stranger^{2,7}, *Psychiatric Genomics Consortium Sex-specific cross-disorder analysis group*. 1) Department of Medicine, The University of Chicago, Chicago, IL; 2) Institute for Genomics and Systems Biology, University of Chicago, Chicago, IL; 3) MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, UK; 4) Vanderbilt Genetics Institute; Vanderbilt University Medical Center, Nashville, TN; 5) Division of Medical Genetics, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN; 6) Department of Psychiatry and Behavioral Sciences, Vanderbilt University Medical Center, Nashville, TN; 7) Center for Data Intensive Science, University of Chicago, Chicago, IL.

Accumulating evidence suggests that widespread phenotypic sex differences in complex traits may include a genetic component beyond sex chromosomes and hormones. Understanding the contribution of sex to complex traits and disorders may have important implications for diagnosis and treatments. Because neuropsychiatric traits moderately share genetic components (Anttila et al., 2016) and are often comorbid, we examined the extent to which males and females share the genetic risk for a range of neuropsychiatric and behavioral traits. Using sex-stratified GWAS summary statistics and the LDSC method, we first assessed genetic correlation between pairs of traits within each sex across several neuropsychiatric and behavioral phenotypes. Next, we assessed whether the genetic correlation of each pair of phenotypes differed significantly between sexes. Overall, the within-sex genetic correlation estimates across each pair of traits were largely similar in males and females. One notable exception is the genetic correlation between educational attainment and alcohol consumption which was significantly higher in females (females: $R_g(SE)=0.28(0.05)$; males: $R_g(SE)=0.04(0.05)$, difference z -score=3.55, FDR-corrected $p=1.1e-2$). This finding was replicated in an independent alcohol consumption dataset (difference z -score=2.27, uncorrected $p=2.3e-2$). For traits with moderate genetic correlation, we performed sex-specific cross-trait meta-analyses then identified variants with sex-heterogeneous effects and characterized their biological function. Finally, we investigated the effect of genetic risk for psychiatric disorders in a large hospital population (BioVU) by calculating polygenic risk scores (PRS) for each disorder. We then performed a phenome-wide association study of the PRS separately by sex. We hypothesized that different clinical manifestations of genetic risk for psychiatric disorders may be more easily observed in a hospital population when individuals have not been ascertained into the study for a specific set of behaviors. Consistent with this, preliminary results from BioVU suggest that a cross-disorder PRS was most significantly associated with substance use disorders in males ($p=5.24e-06$) and most significantly associated with mood disorders in females ($p=5.41e-05$). These results elucidate important similarities and differences across sex in the common variant contribution to neuropsychiatric and behavioral phenotypes.

2441F

Genome-wide meta-analyses reveal novel loci for verbal short-term memory and learning and show genetic correlation with schizophrenia, Alzheimer's disease, and type 2 diabetes. J. Lahti¹, S. Tuominen¹, Q. Yang², S. Medland³, J. Painter³, Z. Gerring³, E. Derks³, J. Bressler⁴, T. Mosley Jr⁵, C. van Duijn⁶, L.J. Launer⁷, S. Debette⁸, I. Deary⁹, S. Seshadri⁹, K. Räikkönen¹, *CHARGE Consortium Cognitive Working Group*. 1) Department of Psychology and Logopedics, University of Helsinki, Helsinki, Finland; 2) Boston University, USA; 3) QIMR Berghofer Medical Research Institute, Australia; 4) University of Texas, USA; 5) University of Mississippi Medical Center, The Netherlands; 6) Erasmus Medical Center University Medical Center, The Netherlands; 7) National Institutes of Health, USA; 8) University of Bordeaux, France; 9) University of Edinburgh, UK.

Objectives: Understanding the biological basis of memory functions may help combat neurodegenerative disorders. Despite moderate heritability in twin studies, consistent findings on the molecular genomic basis of memory functions are scarce. This reflects low power in most of the studies. In a large genome-wide association study (GWAS) based on the CHARGE consortium we investigated common genetic variants associated with short-term verbal memory and learning in adults. **Methods:** We conducted meta-analyses for verbal memory immediate recall test scores and verbal learning scores in a discovery sample across 27 cohorts ($N=44,874$) and 22 cohorts ($N=28,909$), respectively. Respective sample sizes of the replication samples were $N=8763$ and $N=3853$. All participants were adults of European descent and were free of dementia and stroke. **Results:** For verbal memory immediate recall test scores, we observed a genome-wide significant signal in an intron of *CDH18* and for verbal learning, we observed a signal peak in the 3p21.1 region with a large LD block in/near *NT5DC2*, *STAB1*, *ITIH1*, *ITIH4*, and *PBRM1*. For both phenotypes, we also observed a signal near *APOC1*. Associations of lead SNPs in all regions replicated ($p<.05$). MAGMA gene-based analyses implicated an additional nine genes for verbal learning (*CALN1*, *TOMM40*, *SMIM4*, *NEK4*, *GNL3*, *AGXT2*, *MUSTN1*, *GLT8D1*, and *ITIH3*). Several SNPs in the 3p21.1 region are eQTLs for *POC1A*, *GNL3*, *GLYCTK*, *DUSP7*, *ITIH4*, *PPM1M*, and *GLT8D1* in brain tissues in the GTEx or Brainiac databases. S-PrediXcan transcriptome-wide analyses showed that *POC1A* expression in the putamen was associated with lower verbal learning scores. Finally, we show with LDscore regression a negative genetic correlation of schizophrenia and Type 2 Diabetes (T2D) to both memory traits; a negative genetic correlation with anxiety and neuroticism for verbal immediate recall; and with Alzheimer's disease (AD) for verbal learning. **Conclusions:** We showed novel genomic regions in cadherin 18 (*CDH18*) and 3p21.1 that associated with verbal short-term memory and learning, respectively, and we replicated a previously reported finding in *APOC1*. Cadherins are crucial in synaptic plasticity and the 3p21.1 region has been implicated in psychiatric disorders in previous studies. Our results may also partly explain phenotypic links between memory and psychiatric disorders, T2D, and AD.

2442W

Population-scale analysis of copy number variation from whole genome sequencing. R.E. Handsaker^{1,2}, S. Kashin^{1,2}, C. Usher^{1,2}, C. Whelan^{1,2}, M.T. Pato³, C.N. Pato³, M. Boehnke⁴, S.A. McCarroll^{1,2}, WGSPD Consortium. 1) Stanley Center for Psychiatric Research, Broad Institute, Cambridge, MA, 02141 USA; 2) Department of Genetics, Harvard Medical School, Boston, MA, 02115 USA; 3) SUNY Downstate Medical Center, Brooklyn, NY 11203 USA; 4) Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI 48109 USA.

Human genomes exhibit abundant structural variation; this class of variation has been shown to impact many important human phenotypes. However, the full spectrum of structural variation in humans has not been as deeply or as thoroughly ascertained as fine-scale variation (e.g. SNPs and small indels), especially in large populations. We have performed an initial analysis of copy number variation – including deletions, duplications and multi-allelic CNVs – from whole-genome sequencing a cohort of 11,627 individuals from the Whole Genome Sequencing in Psychiatric Disorders (WGSPD) consortium. This cohort is comprised of schizophrenia and bipolar disorder cases, controls and unaffected family members and includes individuals of diverse ancestry including more than 6,500 African Americans and 250 Latinos. Accurate and efficient methods for ascertaining copy number variation at this scale from whole-genome sequencing data have been needed. To address this need, we have enhanced our Genome STRIP research software to efficiently process cohorts of this size in a cloud-computing environment. To date, we have ascertained more than 60,000 copy number variants of length greater than one kilobase, including variants in many repetitive regions of the genome. We have performed uniform genotyping of these variants across the cohort to allow accurate characterization of allele frequencies and to place the structural alleles on SNP haplotypes. Most of this variation appears to be segregating in human populations at allele frequencies below 1%. We estimate that 58% of all protein-coding genes are impacted by at least one copy number variant ascertained in this cohort. This set of broadly genotyped CNVs will make it possible to begin to understand the patterns of rare, segregating copy number variation in large populations and across different ancestry groups. To facilitate access to this resource, we have developed an interactive web interface to allow researchers to search and explore this data set and view the evidence for each variant.

2443T

3q29 deletion syndrome: Neuropsychiatric and behavioral phenotypes. R.M. Pollak¹, M.M. Murphy², M.P. Epstein², M.E. Zwick², C.E. Saulnier³, J.G. Mullen². 1) Genetics and Molecular Biology Program, Laney Graduate School, Emory University, Atlanta, GA; 2) Department of Human Genetics, Department of Medicine, Emory University, Atlanta, GA; 3) Department of Pediatrics, Department of Medicine, Emory University, Atlanta, GA.

3q29 deletion syndrome (3q29DS) is a rare (~1:30,000) genomic disorder characterized by a 1.6 Mb heterozygous deletion on chromosome 3 and is associated with a wide range of phenotypes, including neurodevelopmental and neuropsychiatric disorders, growth deficits, feeding problems, and congenital heart defects. Notably, the 3q29 deletion confers a 40-fold increased risk for schizophrenia, a 30-fold increased risk for Autism Spectrum Disorder (ASD), and a significantly increased risk for intellectual disability and generalized anxiety disorder. In order to develop a more comprehensive picture of ASD in 3q29DS, we used Emory University's 3q29DS registry (3q29deletion.org) to collect self-report data on 3q29DS patients and typically developing controls, including a custom medical and demographic questionnaire (n=94 3q29DS, 64 control); Achenbach Behavior Checklists (CBCL/ABCL, n=48 3q29DS, 57 control); Social Responsiveness Scale (SRS, n=48 3q29DS, 59 control); and the Social Communication Questionnaire (SCQ, n=33 3q29DS, 46 control). Statistical testing and data visualization were performed in R. Self-report data from 3q29DS registry participants shows increased prevalence of ASD diagnosis versus the general population (30.7% vs. 1.47%, $p < 2.2 \times 10^{-16}$). The male:female ratio of self-reported ASD diagnosis in 3q29DS is 1.8:1, strikingly different from the 4:1 ratio in the general population. Additionally, some participants that do not report an ASD diagnosis scored in the clinical range on all self-report scales, indicating substantial impairment in the absence of a clinical diagnosis. Finally, 3q29DS individuals show a novel constellation of ASD features, including less compromised social motivation compared to other domains, which is dramatically different from the profile observed in idiopathic ASD. These results show that the 3q29DS population is significantly enriched for ASD diagnosis and ASD features measured via standardized ASD symptomology surveys. Our findings suggest ASD evaluation should be the standard of care for individuals newly diagnosed with 3q29DS. Our findings also suggest that cognitive behavioral therapy strategies may be successfully deployed in 3q29DS, by leveraging the relatively intact social motivation of 3q29DS individuals.

2444F

In silico evidence for the contribution of expression dysregulation of 16p11.2 genes to schizophrenia. M. Vysotskiy, L.A. Weiss. Department of Psychiatry, Weill Institute for Neurosciences, Institute for Human Genetics, University of California San Francisco, San Francisco, California.

A copy number variant at 16p11.2 is strongly associated with behavioral disorders. Schizophrenia, for example, is observed more frequently in duplication, but not deletion, carriers. Not every gene in the 30-gene region may influence schizophrenia; however, finding a set of primary driver genes for behavioral phenotypes remains unsolved. Duplications are likely to increase transcription levels, a potential mechanism for their pathogenicity. We hypothesize that in the general population (individuals carrying two copies of each CNV gene), psychiatric disease risk will also be affected by expression variation in these genes, but in more modest ways. The effect of transcription on phenotype can be studied *in silico* by using SNP (eQTL) predictive models (PredictDB) to "impute" gene expression into genetic datasets and associating this genetically-predicted expression with phenotype; this can also be done with summary statistics. Here, we use a brain-specific model trained on data from the CommonMind Consortium, with genes not in the model imputed as the median of all other tissues. Using schizophrenia summary statistics from the Psychiatric Genomics Consortium (PGC), we have observed that expression variation at 16p11.2 has a greater contribution to schizophrenia compared to matched permuted sets of adjacent genes ($P < 0.01$). Six genes at or near the region showed significant positive association with schizophrenia: *INO80E* ($P = 2 \times 10^{-6}$), *TMEM219* ($P = 1 \times 10^{-6}$), *DOC2A* ($P = 6 \times 10^{-7}$), *PPP4C* ($P = 2 \times 10^{-6}$), *MAPK3* ($P = 4 \times 10^{-6}$), *MAZ* ($P = 6 \times 10^{-6}$). Because there are 16p11.2 genes that are expected to influence downstream genes, which may have additional impact on schizophrenia, we explored the effects of transcriptional variation in genes downstream of CNV genes. We generated downstream regulatory networks by identifying genes in the Genotype-Tissue Expression Consortium (GTEx) data whose expression was correlated with the genetically-predicted expression of 16p11.2 genes. 91% of 16p11.2 genes tested had at least one downstream gene, and on average, each gene had 28 unique targets; this compares with XX% of genes and YY targets in control regions. Six genes at or near 16p11.2 had networks positively associated with schizophrenia: *BOLA2*, *DOC2A*, *CD2BP2*, *PRRT2*, *SULT1A4*, *MAPK3* (permuted $P < 0.05$). Our results suggest multiple 16p11.2 genes contribute to schizophrenia, directly or by affecting genes downstream, and that this CNV region is unusually influential on additional genes in *trans*.

2445W

Polygenic risk score improves prediction of alcohol-related morbidities.

T.T.J. Kiiskinen¹, N. Mars¹, J. Koskela¹, J.T. Rämö¹, J. Kaprio¹, V. Salomaa², P. Mäkelä², A.S. Havulinna^{1,2}, S. Ripatti^{1,3,4}, GSCAN. 1) Institute for Molecular Medicine Finland (FIMM), HiLIFE, University of Helsinki, Helsinki, Finland; 2) National Institute for Health and Welfare, Helsinki, Finland; 3) Program in Medical and Population Genetics and The Stanley Center for Psychiatric Research, The Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; 4) Department of Public Health, Clinicum, Faculty of Medicine, University of Helsinki, Helsinki, Finland.

Alcohol drinking is a major dose-dependent contributor to global morbidity and mortality. Alcohol drinking behavior is strongly affected by genetic factors: estimated heritability of alcohol consumption is 43% and its genomic heritability-estimate is 13%. A growing number of large scale GWASes have identified multiple loci associated with alcohol consumption. Measures of this polygenic burden may provide more objective prognostic information independent of individual reporting bias. Here, we aim to study if a polygenic risk score (PRS) for alcohol consumption is associated with alcohol-related morbidities and mortality in the Finnish FINRISK population survey cohorts ($n = 27,344$) spanning 1992 to 2012. We used summary statistics from the largest existing GWAS meta-analysis of alcohol consumption (GSCAN consortium, $n = 527,282$ after exclusion of all Finnish samples) and applied a Bayesian PRS method (LDpred) to create a genome-wide PRS for alcohol consumption. To identify individuals with major alcohol-induced health events, we constructed 21 somatic and psychiatric alcohol-related disease endpoints using nationwide hospital discharge, causes-of-death and prescription drug purchase registries that were electronically linked to the cohort baseline data. One SD increase in the PRS of alcohol consumption was associated with a 9.5 g increase in weekly pure alcohol intake ($\beta = 9.47 [8.06-10.90 \text{ g}]$, $p = 6.3 \times 10^{-36}$). The PRS was associated with lifelong alcohol-related morbidities status ($n = 1,430$) with a 32 % increase in the risk per 1 SD increase in the PRS when alcohol consumption was not in the model (OR = 1.32 [1.25-1.40], $p = 6.0 \times 10^{-23}$), and with a 24 % increase when alcohol consumption was included in the model (OR = 1.24 [1.17-1.31], $p = 5.6 \times 10^{-13}$). Using Cox regression, the PRS was associated with a 20% increase in the risk of combined alcohol-related major health events, hospitalization or death ($n = 715$ cases) (HR = 1.20 [1.11-1.29], $p = 2.0 \times 10^{-6}$). We observed a similar increase in the risk of alcohol-related mortality alone ($n = 273$ deaths) (HR = 1.22 [1.09-1.39], $p = 1.0 \times 10^{-3}$). In conclusion, increased polygenic risk for alcohol consumption is associated with alcohol-related health events. The PRS was associated with alcohol-related morbidities even when self-reported baseline alcohol consumption was in the model. These preliminary findings show promise for utilizing genetic information to improve risk estimation for alcohol-related disorders.

2446T

Mutation screening in *SHANK2* and *SHANK3* genes in Brazilian individuals with autism spectrum disorder. D.B.A. Rosan¹, A.L. Bossolani-Martins², S. Ezquina³, L.A. Britto³, M.R. Passos-Bueno³, A.C. Fett-Conte⁴. 1) Universidade Estadual Paulista Julio de Mesquita Filho (UNESP/IBILCE), Sao Jose do Rio Preto, Sao Paulo, Brazil; 2) Universidade Federal do Mato Grosso do Sul (UFMS), Paranaiba, Mato Grosso do Sul, Brazil; 3) Universidade de Sao Paulo (USP), Sao Paulo, Sao Paulo, Brazil; 4) Faculdade de Medicina de Sao Jose do Rio Preto (FAMERP), Sao Jose do Rio Preto, Sao Paulo, Brazil.

Autism spectrum disorder (ASD) is a severe neuropsychiatric disorder characterized by impairment of social communication, with repetitive patterns of behavior, interests or activities. With a complex and heterogeneous etiology, numerous genetics, epigenetics and environmental factors can explain this changes that affect 1-68 individuals. Several genes are linked to the ASD etiology, many of them are responsible for encoding proteins involved in the formation and function of synapses, acting with a probable additive effect, resulting in the behavioral phenotype. Among those genes are the *SHANK* family, composed by *SHANK1*, *SHANK2* and *SHANK3* genes that are expressed in different regions of the brain encoding scaffolding proteins that function as molecular scaffolds in the postsynaptic density of excitatory synapses. This study aimed to investigate the presence of mutations in the exons 11, 13 and 22 of *SHANK2* and exons 2, 6 and 21 of *SHANK3* genes to evaluate a possible association with the behavioral phenotype and the heritability of some variations with the broad autism phenotype (BAP) in the parents. We sequenced 200 Brazilian individuals with idiopathic ASD and the results were compared with those of a control group composed of 566 Brazilian individuals without psychiatric disorders. In *SHANK3*, we found three novel variants, one missense (NM_033517.1:c.206A>G) not present in the parents and two synonymous (NM_033517.1:c.4176C>G and NM_033517.1:c.4152C>T), the first one was also found in the patient's father with BAP. We also found two previously described variation, one synonymous (NM_033517.1:c.3813C>T) and a missense (NM_033517.1:c.4216C>G) that was not found in the parents, but in the patient's sister with BAP, we found another variation in this exon (NM_033517.1:c.4108A>G). Those variants were not found in the control group. In *SHANK2*, we found a previously described synonymous variation (NM_012309.4:c.1716C>T) in an individual with ASD and his mother with BAP, but not in controls. Five missense variations were found in the patients and in controls, but not in the parents (NM_012309.4:g.70666765C>T; NM_012309.4:c.1289C>T; NM_012309.4:c.1706G>A; NM_012309.4:c.1759C>T and NM_012309.4:c.2453G>A). Genetic studies have demonstrated the strong association of variants in those genes with susceptibility to ASD and BAP, even synonymous variants may play a role in the etiology. Further investigations should be carried out to clarify the effect of those variants.

2447F

Whole exome sequencing and linkage analysis in an extended pedigree segregating dyslexia. A.K. Adams¹, D.T. Truong², S.D. Smith³, E.G. Willcutt^{4,5}, R.K. Olson^{4,5}, J.C. DeFries^{4,5}, B.F. Pennington⁶, J.R. Gruen^{1,2,7}. 1) Department of Genetics, Yale University, New Haven, CT; 2) Department of Pediatrics, Yale University, New Haven, CT; 3) Munroe Meyer Institute, University of Nebraska Medical Center, Omaha, NE; 4) Institute for Behavioral Genetics, University of Colorado, Boulder, CO; 5) Department of Psychology and Neuroscience, University of Colorado, Boulder, CO; 6) Department of Psychology, University of Denver, Denver, CO; 7) Investigative Medicine Program, Yale University, New Haven, CT.

With the cost of sequencing continuing to fall, and the remaining difficulty of ascertaining large collections of unrelated subjects for genome wide association studies, recent approaches have returned to sequencing and linkage analysis of large extended families to discover new genes associated with complex disorders. One such family with a long history of use in dyslexia research is family 6432. While 6432 has previously been used in linkage studies to identify DYX1 in 1983, the variant or variants that contributed to the observed linkage results remain unknown. In the present study, whole exome sequencing (WES) was performed on 27 members of family 6432, a forty-three member, four generation pedigree segregating dyslexia in a Mendelian dominant manner with no apparent evidence of sex-biased inheritance. Resulting sequence was aligned, processed, and variants were identified according to the Broad Institute's GATK best practices for sequencing using an in-house pipeline. Variants were annotated using a combination of ANNOVAR and VEP to maximize number of *in silico* tools for variant information. Because of the significant number of variants observed across all members of the family, and the peculiar structure of WES data relative to SNP marker array data, the Pedigree based analysis pipeline (PBAP) was used to select non-overlapping marker subpanels and sample inheritance vectors (in MORGAN) for later linkage analysis. File conversions were handled by either PLINK or MEGA2 as appropriate. Both twopoint and multipoint linkage analysis was performed using SOLAR at 1cM resolution. Variants under corresponding linkage peaks with LOD > 1 were checked for variant segregation and *in silico* prediction of damage to corresponding proteins. Linkage results, and corresponding variants under peaks are forthcoming. Taken together, this study represents the first attempt to sequence and reanalyze a previously collected dyslexia family to discover potentially causal variants underlying an identified dyslexia associated locus.

2448W**Potential role of rare variants in the genetics of tardive dyskinesia. A.**

Alkelai¹, L. Greenbaum^{2,3}, E.L. Heinzen¹, D. Goldstein¹, B. Lerer¹. 1) Institute for Genomic Medicine, Columbia University, New York, NY, USA; 2) The Danek Gertner institute of Human Genetics, Chaim Sheba medical center, Tel Hashomer, Ramat Gan, Israel; 3) The Joseph Sagol Neuroscience Center, Chaim Sheba medical center, Tel Hashomer, Ramat Gan, Israel; 4) Biological Psychiatry Laboratory, Department of Psychiatry, Hadassah-Hebrew University Medical Center, Jerusalem, Israel.

Tardive dyskinesia (TD) is a chronic, irreversible side effect of antipsychotic treatment. TD occurrence is influenced by both clinical and demographic variables, as well as genetic factors. Contribution of common variants to TD susceptibility has been investigated in recent years, including by the genome-wide association study approach, but results are inconsistent. In order to discover the involvement of rare variants in this phenotype, we implemented the whole exome sequencing (WES) method. We performed whole exome sequencing in 20 Ashkenazi and 17 Non-Ashkenazi Jewish schizophrenia patients with severe TD, and 18 Ashkenazi and 25 Non-Ashkenazi patients without any manifestation of TD (total AIMS score of zero) despite more than 10 years of exposure to antipsychotics. We performed a case-control, exome-wide collapsing analysis including 37 patients with severe TD and 43 patients without any manifestation of TD. We also performed an additional analysis including 37 patients with severe TD and 965 unrelated healthy Jewish controls. Comparing genetic variation across 18,668 protein-coding genes we found a significant case-enrichment of rare loss of function variant in gene that seem to be relevant to this disorder (a presynaptic protein that is involved in synaptic membrane exocytosis). Follow-up case-control study in a different TD sample is currently ongoing to confirm these initial results.

2449T**A variant affecting levels of succinylcarnitine is associated with**

HLA-DQB1*06:02-negative essential hypersomnia. T. Miyagawa^{1,2}, S.S. Khor², H. Toyoda², M. Shimada^{1,2}, K. Mishima³, M. Honda^{1,4}, K. Tokunaga².

1) Department of Psychiatry and Behavioral Sciences, Tokyo Metropolitan Institute of Medical Science, Tokyo, Japan; 2) Department of Human Genetics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; 3) Department of Sleep-Wake Disorders, National Institute of Mental Health, National Center of Neurology and Psychiatry, Tokyo, Japan; 4) Seiwai Hospital, Neuropsychiatric Research Institute, Tokyo, Japan.

Essential hypersomnia (EHS) is a lifelong sleep disorder characterized by excessive daytime sleepiness. Cataplexy, which is typical in narcolepsy, is not present in EHS. EHS is associated with human leukocyte antigen (HLA)-DQB1*06:02, similar to narcolepsy. DQB1*06:02-positive and -negative EHS are different in terms of their clinical features and follow different pathological pathways. DQB1*06:02-positive EHS and narcolepsy are associated with the same susceptibility genes. However, there are fewer studies of DQB1*06:02-negative EHS. Therefore, we performed a genome-wide association study with replication for DQB1*06:02-negative EHS in 408 Japanese patients and 2,247 Japanese healthy controls. As a result, a SNP (rs10988217), which is located 15 kb upstream of carnitine O-acetyltransferase (CRAT), showed a significant association with DQB1*06:02-negative EHS ($P = 7.5 \times 10^{-9}$, odds ratio = 2.63). Meanwhile, no significant association was observed between rs10988217 and narcolepsy or DQB1*06:02-positive EHS. The risk allele of rs10988217 was correlated with higher expression levels of CRAT in various tissues and cell types. In human blood, the risk allele was also associated with levels of succinylcarnitine, which is a short-chain acylcarnitine ($P = 1.4 \times 10^{-18}$). The leading SNP in the CRAT region was the same in associations with both DQB1*06:02-negative EHS and succinylcarnitine. Next, to assess the relationship between DQB1*06:02-negative EHS and succinylcarnitine, levels of succinylcarnitine were measured in 48 patients with DQB1*06:02-negative EHS and 20 control subjects using a LC-MS/MS method. We found that succinylcarnitine levels were lower in the patients, compared with the control subjects ($P = 0.031$). CRAT catalyzes the reversible transfer of short-chain acyl groups from a short-chain acyl-CoA thioester to carnitine, thus forming the corresponding short-chain acylcarnitine. These results suggest that DQB1*06:02-negative EHS may be associated with an underlying dysfunction in energy metabolic pathways.

2450F

Schizophrenia-associated functional study of a POU3F2 regulated-TRIM8 signaling pathway in neural progenitor cells. C. Ding¹, C. Chen¹, Q. Meng¹, R. Dai¹, Y. Jiang¹, C. Liu^{1,2}. 1) Center for Medical Genetics, Central South University, Changsha, Hunan, China; 2) Department of Psychiatry, SUNY Upstate Medical University, Syracuse, USA.

Background: Schizophrenia (SCZ) is a highly heritable psychiatric disorder, and is associated with high morbidity and recurrence rate. This is because SCZ is a complex brain illness with unknown pathogenic mechanisms. Genome-Wide Association Studies have identified hundreds of SCZ associated loci, which contain many genes and SNPs. However, their underlying connections and biological effects are not clear. Recently, we identified *POU3F2* as one of the core regulators of gene co-expression network through transcriptome data from patients with schizophrenia and healthy controls. *POU3F2* regulates *TRIM8* through SCZ associated SNPs, which is located in the *TRIM8* promoter region and disrupt *POU3F2* binding sites. However, these bioinformatic results based on transcriptome data require further validation. The function of *POU3F2* and *TRIM8* in the pathogenesis of SCZ remain unknown. Accumulating evidence suggests that SCZ is likely a neurodevelopmental disorder. Several SCZ risk genes contribute to brain development through regulating proliferation and differentiation of neural progenitor cells (NPCs). In this study, we validated the regulatory relationship between transcription factor *POU3F2* and its target *TRIM8* via experiments and studied the potential effects on brain development. **Methods:** 1) We changed *POU3F2* expression and performed RT-qPCR and western blot to detect consequent expression changes of *TRIM8*. 2) The predicted regulatory elements disrupted by SNPs were verified by dual-luciferase reporter assay. We also studied the direct interaction between *POU3F2* and corresponding regions through EMAS. 3) To assess the proliferation ability of NPCs, we performed CCK-8 and Edu incorporation assays. To evaluate differentiation changes caused by expression changes of *POU3F2* and *TRIM8*, we performed neurons and astrocytes differentiation assays. **Results:** 1) Experimental results confirmed *POU3F2* positively regulates *TRIM8* expression at mRNA and protein levels. 2) *POU3F2* regulates *TRIM8* through a regulatory region that harbors the SCZ associated SNP (rs5011218), which affects *POU3F2* binding efficiency. 3) *POU3F2* and *TRIM8* inhibited the proliferation of NPCs through regulating the cell cycle. The neurons differentiation of NPCs was promoted by *POU3F2* and *TRIM8*, but they do not impact astrocytes differentiation. **Conclusion:** *POU3F2* regulates *TRIM8* expression via SCZ associated SNP (rs5011218). This pathway contributes to the risk of SCZ through affecting brain development.

2451W

Whole genome sequencing in specific language impairment (SLI) exposes a shared genetic basis for language ability in autism spectrum disorders. J. Michaelson, T. Koomar, J.B. Tomblin. University of Iowa, Iowa City, IA.

Specific language impairment (SLI) is a developmental language disorder that affects an estimated 7% of school-aged children, and has a lifelong impact on social, educational, vocational, and mental health outcomes. Although SLI shares some similarity with autism spectrum disorder (ASD) in terms of deficits in social communication, ASD is an exclusionary criterion in the diagnosis of SLI. Nevertheless, there is some evidence that there may be elements of shared genetic risk between the conditions. SLI has an estimated heritability of > 0.5, and most genetic studies to date have focused on common variation. Genetic risk for SLI likely resembles that of other neurodevelopmental conditions: a large degree of heterogeneity, with a mix of rare alleles of large effect and common polygenic risk. Here we report the first whole genome sequencing study of SLI (N=400), providing the first glimpse of the full spectrum of genetic variation underlying language ability: from rare to common, single nucleotide variants (SNVs) to structural variants (SVs). Early results show that, as expected, overall burden of SVs is predictive of IQ, and furthermore that SV burden in ASD risk genes is specifically predictive of language ability (even after correcting for IQ). Spatiotemporal expression patterns of ASD risk genes with the strongest associations with language ability converge on regions of the brain well-supported for critical roles in language. Models trained on the SLI sample that used pathway burden as predictors of language generalized successfully in predicting language ability in ASD cohorts, while identifying molecular mechanisms that contribute positively and negatively to language ability, and suggesting that these mechanisms are relevant to language across diagnostic boundaries. We found that genes and pathways that show positive contributions to language with increasing genetic burden demonstrate signatures of human-specific accelerated substitution, while genes and pathways that show a negative effect on language with increasing burden indicated evidence of stricter conservation. Furthermore, putative language-associated genes were enriched for lateralized brain expression, consistent with language as a highly lateralized function in the brain. By overlaying these language-specific findings on larger studies of ASD, we are taking the first steps in illuminating the molecular pathways that support language ability.

2452T

A genome-wide association study of completed suicide in Utah. J.S.

Anderson¹, A.A. Shabalin¹, J. Shade¹, A.V. Bakian¹, D.E. Adkins¹, B. Callor², E.D. Christensen², E. DiBlasi¹, D. Gray¹, C. Hayward², D.J. Porteous⁴, A.C. Edwards⁵, Q. Li⁶, H. Coon¹, A.R. Docherty¹. 1) Psychiatry, University of Utah School of Medicine, Salt Lake City, UT; 2) Utah State Office of the Medical Examiner, Utah Department of Health, Salt Lake City, UT; 3) MRC Human Genetics Unit, University of Edinburgh, Edinburgh, UK; 4) Centre for Genomic and Experimental Medicine, University of Edinburgh, Edinburgh, UK; 5) Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University School of Medicine, Richmond, VA; 6) Johnson and Johnson, New York, NY.

Background: Suicide is the tenth leading cause of death in the United States, with more than 40,000 suicide deaths per year. Importantly, of the approximately 3,600 per 100,000 individuals in the US who exhibit suicidal behavior or ideation, fewer than 1% actually go on to complete suicide. Heritability of *completed* suicide is estimated at 43%, indicating that genetic factors almost certainly play an important role. Because of difficulties in data collection, genetic studies examining completed suicide are rare and underpowered. Here we present results from the largest genome-wide association study (GWAS) of completed suicide to date. **Methods:** Through collaboration with the Utah Office of the Medical Examiner, we collected blood samples from 1321 individuals who died by suicide. Samples were genotyped using the Illumina Infinium PsychArray platform and cases of European ancestry were carefully matched to a large cohort of control samples from Generation Scotland, resulting in 1070 suicide cases and 3532 controls (total $N = 4602$). A mixed-model GWAS of 7 358 674 variants was conducted using GEMMA, and comparison analyses were conducted with the Utah population of the 1000 Genomes Project. Gene set and pathway enrichment analyses were employed to characterize function in the implicated variants. **Results:** A total of 126 variants mapping to 4 genes (*TLR 1*, *TLR 6*, *TLR 10*, *FAM114A1*) met genome-wide significance ($p < 5 \times 10^{-8}$). An additional 63 variants were nominally significant at $q < 0.05$ and mapped to 11 genes, including *HLA-C* and *HLA-V*. *HLA* imputation and pathway analysis further implicated major histocompatibility complex and immune-related genes in risk of suicide completion. **Discussion:** In the context of other recent findings relating major histocompatibility complex variations to psychosis and general inflammatory markers to major depression, these results help suggest that immune-related genes are indicated not only in broad psychopathology, but in extreme, rare outcomes. We discuss potential targets for future research, as well as propose that polygenic risk scores for suicide completion, calculated from these summary statistics, may help address the current gap of genetic research on completed suicide.

2453F

Sex matters: Genome-wide association study reveals sex differences in nicotine metabolism biomarker genetics among African American but not European American smokers. M.J. Chenoweth¹, J.J. Ware², A.Z.X.

Zhu¹, C.B. Cole³, L.S. Cox⁴, N.L. Nollen⁴, J.S. Ahluwalia⁵, N.L. Benowitz⁶, R.A. Schnoll⁷, L.W. Hawk Jr.⁸, P.M. Cinciripin⁹, T.P. George¹⁰, C. Lerman¹¹, J. Knight³, R.F. Tyndale¹. 1) Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health and the Departments of Pharmacology and Toxicology and Psychiatry, University of Toronto, Toronto, Ontario, Canada; 2) MRC Integrative Epidemiology Unit (IEU) and School of Social and Community Medicine, University of Bristol, Bristol, UK; 3) Data Science Institute and Lancaster University Medical School, Lancaster, UK; 4) Department of Preventive Medicine and Public Health, University of Kansas School of Medicine, Kansas City, KS, USA; 5) Department of Behavioral and Social Sciences, Brown University of Public Health, Providence, RI, USA; 6) Departments of Medicine and Bioengineering and Therapeutic Sciences, Division of Clinical Pharmacology and Experimental Therapeutics, University of California, San Francisco, CA, USA; 7) Department of Psychiatry, Perelman School of Medicine, and Abramson Cancer Center, University of Pennsylvania, Philadelphia, PA, USA; 8) Department of Psychology, University at Buffalo, SUNY, Buffalo, NY, USA; 9) Department of Behavioral Science, University of Texas, MD Anderson Cancer Center, Houston, TX, USA; 10) Division of Schizophrenia, Centre for Addiction and Mental Health and Division of Brain and Therapeutics, Department of Psychiatry, University of Toronto, Toronto, ON, Canada; 11) Department of Psychiatry, Annenberg School for Communication, and Abramson Cancer Center, University of Pennsylvania, Philadelphia, PA, USA.

Cigarette smoking behavior and cessation rates differ between men and women. Nicotine, the major psychoactive compound in cigarettes, is metabolically inactivated by CYP2A6 to cotinine and then to 3'-hydroxycotinine. The Nicotine Metabolite Ratio (NMR; 3'-hydroxycotinine/cotinine) is a heritable (~60-80%) biomarker of CYP2A6 activity and nicotine metabolism that prospectively predicts smoking cessation. In genome-wide association studies (GWASs) of the NMR controlling for sex, ~99% of hits in African (AA) and European ancestry (EA) smokers were in chromosome 19 (contains *CYP2A6*). Here we conducted sex-stratified GWASs of the NMR in AA and EA smokers from two smoking cessation clinical trials (NCT01314001 and NCT00666978). Additive genetic models adjusted for population substructure and NMR covariates (e.g. age and BMI). Overall, marked sex differences were observed in AA but not in EA smokers. In AA males ($n=352$), only $n=37/80$ genome-wide significant (GWS) hits were found on chromosome 19; the top overall hit was rs12459249, located 9.5kb 3' of *CYP2A6* (for C vs. T: $\beta=0.61$, $P=4.5 \times 10^{-18}$). Other GWS regions with >1 hit included chromosomes 3, 6, 11, and 22; none of these regions were significant in AA females ($n=503$). The chromosome 3 signal comprised 14 hits located ~150-200kb 5' of *CBLB*, involved in proteosome-mediated protein degradation. The chromosome 6 signal represented two independent loci, with one region containing 6 hits located ~6-15kb 3' of a gene putatively involved in drug transport (i.e. *SLC17A2*). In AA females, $n=66/71$ GWS hits were found on chromosome 19; the top overall hit differed from AA males and was rs11878604, located 16kb 3' of *CYP2A6* (for C vs. T: $\beta=-0.67$, $P=3.0 \times 10^{-24}$). Other GWS regions in AA females were intergenic, including a locus with 3 hits located ~1MB 5' of *TMEM132C* on chromosome 12, and were not significant in AA males. In EA males ($n=541$) and females ($n=389$), all GWS hits ($P < 5 \times 10^{-8}$) were found in chromosome 19 ($n=73$ and $n=43$ hits in males and females, respectively). In each sex, the top hit was the intronic *CYP2A6* SNP rs56113850 (for C vs. T: $\beta=0.72$, $P=4.3 \times 10^{-37}$ in males; $\beta=0.63$, $P=2.9 \times 10^{-21}$ in females). Our findings in AA show sex differences in genetic influences on a nicotine metabolism biomarker predictive of success on smoking cessation treatments, demonstrating intersections between genetic variation, race, sex, and metabolism which may lead to improved smoking treatment strategies for men and women.

2454W

A genome-wide study for interactions of Cytomegalovirus infection with genetic variation affecting age-of-onset of bipolar disorder. *B.J. Coombes¹, B. Larrabee¹, H. Sicotte¹, S.L. McElroy², M.A. Frye³, R. Yolken⁴, J.M. Biernacka^{1,3}.*

1) Department of Health Sciences Research, Mayo Clinic, Rochester, MN; 2) Lindner Center of HOPE, University of Cincinnati, Cincinnati, OH; 3) Department of Psychiatry and Psychology, Mayo Clinic, Rochester, MN; 4) Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, MD.

Bipolar disorder (BD) is a highly heritable disease, but most of the heritability remains unexplained. Identifying genetic risk factors is likely hampered by the extensive heterogeneity in BD and inadequate investigation of gene-environment (GE) interactions. Here, we aim to study GE interactions with Cytomegalovirus (CMV) exposure that are associated with age-of-onset of BD. CMV IgG and IgM antibody levels were measured in 173 patients with early onset BD (age 19 or younger) and 664 patients with late onset BD (age 20 or older). Early onset BD patients had lower seropositive rates of CMV IgG but higher titer levels of CMV IgM than patients with late onset BD. To find GE interactions associated with age-of-onset, we performed a SNP-level and gene-level gene-environment wide interaction scan (GEWIS). At the SNP-level, we used a two-step strategy to reduce the multiple testing burden and found a significant interaction of CMV IgM titer levels with rs138548272 ($p = 1.6 \times 10^{-6}$), which is 150kb upstream of a gene that encodes the glutamate metabotropic receptor (*GRM3*). *GRM3* is involved in brain function and has been found to be perturbed in individuals with schizophrenia. At the gene-level, we performed a biologically-informed GEWIS by testing for association of interactions between antibody titer and estimated gene-expression in a given tissue. Gene expression was estimated from SNP data using PrediXcan, which incorporates tissue-specific expression quantitative trait loci information to predict gene expression for each individual. In a scan of 7866 genes with predicted expression in four candidate brain tissues, one gene, DNA polymerase epsilon (*POLE*), had significant interaction with CMV IgM ($p = 4.4 \times 10^{-6}$). The results suggest that DNA damages caused by CMV infection combined with mutations leading to low expression of *POLE*, which is involved in DNA repair, may increase the likelihood of developing BD at an earlier age. No significant SNP or gene-level interactions were found with CMV IgG. This work reveals that certain exposures may explain some of the clinical and genetic heterogeneity in BD. Further investigation into the role of specific genes in modifying the effect of CMV infection on BD and its sub-phenotypes is warranted.

2455T

Genome-wide association study in an Australian population identifies *FZD1* and ionotropic glutamate receptor pathway associated with schizophrenia. *X. Liu¹, J.R. Atkins^{2,3}, W.R. Reay^{2,3}, H.M. Cairns^{2,3}, C. Fitzsimmons^{2,3}, M.J. Green^{4,5,7}, R.J. Scott⁶, V.J. Carr^{4,7}, A.V. Boddy¹, S. Low^{1,6}, M.J. Cairns^{2,3,7}.*

1) Faculty of Pharmacy, University of Sydney, Sydney, NSW, Australia; 2) School of Biomedical Sciences and Pharmacy, University of Newcastle, Newcastle, Australia; 3) Hunter Medical Research Institute, Centre for Brain and Mental Health Research, Newcastle, Australia; 4) School of Psychiatry, University of New South Wales, Sydney, Australia; 5) Neuroscience Research Australia, Sydney, Australia; 6) Cancer Precision Medicine Center, Japanese Foundation for Cancer Research, Tokyo, Japan; 7) Schizophrenia Research Institute, Sydney, Australia.

Schizophrenia is a genetically complex psychotic disorder with a high heritable component and a substantial health burden. To investigate genetic loci associated with schizophrenia in an Australian population, we conducted a genome-wide association study (GWAS) for schizophrenia (466 schizophrenia cases and 280 controls). Gene-based analysis identified *FZD1* (Frizzled Class Receptor 1) ($P=4.79 \times 10^{-6}$) as being involved in the pathogenesis of schizophrenia and pathway enrichment analysis also implicated genes regulating the ionotropic glutamate receptor pathway ($P=4.92 \times 10^{-3}$). These findings could contribute to a better understanding of schizophrenia susceptibility, pathogenesis and shed light on targets for new therapeutics.

2456F

Treatment remission in late-life depression: Post-GWAS *in silico* biological characterization of variants associated venlafaxine treatment. V.S. Marshe^{1,2}, M. Maciukiewicz², A.K. Tiwari^{2,3}, E. Sibille^{2,4}, D. Blumberg², J.F. Karp^{5,6}, E.J. Lenze⁷, C.F. Reynolds, III⁵, J.L. Kennedy^{1,2,3}, B.H. Mulsant^{1,2,3}, D.J. Mueller^{1,2,4}. 1) Institute of Medical Science, University of Toronto, Toronto, ON, Canada; 2) Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, ON, Canada; 3) Department of Psychiatry, University of Toronto, Toronto, ON, Canada; 4) Department of Pharmacology, University of Toronto, Toronto, ON, Canada; 5) Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA, USA; 6) VA Pittsburgh Health System, Geriatric Research Education and Clinical Center, Pittsburgh, PA, USA; 7) Healthy Mind Lab, Department of Psychiatry, Washington University, St. Louis, MO, USA.

Background. In geriatric depression, antidepressant treatment response is often slow and incomplete. Therefore, biomarkers are needed to personalized treatment approaches. We conducted a GWAS for late-life depression remission after antidepressant treatment and leveraged post-hoc, gene-based/network approaches to understand the genetic architecture of antidepressant response. **Methods.** We included 307 older adults (>60 years) with MDD and treated with venlafaxine XR for 12 weeks. A standard GWAS was conducted for post-treatment remission status, followed by *in silico* biological characterization of associated genes. We conducted gene-based association using MAGMA for 17,651 genes. Subsequently, we explored tissue specificity for differentially expressed genes using RNA-seq expression data from 53 specific tissues from GTEx. For biological pathway enrichment, we explored gene sets from KEGG, GO, Reactome, and BioCarta. To further understand venlafaxine remission in the context of protein interactions, we conducted protein-protein interaction (PPI) network analysis using experimentally validated data for PPIs from the STRING Interactome. **Results.** We extracted 886 genes found to be associated with remission prior to correction for multiple testing ($p < 0.05$). At the gene level, there was a significant over-representation of genes expressed in whole blood (FDR adjusted, $p = 0.04$). Although there was no significant enrichment of differentially expressed genes, we observed down-regulation in the anterior cingulate cortex, frontal cortex and the spinal cord (unadjusted $p < 0.05$). We observed a nominal association (unadjusted $p < 0.05$) of intracellular signaling networks including (1) Biocarta ErbB2, TNFR2, and β -arrestin; (2) KEGG ErbB and neurotrophin; and (3) Reactome EGFR. Our PPI network analysis found 1 core network, consisting of 312 seeds, 711 interacting proteins, and 1730 overall interactions. The ten central 'hub' proteins included UBC, EGFR, GRB2, PIK3CA, CDK5, APP, RPA1, IKBKB, RAF1 and NCOR1 which have previously associated with AD and MDD. This subnetwork showed significant enrichment for signaling pathways including neurotrophins, ErbB, MAPK, and Wnt ($p < 5.0e-07$). **Conclusions.** Overall, our post-GWAS analyses support the involvement of known processes in synaptic plasticity, as well as, intracellular signaling pathways including MAPK and neurotrophins (e.g., BDNF) as being potentially involved in venlafaxine treatment remission in older adults.

2457W

GRM8 association with response to antipsychotic iloperidone treatment. M. Polymeropoulos, S. Smieszek, C. Polymeropoulos. Vanda Pharmaceuticals, DC, DC.

In order to understand the pharmacogenomic basis of response to iloperidone treatment administered to schizophrenia patients, we conducted a genome-wide association study. Genotypes of 205 patients were utilized to evaluate the treatment with antipsychotic iloperidone. Using linear regression, we have directly tested the association between SNPs and PANSS-T (day 28), correcting for essential covariates including principal components. We identified SNPs associated with iloperidone efficacy in the Glutamate Metabotropic Receptor 8 (GRM8), top variant: rs13222700 ($P < 1 \times 10^{-6}$). Metabotropic glutamate receptor type 7 (GRM7) and type 8 (GRM8) are involved in the neurotransmission of glutamate which is supposed to play an important role in the development of schizophrenia and major depressive disorders. Minor allele associates with lower PANSS-T, effect observed in treated and not the placebo MAF in pop (Include). Furthermore there is a significant association (pval 0.0008) of the rs13222700 with MR (ilo+88)/95, where the minor allele seems to be a marker for slow metabolizers. Furthermore, Haploreg v4.1 was used to investigate the regulatory potential. In accordance the identified locus maps to an enhancer region. It shows evidence of changed motifs for STAT and Sox. Furthermore, analysis of a library of transcription factor binding site position weight matrices predicts that the SNP alters the binding site via 6 transcription factors sites. The locus has been shown to be a significant (0.00008) eQTL for GRM8 and there are 2 other strong loci. In addition, we confirmed variants previously reported in NPAS3. To investigate further we conducted a logistic regression based on binary cutoffs at PANSS 30 and on the top ends of the distribution logistic model, splitting dataset into two groups, robust responders (≥ 30 PANSTchange) vs weak (< 20 PANSS-Tchange). GRM8 signal was present. Other top loci directed us NRXN3, GRIP1, NPAS3, SHANK2, NTRK2. To leverage the top results we conducted enrichment analysis which showed enrichment in categories such as regulation of neurotransmitter levels (GO:0001505), regulation of glutamate secretion GO:1903294, impaired synaptic plasticity (P-value: 0.00001), hyperactivity: NTRK2 (0.0002), decreased prepulse inhibition (0.0009). The study provides greater understanding of the variability in treatment response and may lead to further optimization of the benefit to risk ratio for the affected patients. .

2458T

A genome-wide association study of emotion recognition and theory of mind. *M.R. Woodbury-Smith^{1,2}, P. Szatmari^{2,3}, S.W. Scherer^{2,4}, A.D. Paterson^{2,5}.*

1) Newcastle University, Newcastle upon Tyne, United Kingdom; 2) The Centre for Applied Genomics, Hospital for Sick Children, Toronto, ON, Canada; 3) Division of Child and Youth Mental Health, University of Toronto, Toronto, ON, Canada; 4) McLaughlin Centre for Molecular Medicine, University of Toronto, Toronto, ON, Canada; 5) Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada;.

Our capacity to negotiate the complex social world is mediated by numerous brain mechanisms working together. At a neuropsychological level, there is evidence for genetic contributions for several of these mechanisms. The Avon Longitudinal Study of Parents and Children (ALSPAC) has collected data on two such neuropsychological domains, emotion recognition and theory of mind (ToM), in more than 5,000 typically developing children (M:F ~ 1:1), who have also been genotyped with imputation to the 1000 genomes project. We have undertaken a genome-wide association study (GWAS) of both of these traits. Previously, genome-wide significant signals for ToM have been published from 23andMe data (N ~80,000). Using non-imputed SNPs, genome-wide estimation of heritability was ~5-7% for both traits, with very little shared heritability. Three loci with suggestive evidence of association for emotion recognition were identified; these did not overlap the published 23andMe results for ToM. GWAS of ToM in ALSPAC identified five signals at the suggestive level of significance; these also did not overlap with the published ToM results, or with the ALSPAC emotion recognition signals. Our results suggest unique rather than shared genetic mechanisms for these two traits, and a number of loci that may harbour genes underlying social cognition have been identified. Our study therefore has important implications for the understanding of the genetic underpinning of everyday social interaction.

2459F

Causal network relationships of multiple brain structure and behavior phenotypes based on BSNIP genetic association data. *Y. Xia^{1,2}, N. Alliey-Rodriguez², M. Keshavan⁴, G. Pearson⁵, B. Clementz⁶, J. Bishop⁷, C. Tamminga⁸, P. Buckley⁹, J. Sweeney¹⁰, C. Chen², E. Gershon³, C. Liu¹, BSNIP consortium.*

1) Department of Psychiatry, SUNY Upstate Medical University, Syracuse, NY; 2) Center for Medical Genetics, School of Life Sciences, Central South University, Changsha, Hunan, China; 3) Department of Psychiatry and Behavioral Neuroscience, University of Chicago, Chicago, IL; 4) Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA; 5) Olin Neuropsychiatry Research Center, Institute of Living, Hartford Hospital, Hartford, CT; 6) Psychology Department, Psychology Building University of Georgia Athens, Georgia; 7) PHARM Experimental&Clin Pharm, College of Pharmacy, University of Minnesota, Minneapolis, MN; 8) Paul M. Bass Administrative and Clinical Center, Dallas, Texas; 9) VCU School of Medicine, Richmond, VA; 10) Department of Psychiatry, UT Southwestern Medical Center in Dallas, Texas.

Developing quantifiable intermediate neurobiological and behavioral phenotypes (Phenos) for reclassification of psychiatric disorders has attracted much attention. However, the genetic contribution to these Phenos and the causal relationship among them remain unknown. Illustrating the genetic contribution and causal relationship among the Phenos will present an opportunity for understanding the etiology of the disease. The Bipolar and Schizophrenia Network for Intermediate Phenotypes (B-SNIP) consortium collected 1,115 unrelated individuals along with genetic analysis on 463 extensive Phenos from clinical and behavioral interview data, brain structural magnetic resonance imaging, diffusion tensor imaging data. To estimate the heritability of the Phenos and the causal relationship among them, MR-base platform was used. We found 121 Phenos with the estimated heritability larger than 0.2 while 78 Phenos with at least one genome-wide significant SNP, involving 231 SNPs. Among the 231 SNPs, 42 SNPs are in hub genetic factors which are defined as SNPs achieve significance in more than two Phenos. Combining with brain QTL data from multiple resources, we found 27 out of the 42 SNPs are also eQTL SNPs. Genetic similarity results showed that 17 out of the 78 Phenos were highly correlated ($r > 0.6$, $p < e-5$, corrected for 3,003 tests). We found 61 Phenos pairs with significant causal effect ($p < e-4$, corrected for 272 tests). We built up a directional Phenos network based on the causal results. Two Phenos (white matter cortical volumes from the left frontal pole and right entorhinal C.) were the root Phenos, which only have a causal effect on the others. Four Phenos (grey matter thickness of right insula, cortical gyrification indices of superior parietal and fusiform, grey matter of left postcentral) are top Phes with no causal effect to other Phenos. Meanwhile, the cortical gyrification indices of left paracentral, precentral, postcentral, frontal pole, superior frontal, and rostral middle frontal are middle Phenos, which means they are both products and contributors of the causal effect to each other. This is an investigation for genetic contribution and causal effect of the brain and behavior related Phenos. We identified brain and behavioral quantitative traits that are genetically influenced. We further illustrated the causal relationships among Phenos, which suggests the basis for genetic dissection and developmental biology of psychiatric disorders. .

2460W

Predicting the genetic regulation of schizophrenia in African American populations. *P.N. Fiorica^{1,2}, H.E. Wheeler^{2,3}*. 1) Department of Chemistry and Biochemistry, Loyola University Chicago, Chicago, IL; 2) Department of Biology, Loyola University Chicago, Chicago, IL; 3) Department of Computer Science, Loyola University Chicago, Chicago, IL.

Due to the dearth of genomic studies in people of recent African ancestry, determining the transferability of genetic association study results across populations has been challenging. Recent African ancestry populations have more genetic variation than European and Asian populations. Adding to the complexity of analysis, African American genetic ancestry is usually admixed, including chromosome blocks of both West African and European ancestry. This genetic variation, combined with environmental variation, may lead to differences in the way people in these populations express complex traits. Specifically, neuropsychological disorders, such as schizophrenia have been determined to be heritable diseases regulated by many loci. In a cohort of 2,220 African American (1241 Cases-979 Controls) individuals acquired from the Genetic Association Information Network, we performed genotype quality control, principal component analysis, imputation, and a case-control GWAS for schizophrenia. No SNPs significantly associated with schizophrenia susceptibility in this cohort. In addition to using SNP-level association tests, we performed PrediXcan, a gene-based association method for trait mapping using transcriptome prediction models from all 44 tissues of the GTEx Project. PrediXcan measures association between genetically regulated levels of expression and the targeted phenotype. Additionally, by associating a phenotype with the predicted level of expression, we can use PrediXcan to identify the direction of effect of each gene. In doing so, we identified significantly associated levels of predicted expression for three genes: increased expression of *SMPD3* (Bonferroni adjusted $P=0.043$) and decreased expression of *ZBTB1* ($P=0.040$) and *DHH* ($P=0.022$) associated with schizophrenia susceptibility. None of these genes have been reported in the NHGRI-EBI GWAS Catalog to be associated with schizophrenia, but they were associated with schizophrenia in a meta-analysis of the Psychiatric Genomics Consortium data in gene2phenotype.org, thus highlighting the advantages of PrediXcan application to smaller cohorts in comparison to GWAS. Furthermore, due to a lower multiple tests burden, PrediXcan was able to identify significant associations, where a traditional GWAS could not. These results augment previous data identifying the potential for PrediXcan to be used as a superior method for identifying genetic susceptibility compared to other genetic association tests.

2461T

Study of alcohol metabolizing genes (ADH1C and ALDH2) polymorphism in alcohol dependence. A case control study from India. *B. Shankarappa^{1,2}, S. Byrappa², R. Chandra², M. Purushottam², S. Jain², P. Murthy²*. 1) Department of Psychiatry, St Johns Medical College Hospital, Bangalore, 560034, India; 2) Molecular Genetics Lab, Department of Psychiatry, National Institute of Mental Health and Neurosciences, Bangalore, 560029, India.

Aim: □ To study the association of ADH1C(rs698) and ALDH2 (rs671) polymorphism in alcohol dependent subjects compared to controls. **Background:** Alcohol dependence (AD) is a major public health problem affecting many people and families. It is a risk factor for many health problems and several neuro-psychiatric syndromes. The symptoms of alcohol withdrawal can vary from being relatively mild, to being severe, such as Delirium tremens (DT) and seizures (SZ). Alcohol dehydrogenase (ADH) and Aldehyde dehydrogenase (ALDH) are enzymes involved in alcohol metabolism. Alcohol is converted to acetaldehyde by ADH then to acetic acid by ALDH. Km of the ADH1C is very high; In-vitro, in presence of hydrophobic substance its activity increases, which is useful for the first pass metabolism. Polymorphisms of these enzymes are known to play a role in alcohol dependence. ADH1C*1 allele is more active (2.5 fold) than ADH1C*2. **Method:** Male participants were selected from patients seen at the Centre for Addiction Medicine, NIMHANS. Individuals who met the criteria for alcohol dependence (ICD 10) were recruited into the study after obtaining informed consent. The clinical instruments applied included SAGA-IV, SADQ, MINI, FHAM and RCCL1 for alcoholic liver disease. In the present study ALDH2 SNP was studied in AD (N=223) and controls (N=150) and ADH1C polymorphism among AD subjects (N=335) and age matched controls (N=150). Of these AD subjects there were some who had experienced DT (N=154) and others who had not (N=178). DNA isolated from the peripheral blood was used for genotyping. Taqman assay and PCR-RFLP method was used for genotyping. **Results:** Genotype and allele frequencies did not differ between AD subjects and controls. The ADH1C*2 allele frequency (*2=0.48; $P=0.03$) was significantly higher in AD with DT compared to AD without DT. The less active allele, ADH1C*2, was more prevalent in the those who experienced DT. ALDH2 polymorphism was monoallelic in both the groups. **Discussion:** ADH1C activity is known to be dynamically regulated through induction or kinetic activation. The ADH1C*2 allele has a higher Km for alcohol (6.5) with lower activity. Occurrence of ADH1C*2 could be associated with a deleterious change in acetaldehyde concentration during withdrawal. It is interesting that we have found higher frequency of ADH1C*2 allele in AD subjects with DT. Studying other SNPs of ADH1C & ALDH2 might further help to understand the phenomenon better. .

2462F

Identification of risk genes in sub-threshold genome-wide association study loci in schizophrenia. R. Chen, Q. Wang, Y. Ji, Q. Wei, H. Yang, X. Zhong, B. Li. Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN.

Genome-wide association studies (GWAS) have identified numerous genetic variants associated with human complex traits/diseases, however translating the findings into understanding is a big challenge. A reason is that genetic architecture of complex traits/disease is far to be comprehensive due to the discovery power. Most complex traits/diseases involve in a large number of genetic risk contributors and each accounts for a modest effect, which require a large sample size to achieve an adequate power in GWAS. Although consortium-level collaboration has greatly overcome the issue, the discovered variants only explain a modest proportion of heritability. Many plausible associations almost reach whole genome significance are neglected. Another reason is that most associating variants are located in non-coding regions and exert effects in regulatory ways. Identifying genes that variants manifest on is still challenging. Emerging evidence shows that genetic association can be validated by biology evidence, so integration of multiple lines of evidence should help identifying high confidence risk genes from a set of GWAS loci. In this study we applied a network-based framework by the integration of multiple dimension data to identify risk genes from sub-threshold GWAS loci (subGWAS) leveraging the information from significant GWAS loci (sigGWAS) and then use the extended gene list to gain new insights about the etiology of schizophrenia. Our results show that subGWAS loci are enriched for signals: 1) subGWAS loci from PGC enrich heritability in an independent cohort; 2) Many subGWAS loci reach the whole genome significance in the following larger sample size study. We also found that the high confidence genes (HRG) identified from subGWAS were enriched in risk signals: HRG explained more heritability than a local background and enriched in previously schizophrenia-associating gene sets, like miR137 target genes; shows tissue specificity in brain and higher expression in prenatal brain; defects of HRG showed central nervous system phenotypes in mouse. Studies have established that psychiatric disorders share genetics at variant level but not at gene level because it's hard to identified risk genes from other disorders like bipolar and major depression given their GWAS is behind schizophrenia. Using more candidate genes from subGWAS loci, we aimed to explore the sharing of these disorders at gene level to advance our knowledge of psychiatric disorders.

2463W

Copy-number variations in children with disorders of spoken and written language point to genes with prenatal cerebellar expression. B. Peter^{1,2}, T. Hogan³, M. Alt⁴, S. Green⁵, N. Cowan⁶, I. Schrauwen⁷, M. Naymik⁸, M. Sacchetta¹, C. Vose¹, K. Deshpande¹, J. Guido¹, S. Gray¹. 1) Dpt. of Speech & Hearing Science, Arizona State University, Tempe, AZ; 2) Dpt. of Communication Sciences and Disorders, Saint Louis University, Saint Louis, MO; 3) Dpt. of Communication Sciences and Disorders, MGH Institute of Health; 4) Dpt. of Speech, Language, and Hearing Sciences, University of Arizona, Tucson, AZ; 5) T. Denny Sanford School of Social and Family Dynamics, Arizona State University, Tempe, AZ; 6) Department of Psychological Sciences, University of Missouri, Columbia, MO; 7) Center for Statistical Genetics, Baylor College of Medicine, Houston, TX; 8) Neurogenomics Division, Translational Genomics Research Institute, Phoenix, AZ.

Recent studies have reported cases of partial chromosomal deletions or duplications in children with language impairment and dyslexia [1, 2]. These copy-number variations (CNVs) involved genes with relevance for neurodevelopmental traits. The purpose of the present study was to validate these findings and to investigate possible gene-brain-phenotype associations. In a sample of 21 children with carefully phenotyped language impairment and/or dyslexia and 54 typical controls, DNA was genotyped with the Illumina Human PsychArray-24 v1.1. CNVs were detected by consensus of at least two of three hidden-Markov model algorithms using the online service CNV-Web-store. Results were filtered against a list of 107 candidate genes and regions compiled from previous reports of CNVs as well as single nucleotide variants thought to influence language functions. Gene expression levels in brain tissue across the lifespan were obtained from the Genotype-Tissue Expression and the Human Brain Transcriptome databases. CNVs overlapping with four previously reported CNVs were found. Heterozygous deletions overlapping with *DNAH14*, *CNTNAP2*, and *CTNNA3* were found in five children with language impairment and/or dyslexia. A heterozygous duplication overlapping with parts of *OTUD7A* and *CHRNA7* was found in one child with dyslexia. Most severe phenotypes involved the *CTNNA3* deletion. Only one typical control child had a CNV of interest, in this case a deletion involving *DNAH14*. These findings thus validate previously published CNV results. A look at gene expression levels in the brain across the lifespan shows that four of the five genes of interest in this study (the exception is *CHRNA7*) are highly expressed in the cerebellum during early prenatal development. Most striking prenatal cerebellar expression patterns were found for *CTNNA3*. These findings are consistent with the recent literature on cerebellar involvement in linguistic functions including spoken and written language processes [3]. References 1. Gialluisi, A., et al., *Investigating the effects of copy number variants on reading and language performance*. J Neurodev Disord, 2016. **8**: p. 17. 2. Pettigrew, K.A., et al., *Copy Number Variation Screen Identifies a Rare De Novo Deletion at Chromosome 15q13.1-13.3 in a Child with Language Impairment*. PLoS One, 2015. **10**(8): p. e0134997. 3. Marien, P., et al., *Consensus paper: Language and the cerebellum: an ongoing enigma*. Cerebellum, 2014. **13**(3): p. 386-410.

2464T**A new phenotyping algorithm for identifying cases of developmental language disorder in large-scale electronic health record systems.**

C.E. Walters^{1,5}, K. Margulis^{2,5}, J.E. Below³, N.J. Cox³, S.M. Camarata², R.L. Gordon^{4,3,5}. 1) Neuroscience Program, Vanderbilt University, Nashville, TN; 2) Dept. of Hearing & Speech Sciences, Vanderbilt University Medical Center, Nashville, TN; 3) Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN; 4) Dept. of Otolaryngology, Vanderbilt University Medical Center, Nashville, TN; 5) Program for Music, Mind & Society at Vanderbilt, Nashville, TN.

Developmental language disorder (DLD), also known as specific language impairment, is a heritable communication disorder characterized by language acquisition delay and difficulties with grammar and vocabulary. Despite its high prevalence in the population (7% of children) and heritability, prior GWAS of DLD have been characterized by small sample sizes. The aim of this study is to utilize an electronic health record (EHR) phenotyping approach as a first step towards large-scale GWAS of DLD. We developed a phenotyping rubric to identify cases of DLD in EHRs in a major biobank based in an academic medical center. Our initial search retrieved a set of 983 EHRs containing the following ICDs: 315.32/F80.2 (mixed-expressive receptive language disorder), 315.31/F80.1 (expressive language disorder), and 315.39/F80.89 (other developmental language disorders) and that have genetic data on file; cases were excluded from the initial search where hearing loss, intellectual disability, or chromosomal anomalies were indicated. Manual review was then performed on the EHRs to sort them into one of four categories: Category 1 consists of cases of DLD when no other medical condition that could be related to the DLD is present (normal childhood illnesses such as asthma might be present). Category 2 includes cases of DLD co-occurring with a major medical problem that is not thought to cause the DLD (e.g. diabetes, but not neurodevelopmental disorders). Therefore, cases in Categories 1 and 2 (marked for inclusion) had no other medical condition is thought to share a biological basis with DLD is indicated. Category 3 (neurological) and Category 4 (trauma) consisted of co-morbid or already present conditions that could have caused the DLD, and were marked for exclusion. Results of manual review showed that N=481 (49%) of the records met inclusion criteria (Categories 1 and 2). Automation of the algorithm, solely using ICD codes, to sort the EHRs into the four categories is underway using the KNIME programming tool. The preliminary analysis shows 75% positive predictive value of the automated algorithm for determining inclusion vs. exclusion when compared to the manual coding. Ongoing work is focused on improving the performance of the automated algorithm. This groundwork will facilitate the future direction of the study, in which we plan to validate the algorithm in biobank data (>2000 records) from multiple sites and conduct a full GWAS on associated genetic data.

2465F**Assessment of associations between mitochondrial DNA haplogroups and attention deficit and hyperactivity disorder in Korean children.** *I.W*

Hwang¹, S.H Han¹, B.N Kwon¹, H.J Kim¹, N.R Lee¹, M.H Lim², H.J Kwon³, H.J Jin¹. 1) Biological sciences Dept, Dankook University, Cheonan, South Korea; 2) Psychology and Psychotherapy Dept, Dankook University, Cheonan, South Korea; 3) Preventive Medicine Dept, Dankook University, Cheonan, South Korea.

Attention deficit hyperactivity disorder (ADHD) is a multifactorial disorder with multiple environmental and genetic factors. Until now, many genetic variants have been reported to be associated with ADHD, but there are no reports pertaining to the genetic association between mtDNA haplogroups and ADHD. Therefore, we performed an mtDNA haplogroup analysis of a total of 472 Korean children. The 20 East Asian specific mtDNA haplogroups were determined using the SNaPshot assay. We also sequenced the displacement loop (D-loop) region, position 15,971–613. Our results showed that haplogroup B4 was significantly associated with ADHD (OR, 1.90; 95% CI, 1.055 – 3.429; $p = 0.031$). Marginally significant association with ADHD was found in haplogroup B5 (OR, 0.26; 95% CI, 0.059 – 1.139; $p = 0.041$), and a similar result was also observed in boys diagnosed with ADHD (OR, 0.17; 95% CI, 0.022 – 1.340; $p = 0.048$). Compared with the girls, those with ADHD carried an excess of the haplogroup D4b (OR, 4.83; 95% CI, 1.352 – 17.272; $p = 0.014$). Stratified analysis of subtypes also showed significant results (combined: haplogroup B4, $p = 0.007$; inattentive: haplogroup F, $p = 0.022$). In the BASC-2 analysis, haplogroup D4 was significantly associated with adaptive scales ($p < 0.05$). Our results showed a possible role of mtDNA haplogroups in the genetic etiology of ADHD and ADHD symptoms in Korean children.

2466W

Comparative prevalence of copy number variation in common childhood neurodevelopmental disorders. *J.N. Kaplanga^{1,2}, S. Vanderstell¹, H. Ahmad².* 1) Paediatrics, Western University, Owen Sound, Ontario, Canada; 2) Paediatrics, McMaster University, Hamilton, Ontario, Canada.

CNV have been observed in neurodevelopmental disorders (NDD) in past studies and many pathogenic variants are shared among certain NDD. In pediatric practice, patients with a certain constellation of behavioral, emotional, cognitive and social difficulties are investigated by microarray as a first-tier test to determine a diagnosis. In the majority of cases the clinical constellation is consistent with a pre-test differential diagnosis that includes four NDD, viz; ADHD, FASD, ASD, and FXS. While prenatal alcohol exposure is the known cause of FASD and mutations in the FMR gene are the known causes of FXS, a single cause has not been identified in ADHD and ASD. And yet a large number of patients with a diagnosis of any of the four NDD share a common phenotype. Studies have suggested that many pathogenic CNVs are shared among NDD. This suggests that certain CNVs play an important role in an etiopathogenetic mechanism resulting in a shared phenotype. The purpose of this study is to: 1) determine the prevalence of CNV in ADHD, FASD, ASD and FXS patients with a shared constellation of behavioral, emotional, cognitive and social relatedness clinical features, 2) to determine whether any specific CNVs are shared across diagnoses 3) to determine whether any shared CNVs are pathogenic, and to compare the findings with data from other studies. As described previously a multicenter and multiclinic patient sample, of 89 FASD, 98 ADHD, 79 FXS and 82 ASD patients aged 6 to 18 and 50 controls were studied. We performed genomic arrays using the Illumina BeadChip platform employing Infinium cytoSNP-850K v1.1 array type. The data was analyzed by the BlueFuse Multi 4.3 software. We compared the prevalence of CNV between the patients with NDD, and controls. We report that the prevalence of CNV was 9, 13, 11, 7.8 % in ADHD, FASD, ASD and FXS patients, respectively, compared to 5% in controls ($P < 0.05\%$). For FXS the full mutation and premutation CNV were excluded in the count. The CNV were found throughout the genome with no hot spots at any one region. We compared our findings with data from large sample studies. We identified known disease loci and loci not previously reported. Although CNVs are commonly found in normal individuals, the finding of significantly higher CNV prevalence across NDD with a shared phenotype suggests that the CNVs identified in the patient group, play some role in disease etiopathogenesis. This is in agreement with previous studies.

2467T

ABCC1 genetic variation and antiepileptic mood stabilizer pharmacoresistance in bipolar disorder. *J.M. Biernacka^{1,2}, B. Coombes¹, A.M.-C. Ho³, B. Larrabee¹, M. Nassan², B. Singh², C. Colby¹, S. McElroy⁴, M. Frye².* 1) Health Sciences Research, Mayo Clinic, Rochester, MN; 2) Department of Psychiatry & Psychology, Mayo Clinic, Rochester, MN; 3) Department of Molecular Pharmacology & Experimental Therapeutics, Mayo Clinic, Rochester, MN; 4) Lindner Center of HOPE/University of Cincinnati, Cincinnati, Ohio.

Bipolar disorder (BD) is a highly heritable psychiatric disorder characterized by recurrent episodes of mania/hypomania and depression. BD is treated with various medications including certain anti-epileptic drugs (AEDs), such as valproate and lamotrigine, which have mood stabilizing effects. However, AED mood stabilizers (AED-MSs) do not lead to symptom improvement in all BD patients. We investigated the pharmacogenomics of AED-MSs by performing a genome-wide search for genetic variants associated with response to AED-MSs in BD. We used clinical and genetic data from 199 participants in the Mayo Clinic Bipolar Disorder Biobank with retrospectively defined clinical response to AED-MSs. Response to AED-MSs was quantified using the total Alda score, a semi-quantitative metric of clinical improvement. Of the 199 patients with AED-MS Alda scores, most had an Alda score for divalproex/valproate ($N=110$) or for lamotrigine ($N=99$), with a small proportion representing other AED-MSs including carbamazepine and oxcarbazepine. Gene-level tests of association were performed using MAGMA. These gene-level analyses identified two genes significantly associated with AED-MS response: *DISP1* ($p=7.8E-07$) and *ABCC1* ($2.0E-06$). *ABCC1* is a member of the ATP-binding cassette (ABC) gene family, which encode a set of ABC transporters that play a significant role in drug disposition and response. Various studies have implicated these transporters in epilepsy pharmacoresistance, but little is known about their contribution to AED-MS response in BD. Further analyses of our data showed that the gene-level finding was driven largely by two SNPs: rs875740 ($p\text{-value} = 3.3 \times 10^{-6}$) and rs35590 ($p\text{-value} = 3.8 \times 10^{-6}$). Application of PrediXcan to test the association of clinical phenotype with tissue-specific gene expression estimated from SNP data also demonstrated association of *ABCC1* with AED-MS response and indicated that higher gene expression is associated with worse treatment outcomes (whole blood: $\beta = -2.2$, $p = 0.009$). These findings are highly consistent with the drug transporter hypothesis in AED research, which proposes that overexpression of ABC transporters contributes to non-responsiveness. While our results warrant replication and follow-up in independent samples, the observed association of *ABCC1* genetic variation with AED-MS response suggests a pharmacokinetic role of *ABCC1* in response to these medications in the treatment of BD.

2468F

Common variants derived from PGC GWAS on schizophrenia have limited power to contribute to the treatment response or resistance to antipsychotic drugs. J. Li. Department of Psychiatry, Northwestern University, Chicago, IL.

Background: Drug targeted gene-set enrichment analysis support polygenic overlap between pathogenesis of schizophrenia and mechanism of action of antipsychotic drugs (Ruderfer DM et al. 2016). However, whether this accumulated genetic burden for SCZ confers increased risk for treatment response or resistant has not been determined. Many of genetic markers identified from pharmacogenomics studies of ours (Li, J et al 2018) or others (e.g. Stevenson J, et al. 2016) have been linked to schizophrenia or other psychiatric disorders. **Method:** The purpose of this study is 1) to determine whether SCZ risk loci identified by the PGC GWAS, individually and collectively by creating polygenic risk scores (PRS), had a significant effect on predicting treatment response to antipsychotic drugs in a combined dataset of three double-blind registration trial of lurasidone; 2) to test the association between PRS and treatment-resistant schizophrenia (TRS) in an independent cohort of TRS and Non-TRS (Li J et al. 2014), the phenotype of which was determined by is operationally defined as persistence of moderate to severe positive symptoms despite two or more trials of 4–6 week duration with typical or atypical APDs other than clozapine. **Result/Conclusion:** Some SNPs showed a nominal association with Δ PANSS-TOT_{LOCNEW} ($p < 0.05$), but none survived Bonferroni correction. The alleles, from those SNPs annotated for *CACNA1C*, *TCF4*, *SNAP91*, '12q24', '17p11' which showed increased risk for SCZ, were associated with a poor response to lurasidone. Alleles from those loci at *SNX19*, '3p21', '1q21', which showed increased risk for SCZ, were associated with a better response to lurasidone. Although we observed a trend for an association between PRS and treatment response with increased overall SCZ risk having poor response to lurasidone ($\beta > 0$), particularly in Δ PANSS-POS ($p < 0.05$), this association did not consistently show a statistical significance throughout all levels of partitions of genetic loci, suggesting that the common variants for SCZ risk identified by PGC GWAS, collectively, have limited power to predict response to lurasidone. No significant association between the PRS and Treatment-Resistant Schizophrenia in an independent cohort of TRS vs NTRS. This was consistent with the result from others (Wimberley T et al. 2017), suggesting that the polygenic risk score has limited power to predict the treatment response or resistant to antipsychotic drugs at the individual patient level.

2469W

MR-PheWAS: Causal association between depression and multiple diseases outcome using depression genetic variants as instrument. A. Mu-lugeta^{1,2}, A. Zhou¹, E. Hypponen^{1,3,4}. 1) Australian Centre for Precision Health, University of South Australia Cancer Research Institute, Adelaide, Australia; 2) Department of Pharmacology, School of Medicine, College of Health Science, Addis Ababa University, Addis Ababa, Ethiopia; 3) Population, Policy and Practice, UCL Great Ormond Street Institute of Child Health, London, UK; 4) South Australian Health and Medical Research Institute, Adelaide, Australia.

Background: Depression has profound effects on an individual's life but evidence relating to the causal effects on health is limited. Previous studies on causal prediction challenging due to confounding and reverse causation. The aim of this study is to explore the causal association between depression and a range of disease outcomes using linked data on hospitalisations and mortality in the UK Biobank. **Methods:** We performed hypothesis-free phenome-wide association analyses using depression genetic risk score (GRS) on 925 disease outcomes using data from 337,536 participants in the UK Biobank. GRS – disease outcome associations passing the multiple-test corrected significance threshold ($P < 0.0019$) were followed by Mendelian randomization (MR) analyses. MR analyses were conducted using inverse-variance weighted MR (MR IVW), weighted median MR and MR-Egger. We conducted multiple sensitivity analyses, including identification of potential pleiotropic variants through running phenome-wide association analysis of each depression variants.

Results: Depression GRS was associated with 42 distinct disease outcomes in the phenome-wide association analysis, with MR IVW analyses supporting a possible causal link between depression and 36 of these disease outcomes. As expected, GRS was strongly associated with depression diagnosis and mood disorders ($P < 8.5 \times 10^{-14}$), with signals also seen for related co-morbidities including anxiety (OR 1.89, 95%CI 1.50 to 2.38, $P = 6.5 \times 10^{-9}$), and sleep disorders (OR 1.59, 95%CI 1.27 to 1.98, $P = 4.9 \times 10^{-5}$). There was evidence for a causal effects on cardiovascular diseases including IHD (OR 1.30, 95%CI 1.15 to 1.47, $P = 4.0 \times 10^{-5}$) and atherosclerosis (OR 1.28, 95%CI 1.10 to 1.48, $P = 1.4 \times 10^{-3}$), and disorders of lipid metabolism like hypercholesterolemia (OR 1.22, 95%CI 1.12 to 1.34, $P = 1.6 \times 10^{-5}$). GRS also suggested associations with several disorders of the digestive system (including diseases of oesophagus (OR 1.30, 95%CI 1.17 to 1.46, $P = 3.3 \times 10^{-6}$, and gastrointestinal haemorrhage (OR 1.26, 95%CI 1.11 to 1.43, $P = 3.5 \times 10^{-4}$) and asthma (OR 1.23, 95%CI 1.06 to 1.44, $P = 6.2 \times 10^{-3}$, as well as associating with symptoms/disorders of urinary system (OR 1.36, 95%CI 1.19 to 1.56, $P = 9.2 \times 10^{-6}$). **Conclusion:** This finding may help policymakers implement early detection and management strategies of depression for preventing other diseases occurrence.

2470T

Whole genome sequencing analysis identifies multiple candidate variants in a family with autism spectrum disorder. K. Calli¹, Y. Qiao¹, R.K.C. Yuen², R.M. Stowe³, S. Martell¹, C. Chijiwa¹, E. Rajcan-Separovic¹, S.W. Scherer², M.E.S. Lewis¹. 1) Medical Genetics, University of British Columbia (UBC), Vancouver, British Columbia, Canada; 2) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Ontario, Canada; 3) Psychiatry and Medicine (Neurology), UBC, Vancouver, British Columbia, Canada; 4) Pathology and Laboratory Medicine, UBC, Vancouver, British Columbia, Canada.

Whole genome sequencing (WGS) has been more widely used as a tool in the clinical diagnosis and it helps increase the diagnosis rate up to 20% in autism spectrum disorders (ASDs). Thus far, more than 100 genes and CNV loci have been reported to be associated with ASD susceptibility. However, none are found in >1% of cases with ASDs, suggesting a diverse genetic heterogeneity of the disorders. We have recruited >700 families with simplex and multiplex ASD and are performing chromosome microarray and whole genome sequencing studies as part of the iTARGET Autism project (<http://www.itargetautism.ca/>) and the MSSNG project (<https://www.mss.ng/>). Using WGS in a trio ASD family in combination with internal bioinformatics pipelines and a commercial software VarSeq, we identified 4 inherited rare damaging missense single nucleotide variants (SNVs) in 4 genes (DNMT3A, PHF2, NRXN2, and SNRPN), and one rare copy number variant (CNV). The CNV is a paternally inherited 14 Kb microdeletion in ZNF517, which is also confirmed by DNA microarray. The variants in DNMT3A and SNRPN were also confirmed by Sanger sequencing. Clinically, the male proband has ASD, moderate intellectual disability, developmental delay, verbal apraxia, post-natal macrocephaly, large stature, adult-onset epilepsy (age 22 years), and mild facial dysmorphism (round facies, bitemporal narrowing, narrow palpebral fissures, low-set and protuberant ears, hypotonia, high arched palate). Neither parent has intellectual disability or ASD, though the father has macrocephaly and over-growth. All of the genes involved in these rare variants and CNV are reported to be ASD-related and involved in brain/neuron development. De novo mutations in the above SNV genes and a recurrent deletion in ZNF517 gene have been identified in cases with ASDs. In our proband, mutations in DNMT3A, PHF2, and SNRPN are paternally inherited while NRXN2 is maternal. DNMT3A is a newly identified ASD candidate gene and its mutation is associated with Tatton-Brown-Rahman Syndrome. Some of the phenotypes are shared in our proband. Functional analysis is in progress including whole transcriptome analyses. Conclusion: DNMT3A is likely the most relevant gene accounting for both ASD and features concordant with Tattan-Brown-Rahman syndrome. Alternatively, our subject's ASD phenotype reflects a collection of quantitative phenotypic traits associated with each of the multiple ASD risk genes identified and its complex genetic origins.

2471F

Targeted sequencing of postsynaptic protein genes associated with schizophrenia: A role for multiple rare risk mutations in schizophrenia susceptibility. M. Cheng, T. Hu, S. Hsu, H. Tsai, S. Tsai. Yuli Branch, Taipei Veterans General Hospital, Hualien County, Taiwan.

Schizophrenia is a chronic debilitating mental disorder with a high genetic component in its etiology. It is a neurobiological disorder with aberrant synaptic connectivity and synaptogenesis. Postsynaptic scaffolding proteins in the NMDA receptor-postsynaptic signaling complexes play a crucial role in regulating the synaptic transmission and the functions of various synaptic receptors. Genes involved in the formation and functional integrity of synapses can be considered as the candidate genes for schizophrenia. This study aims to identify the rare and pathogenic mutations of postsynaptic protein genes in patients with schizophrenia. We sequenced the exon regions of 18 genes encoding the synaptic proteins that belong to DLGAP, NRXN, NLGN, SHANK, and HOMER family in 98 schizophrenia patients on the Ion PGM system. We identified 50 missense heterozygous mutations, and *in silico* analysis showed some of these mutations were damaging or pathological to the protein function. Ten missense mutations, *DLGAP3*:p.H22N; *DLGAP3*:p.G352E; *DLGAP3*:p.S455R; *NRXN2*:p.642T; *SHANK2*:p.L158F; *NLGN2*:p.R309Q; *DLGAP1*:p.F70L; *SHANK3*:p.P209A; *NLGN3*:p.V346M; *NLGN3*:p.Q691R, were absent in 1517 healthy controls from Taiwan BioBank, and ExAC data set. We conducted genotyping assay in an independent sample set of 500 cases and 500 healthy controls, and none of these mutations were detected, suggesting these ten missense mutations are schizophrenia-specific variants. Immunoblot analysis revealed the protein expressions of three mutants, *SHANK2*:p.L158F; *SHANK3*:p.P209A; *NLGN3*:p.V346M, were increased in HEK293 cells compared with wild-type, while that of four mutants, *DLGAP3*:p.H22N; *DLGAP3*:p.G352E; *DLGAP3*:p.S455R; *NLGN2*:p.R309Q, were decreased. The study suggests that postsynaptic density genes harbor rare functional mutations that affect proteins expressions in some schizophrenia patients, support rare coding variants may contribute to the genetic architecture of schizophrenia. The identified postsynaptic density gene mutations, especially, the NLGN gene family, the SHANK gene family, and the DLGAP family shed light on future functional genomic investigations of the genes and related biological pathways of schizophrenia.

2472W

BrainSeq phase II: Schizophrenia-associated expression differences between the hippocampus and the dorsolateral prefrontal cortex. L. Collado Torres^{1,2}, E.E. Burke¹, J.H. Shin¹, A. Rajpurohit¹, S.A. Semick¹, A. Peterson¹, C. Valencia^{1,3}, R. Tao¹, A. Deep-Soboslay¹, T.M. Hyde^{1,4,5}, J.E. Kleinman^{1,5}, D.R. Weinberger^{1,4,5,6,7}, A.E. Jaffe^{1,2,4,6,8,9}, *BrainSeq Consortium*. 1) Lieber Institute for Brain Development, Baltimore, MD; 2) Center for Computational Biology, Johns Hopkins University, Baltimore, MD; 3) Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; 4) Department of Psychiatry, Johns Hopkins School of Medicine, Baltimore, MD; 5) Department of Neurology, Johns Hopkins School of Medicine, Baltimore, MD; 6) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 7) Department of Neuroscience, Johns Hopkins School of Medicine, Baltimore, MD; 8) Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; 9) Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD.

Background: We previously identified widespread genetic, developmental, and schizophrenia-associated changes in polyadenylated RNAs in the dorsolateral prefrontal cortex (DLPFC), but the landscape of hippocampal (HIPPO) expression using RNA sequencing is less well-explored. **Methods:** We performed RNA-seq using RiboZero on 900 tissue samples across 551 individuals (286 with schizophrenia) in DLPFC (N=453) and HIPPO (N=447). We quantified expression of multiple feature summarizations of the Gencode v25 reference transcriptome, including genes, exons and splice junctions. Within and across brain regions, we modeled age-related changes in controls using linear splines, integrated genetic data to perform expression quantitative trait loci (eQTL) analyses, and performed differential expression analyses controlling for observed and latent confounders. **Results:** We identified widespread transcriptional regulation in the DLPFC and the hippocampus over development, with 10,839 genes differentially expressed across age and brain regions (Bonferroni < 0.01) that are nominally replicated in BrainSpan (n = 79 tissue samples, 40 DLPFC and 39 HIPPO). Of these genes, 5,982 (~55%) contain differentially expressed exons and splice junctions that replicated in BrainSpan. We found 48 genes differentially expressed (at FDR < 5%) in hippocampus and 243 in DLPFC between schizophrenia patients and non-psychiatric controls. 36 of the 332 (10.8%) differentially expressed genes in HIPPO at FDR<20% are also differentially expressed in DLPFC, suggesting regional heterogeneity of the molecular correlates of schizophrenia diagnosis. We characterized extensive genetic regulation of gene expression with significantly different effects across the two brain regions: we identified 115,787 region-dependent eQTLs (at FDR < 1%), corresponding to 1,484 genes, at the gene, exon, or splice junction level (99 across all three). These region-dependent eQTLs included clinically relevant risk variants – five schizophrenia risk loci showed significant differential regional regulation. The eQTL results are available via <http://eqtl.brainseq.org/>. **Discussion:** We show extensive regional specificity of developmental and genetic regulation, and schizophrenia-associated expression differences between the hippocampus and DLPFC. These findings suggest that some components of the transcriptional correlates of developmental and genetic risk for schizophrenia may be brain region-specific.

2473T

Genomic prediction of depression risk and resilience under stress. Y. Fang¹, L. Scott², P.X.K. Song², S. Sen¹. 1) Molecular & Behavioral Neuroscience Institute, University of Michigan, Ann Arbor, MI; 2) Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI.

Background: Stress is a key trigger for major depression but there is substantial variation in response to stress between individuals. Advancing our ability to predict depression under stress holds great potential in reducing the burden of the disorder. Large-scale depression genome-wide association studies has, for the first time, provided basis for meaningful depression genomic risk scores, which have not yet been utilized for prediction of depression under stress. **Study Design:** A prospective longitudinal study utilizing physician training as a model stressor. **Participants:** 5227 interns starting residency training at 412 institutions in 14 specialties. **Measures:** Training interns were assessed for depressive symptoms with the Patient Health Questionnaire-9 (PHQ-9) and a series of demographic, psychological and historical measures before the start of the high-stress intern year (baseline). Subsequently, subjects were assessed for depressive symptoms using PHQ-9 quarterly through intern year. Depression polygenic risk score (MDD-PRS) was derived from the PGC2/23andMe dataset and assessed as a predictor of baseline and internship PHQ-9. Further, to identify whether genetic risk was acting through known risk factors, we incorporated previously identified risk factors into the model. **Results:** We found that MDD-PRS significantly predicted PHQ-9 both at baseline and under internship stress, while MDD-PRS was better predictive of depression under stress (beta=0.083, p=5.11e-13) than of baseline depression (beta=0.052, p=4.99e-5). 86.5% of the association between MDD-PRS and depression at baseline was explained by previously identified risk factors (gender, neuroticism, depression history, childhood stress). In contrast, these factors only explained 57.0% of the association under internship stress. Further, MDD-PRS better differentiated interns at low risk of depression from the rest of interns (bottom 2.5% vs. remaining 97.5%: t=3.9, p=1.4e-4) than those at high risk from the rest of interns (top 2.5% vs. remaining 97.5%: t=1.6, p=0.12). **Comment:** Genomic polygenic risk score has meaningful value in predicting future depression under stress. The PRS prediction is only partially overlapping with known psychological, demographic and historical predictors of depression, suggesting that MDD-PRS could add unique predictive power to existing models. Further, MDD-PRS may have particular utility in identifying individuals with high resilience to depression under stress.

2474F**Metabotropic glutamate receptor association in attention deficit**

hyperactivity disorder. J.T. Glessner^{1,2}, M.E. Khan^{1,2}, X. Chang^{1,2}, Y. Liu^{1,2}, N. Williams^{1,2}, J. Li^{1,2}, H. Hain^{1,2}, F. Mentch^{1,2}, C. Kao^{1,2}, P.M.A. Sleiman^{1,2}, H. Hakonarson^{1,2}. 1) Children's Hospital of Philadelphia, Philadelphia, PA., USA; 2) Perelman School of Medicine, University of Pennsylvania, PA., USA.

A polygenic disease model for most psychiatric conditions is undeniable. However, the era of agnostic genome wide scans evaluating increasing sample sizes cannot continue indefinitely. After a decade of detecting CNVs and associating them to disease phenotypes through large case-control cohorts and families, we have elucidated the pathophysiology of neuropsychiatric conditions and circumscribed relevant gene networks. Whether the singular base, <50bp, or larger span variant, we have detected it using array, whole exome sequencing or whole genome sequencing. The complex disease paradigm and spectrum of phenotypes is explicitly modeled in the nature of disease gene association. Metabotropic glutamate receptor (mGluR) genes have been especially pivotal hubs of psychiatric gene association networks. Here we explore the phenotypic and genotypic spectrum of psychiatric disorders in regards to targeted drug therapies. Using modern high-density SNP arrays allows for cost-effective large-scale evaluation of the mGluR network in neuropsychiatric conditions. We evaluated 1,500 samples with ADHD and called their mGluR network CNVs. The mGluR network constitutes 273 genes from first and second degree protein-protein interactors of the 8 GRM genes, 4 of which achieved individual statistical significance in our previous studies (Elia et al, 2011). An 8 member abbreviated mGluR gene network was introduced to model the most prevalent CNVs in the ADHD population demonstrating the greatest treatment response in neuropsychiatric symptoms to the mGluR modulator fasoracetam. A total of 154 positive individuals were found for the 8 gene set from the 1500 samples analyzed. A script was developed to allow highly efficient turn-around time for genotyping and analysis. Follow-up visual validation was performed on CNVs meeting minimum confidence score requirements. CNV detection on high-density SNP arrays, annotation, visual review, and reporting reached >98% accuracy based on Taqman validation. We first discovered significant mGluR CNV associations in ADHD and autism. Subsequent analysis found associations to phenotypes including anxiety, mood disorder, Tourette's, anorexia, conduct disorder, oppositional defiant disorder, and schizophrenia were also associated with an increased burden of mGluR CNVs compared to control subjects.

2475W**A burden of rare damaging coding variants in obsessive compulsive**

disorder. M. Halvorsen¹, J. Samuels², B. Maher², Y. Wang², F. Goes², G. Nestadt², D. Goldstein². 1) UNC Chapel Hill, Chapel Hill, NC; 2) Johns Hopkins University, Baltimore, MD; 3) Columbia University, New York, NY.

Obsessive Compulsive Disorder (OCD) is a complex neuropsychiatric disorder characterized by persistent intrusive thoughts and repetitive behavior. Like many other neuropsychiatric disorders, heritability studies offer support for genetics explaining some portion of OCD risk. Most genetic work on OCD has focused on common variation, leaving the rare variant contribution to OCD risk comparatively unexplored. To assess the burden of rare variants in OCD, we have generated what to our knowledge is currently the largest single collection of exome sequence data for OCD 'trios' and singleton cases. We have called *de novo* mutations in nearly 600 OCD trios, and report an excess of damaging coding mutations in line with what has been observed in other neuropsychiatric disorder cohorts. Combining the cohort with an OCD patient *de novo* mutation callset recently made available by a separate group (Cappi et al. 2017) we observe that *CHD8* is hit 3 times across 776 samples (2x loss of function, 1x damaging missense, $p = 2.5 \times 10^{-5}$). This gene is the only one in the combined cohort with recurrent loss of function mutation. Furthermore, we have conducted gene-based 'collapsing' analyses of rare damaging coding variants in over 1,200 OCD cases relative to a cohort of controls without neuropsychiatric disorders. While we have not identified a single gene in collapsing analysis results that passes the exome-wide significance threshold, we note that *SLITRK5*, whose knockout in mice has previously been reported to generate an OCD phenotype, is the top-ranking single result (odds ratio = 8; $p = 1 \times 10^{-4}$). Results from both *de novo* and case/control analyses are the strongest evidence to date for the presence of rare variants of large effect in OCD genetic architecture.

2476T

Visual masking deficits in schizophrenia: A view into the genetics of the disease through an endophenotype. F.A. Hodel^{1,2}, A. Shaqiri³, M. Roinishvili^{4,5}, E. Chkonia^{5,6}, A. Brand⁷, J. Fellay^{1,2}, M. Herzog³. 1) Global Health Institute, School of Life Sciences, EPFL, Lausanne, Switzerland; 2) Swiss Institute of Bioinformatics, Lausanne, Switzerland; 3) Laboratory of Psychophysics, Brain Mind Institute, EPFL, Lausanne, Switzerland; 4) Laboratory of Vision Physiology, Ivane Beritashvili Center of Experimental Biomedicine, Tbilisi, Georgia; 5) Institute of Cognitive Neurosciences, Free University of Tbilisi, Tbilisi, Georgia; 6) Department of Psychiatry, Tbilisi State Medical University, Tbilisi, Georgia; 7) Institute of Psychology and Cognition Research, University of Bremen, Bremen, Germany.

Schizophrenia is a chronic disorder determined by a complex mix of genetic and environmental factors. To better understand the contributions of human genetic variation to schizophrenia, we performed a genome-wide association study of a validated endophenotype: a backward masking task, in which two vertical bars, slightly offset in the horizontal direction, are shortly presented to participants, who are asked to indicate the offset direction (left vs. right). The bars are followed by a mask, which makes the task difficult. The time between the bars and mask was determined in 214 schizophrenia patients, 113 first-order relatives, and 148 controls. In comparison to the response times observed in controls (42.1 ± 30.6 ms), performance was significantly reduced in healthy relatives (61.3 ± 57.3 ms), and more strongly deteriorated in schizophrenia patients (129.4 ± 113.2 ms), supporting the notion that masking is an endophenotype of schizophrenia. We genotyped all participants on the Infinium Global Screening Array with Multi-Disease drop in. To search for associations between human polymorphisms and our endophenotype, we performed a single-marker genome-wide association study using linear mixed model on log-transformed data, implemented in GCTA software. We did not identify any genome-wide significant association ($p < 5 \times 10^{-8}$), indicating that common variants with strong effects are unlikely to contribute to the large inter-individual differences observed in schizophrenia-related visual masking deficits.

2477F

Whole genome sequencing of 1,748 veteran twins to detect risk and measure rare variation associated with major depressive disorder. D. Hupallo¹, C. Forsberg², J. Goldberg^{2,3}, W. Kremen^{4,5}, M. Lyons⁶, A. Soltis¹, C. Viollet¹, R. Ursano⁷, M. Stein⁴, C. Franz⁴, V. Vaccarino⁸, Y. Sun⁸, N. Smith^{2,3}, C. Dalgard¹, M. Wilkerson¹, H. Pollard¹. 1) The American Genome Center, Collaborative Health Initiative Research Program, Uniformed Services University, Bethesda, MD; 2) Seattle Epidemiologic Research and Information Center, Office of Research and Development, U.S. Department of Veteran Affairs, Seattle, WA; 3) Department of Epidemiology, University of Washington, Seattle, WA; 4) Department of Psychiatry, University of California, San Diego, San Diego, CA; 5) VA San Diego Center of Excellence for Stress and Mental Health, San Diego, CA; 6) Department of Psychological & Brain Sciences, Boston University, Boston, MA; 7) Department of Psychiatry, Uniformed Services University, Bethesda, MD; 8) Department of Epidemiology, Emory University, Atlanta, GA.

Major Depressive Disorder (MDD) is a complex neuropsychiatric disease with known genetic associations but without known links to rare variation in the human genome. Initial studies, based on SNP-arrays of common alleles, have identified 47 regions associated with MDD among individuals of European ancestry. Here, we report the sequencing of 1,711 whole genomes at 30X mean depth, including 988 monozygotic twins, 632 dizygotic twins, and 91 singletons sequenced in an effort to identify rare genetic variation associated with MDD. Sequencing results were hierarchically clustered using 54 quality metrics to identify and remove bias, resulting in 33 samples being excluded. Population structure was characterized via principle component analysis. Specimen organization within the cohort was also used to observe other potential biases, concluding no batch effect was present as a result of collection date, sequencing date, or sample processing. Kinship among twins was quantified and its impact on statistical significance of associations measured via association trials, finding a moderate inflation of test statistics when including twins. Veterans were classified as having MDD based on psychiatric interview diagnostic criteria. Using previously published genome-wide association summary statistics from one independent study, we calculated a polygenic risk score (PRS) using 13 loci. In preliminary, unadjusted analyses, we found a modestly higher PRS among Veterans with a history of MDD than among those without a history: -0.101 versus -0.112 (p -value 0.17). An initial look at average rare variant burden across loci was found to be 9,320 rare variants per 0.5Mb window, including an average of 174 exonic rare variants per 0.5Mb window within our patient cohort. A deeper exploration of the 47 loci is underway. These preliminary results may represent a replication of the MDD associations in a new independent cohort and provide support for undertaking a deeper investigation of MDD rare variation. These initial investigations together set the stage for discovering new rare variants in this complex and chronic disease and potentially expanding the genetic risk landscape for MDD.

2478W

Rare variant associations with cigarette smoking in the Trans-Omics for Precision Medicine Whole Genome Sequencing Program (TOPMed). S.K. Jang¹, H.E. Young¹, M. Liu¹, D. Becker², B. Cade³, S. David⁴, L. Emery², M. Foreman⁵, E. Fox¹, S. Gharib², D. Glahn⁶, M. Hall⁷, J. He⁸, J. Hokanson¹⁰, S.J. Hwang¹¹, A. Justice¹³, R. Kaplan¹⁶, C. Laurie², D. Levy¹¹, L. Martin¹², K. North¹³, M. Ragland¹⁰, R. Reed¹⁴, A. Shadyab¹⁵, T. Wang¹⁶, W. White¹⁷, L. Yanek², K. Young¹³, W. Zhao¹⁸, S. Vrieze¹, *Trans-Omics for Precision Medicine Whole Genome Sequencing Program (TOPMed)*. 1) University of Minnesota, Twin Cities, Minneapolis, MN; 2) University of Washington, Seattle, WA; 3) Johns Hopkins University, Baltimore, MD; 4) Brigham & Women's Hospital, Boston, MA; 5) Stanford University, Stanford, CA; 6) Morehouse School of Medicine, Atlanta, GA; 7) University of Mississippi, University, MS; 8) Yale University, New Haven, CT; 9) Tulane University, New Orleans, LA; 10) University of Colorado at Denver, Denver, CO; 11) NIH National Heart, Lung and Blood Institute, Bethesda, MD; 12) George Washington University, Washington, DC; 13) University of North Carolina, Chapel Hill, NC; 14) University of Maryland, College Park, MD; 15) University of California, San Diego, La Jolla, CA; 16) Albert Einstein College of Medicine, Bronx, NY; 17) Tougaloo College, Tougaloo, MS; 18) University of Michigan, Ann Arbor, MI.

Smoking is a major preventable cause of morbidity and mortality in the United States and is moderately heritable. Genome wide association studies (GWAS) has allowed the discovery of hundreds of common variants associated with smoking. Although low frequency variants are expected to influence tobacco use, these rare variants are yet to be identified possibly due in part to the high statistical power required to test rare variants. Most of the GWAS for smoking have been conducted in individuals with European ancestry and similar associations have not been studied in non-European populations. As a part of the Trans-Omics for Precision Medicine Whole Genome Sequencing Program (TOPMed), we seek to identify rare variants associated with four measures of smoking behaviors in a multi-ethnic sample. Among the four phenotypes, two relate to the initiation of smoking, capturing the age at which they started smoking regularly and whether an individual has ever become a regular smoker in their life. Another relates to smoking cessation which contrasts current and former smokers. Finally, the number of cigarettes smoked per day reflects the heaviness of tobacco use among smokers. At the time of writing, the TOPMed project has called genotypes from 65,000 individuals with 30x whole genome sequences. A subset of 45,000 individuals from this data freeze have one or more smoking phenotypes. We are conducting single variant association tests for age of initiation, smoking initiation, smoking cessation, and cigarettes smoked per day with 154,176,919 variants with minor allele count ≥ 1 using whole genome sequences from 7,067 individuals. The preliminary results for age of initiation showed no significant associations, suggesting that additional samples will be required to raise the statistical power to detect significant associations. We will present updated results from single variant tests and gene based tests on larger samples, up to the full sample of 45,000 for one or more phenotypes. Stronger power to detect effects will be achieved as the remainder of our sample is analyzed and as the data available to us continues to grow with the release of new data freezes through TOPMed.

2479T

Association of HLA locus alleles with posttraumatic stress disorder. S. Katrinli¹, A. Lori², K.J. Ressler^{2,4}, A.K. Smith^{1,4}. 1) Department of Gynecology & Obstetrics, Emory University, Atlanta, GA; 2) Department of Human Genetics, Emory University School of Medicine, Atlanta, GA; 3) Department of Psychiatry, Harvard Medical School and McLean Hospital, Belmont, MA; 4) Department of Psychiatry and Behavioral Sciences, Emory University, Atlanta, GA.

Background: Immune dysfunction is commonly reported among those with posttraumatic stress disorder (PTSD). An individual's immune response is shaped, in part, by the highly polymorphic Human Leukocyte Antigen (HLA) locus, which has been implicated widely in major psychiatric disorders. The goal of this study was to test for association between common HLA genotypes and PTSD. **Methods:** Genome-wide data (HumanOmni1-Quad) was used to impute classical HLA alleles in 403 lifetime PTSD cases and 399 controls of African ancestry. HLA imputation was performed using HIBAG and multi-ethnic parameter estimates. Posterior probabilities for HLA alleles in each locus was ≥ 0.8 . Validation of the model with HAPMAP ASW and LWK data also yielded ≥ 0.92 overall accuracy. Association with lifetime PTSD was evaluated using logistic regression, controlling for the first two principal components (ancestry) and gender. A Bonferroni correction was used to adjust for multiple testing in each HLA gene. The effect of HLA alleles on gene expression was assessed by weighted correlation network analysis (WGCNA), using 312 subjects with available expression data (HT12). For associated modules, enrichment analysis was performed using anRichment (Internal Collection). **Results:** Among the 7 HLA genes tested, HLA-C*07:01 (OR=1.8, $p=1.7 \times 10^{-3}$), DQA1*01:01 (OR=1.8, $p=7.9 \times 10^{-3}$) and DQB1*05:01 (OR=1.7, $p=2.1 \times 10^{-3}$) were more common in those with PTSD. WGCNA was used to explore expression patterns in those with each of the PTSD risk alleles. HLA-C*07:01 associated with 7 expression modules enriched (FDR<.05) for multiple biological pathways including general transcriptional processes and B cell receptor signaling. Interestingly, there was also enrichment in brain-specific expression highly relevant to PTSD (nucleus accumbens and amygdala). HLA-DQA1*01:01 associated with 2 expression modules, one enriched in NOD-like receptor signaling. HLA-DQB1*05:01 correlated with 2 expression modules one of which was enriched for cytokine signaling. These PTSD alleles have also been associated with other major psychiatric disorders. **Conclusions:** These results support an effect of HLA alleles on immune system and neural mechanisms to PTSD.

2480F

Genome-wide association study of heritable psychological distress

symptoms. S. Kim^{1,8}, H.J. Jang^{1,2,8}, W. Myung^{3,9}, S. Cha¹, H. Lee^{3,4}, S.K. Cho^{1,5}, B. Kim¹, J.W. Kim², D.K. Kim⁶, E.A. Stahl⁷, H.H. Won^{1,9}. 1) Samsung Advanced Institute for Health Sciences and Technology (SAIHST), Sungkyunkwan University, Samsung Medical Center, Seoul, South Korea; 2) Department of Laboratory Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, South Korea; 3) Department of Neuropsychiatry, Seoul National University Bundang Hospital, Seongnam, South Korea; 4) Institute of Health and Environment, Seoul National University, Seoul, South Korea; 5) Molecular Genetic Epidemiology Section, National Cancer Institute (NCI), Frederick, MD 21701, USA; 6) Department of Psychiatry, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, South Korea; 7) Department of Genetics and Genomic Sciences, Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 8) Co-first authors; 9) Co-corresponding authors.

Background Psychological distress symptoms are known to affect psychiatric disorder and medical illness. In addition to environmental factors, substantial interindividual variation also influences psychological distress, suggesting that genetic variation may be an underlying factor. In this study, we aim to examine heritability and genetic determinants of individual symptoms of psychological distress. **Methods** Psychological distress symptoms of 12,680 individuals in the Korean Genome and Epidemiology Study were assessed using a validated 18-item scale of psychological well-being index short form. We genotyped these individuals and estimated the heritability of individual psychological distress symptoms using the GCTA-REML method. We performed a genome-wide association analysis of individual psychological distress symptoms that showed significant heritability. Cell-type group analysis was conducted to investigate enrichment of heritability in particular cell types using linkage disequilibrium (LD) score regression. We used the GTEx expression quantitative trait locus (eQTL) database to investigate association of detected variants and their LD proxies ($r^2 > 0.8$ in East Asians) with expression levels of genes. **Results** Three psychological distress items showed significant heritability: subjective well-being (9%), fatigue and appetite (11%), and enjoying daily life (8%). Partitioned heritability analysis showed that heritability of subjective well-being was enriched in the central nervous system (CNS) including brain ($P < 0.05$). Two genome-wide significant associations were observed: rs6735649 with subjective well-being ($P = 1.32 \times 10^{-8}$, odds ratio [OR] = 1.18, 95% confidence interval [CI]: 1.12–1.25), and rs35543418 with enjoying daily life ($P = 1.33 \times 10^{-8}$, OR = 1.59, 95% CI: 1.35–1.86). Variants in high LD with rs6735649 were associated with expression levels of *STEAP3* and *C1QL2* in brain tissues. **Conclusions** Our findings demonstrate that psychological distress symptoms are partly heritable and two genetic loci are associated with subjective well-being or enjoying daily life. Further analysis in larger cohorts and different populations is required to replicate our findings and to identify additional genetic variants.

2481W

Dopamine perturbation of gene co-expression networks reveals differential response among schizophrenia subjects for translational machinery.

M.Z. Kos¹, J. Duan^{2,3}, A.R. Sanders^{2,3}, L. Blondell¹, E.I. Drigalenko⁴, M.A. Carless¹, P.V. Gejman^{2,3}, H.H.H. Goring¹, *Molecular Genetics of Schizophrenia (MGS) Collaboration*. 1) South Texas Diabetes and Obesity Institute, Department of Human Genetics, University of Texas Rio Grande Valley School of Medicine, San Antonio, TX; 2) Center for Psychiatric Genetics, NorthShore University HealthSystem, Evanston, IL; 3) Department of Psychiatry and Behavioral Neuroscience, University of Chicago, Chicago, IL; 4) Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX.

The dopaminergic hypothesis of schizophrenia (SZ) postulates that positive symptoms of SZ, in particular psychosis, are due to disturbed dopaminergic neurotransmission via the dopamine (DA) receptor D2 (DRD2), for which most antipsychotic drugs antagonize. However, dopaminergic abnormalities in SZ tend to be pre-synaptic and likely non-receptor-mediated, with the oxidative effects of DA linked to cellular toxicity and neurodegeneration. To investigate the non-receptor-mediated effects of DA on gene expression patterns in SZ, we used a cell perturbation approach *ex vivo*, generating transcriptomic profiles in B-cell transformed lymphoblastoid cell lines from 514 SZ cases and 690 controls, both before and after DA exposure (100 μ M). Weighted Gene Co-expression Network Analysis (WGCNA) of RNA-seq data identified seven co-expression modules under baseline conditions, of which six were preserved in DA-stimulated data. One of these modules shows significantly increased association with SZ after DA perturbation (baseline: $P = 0.023$; DA-stimulated: $P = 7.8 \times 10^{-5}$; Δ AIC = -10.5). This module is highly enriched for ribosomal proteins ($FDR = 2 \times 10^{-110}$) and genes involved in co-translational targeting ($FDR = 4 \times 10^{-141}$), as well as mitochondrial oxidative phosphorylation and neurodegeneration. Genome-wide SNP association testing revealed tentative QTLs underlying the module co-expression, most notably at *FASTKD2* (top $P = 2.8 \times 10^{-6}$), a gene involved in mitochondrial translation. These results provide key insights into the long-standing dopaminergic hypothesis of SZ by substantiating the role of translational machinery in SZ pathogenesis, with a potential dopaminergic mechanism disrupting mitochondrial function, and demonstrates the utility of disease-relevant functional perturbation to the study of complex genetic etiologies.

2482T

Whole blood transcriptome analysis in bipolar disorder reveals strong lithium effect. C.E. Krebs¹, A.P.S. Ori², A. Vreeker², T. Wu², R.M. Cantor¹, M.P. Boks³, R.S. Kahn³, L.M. Olde Loohuis³, R.A. Ophoff^{1,2,4}. 1) Human Genetics, UCLA, Los Angeles, CA; 2) Center for Neurobehavioral Genetics, UCLA, Los Angeles, CA; 3) Department of Psychiatry, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands; 4) Department of Psychiatry and Biobehavioral Sciences, UCLA, CA.

Bipolar disorder (BP) is a highly heritable mood disorder with a complex genetic architecture. It is commonly treated prophylactically with the mood stabilizer lithium (Li), although treatment responses vary widely across patients. Both how bipolar disorder genetic variants confer risk and the molecular mechanisms underlying Li's therapeutic effects remain poorly understood. We performed a whole blood transcriptome analysis in a BP case-control sample ($N_{\text{subjects}} = 480$) by RNA sequencing. After correcting for technical and biological covariates including medication use with lamma, very few genes showed significant case-control differential expression (DE; $\text{FDR} < 0.05$, $N_{\text{genes}} = 6$). Further analysis, however, revealed widespread changes in gene expression linked to Li treatment at the time of blood draw ($\text{FDR} < 0.05$, $N_{\text{genes}} = 976$). These genes were enriched for immune response and calcium signaling functional annotations, among others. In a gene co-expression network constructed from all RNA-sequencing samples, 27 modules of co-expressed genes were discovered, five of which were significantly associated with Li use. To determine if any of the DE profiles could be explained by differences in blood cell-type composition, we applied CIBERSORT to deconvolute cellular heterogeneity based on an external panel of 22 blood cell type gene expression signatures. After accounting for cell-type composition, the number of Li-induced differentially expressed genes decreased ($\text{FDR} < 0.05$, $N_{\text{genes}} = 233$), but still showed enrichment for immune response and calcium signaling functional annotations. One of the five Li-associated co-expression modules was also associated with a cell type proportion estimate ($P_{\text{lithium}} = 9.40\text{E-}04$, $P_{\text{neutrophils}} = 3.19\text{E-}4$) and was enriched for immunity functional annotations. We performed gene set analysis with MAGMA to examine whether the DE genes were enriched for BP or Li-response genetic risk, but found no such evidence. Overall, these results suggest that Li acts on immune response and calcium signaling pathways. While we observe elevated levels of neutrophils in Li users, changes in cell-type composition do not fully account for these patterns of DE. Our findings add to mounting evidence of Li's immune-related and calcium regulatory mechanisms, providing novel molecular validation with the largest Li-treated transcriptome dataset to date.

2483F

Genetic associations with suicide attempt severity. D.F. Levey^{1,2}, R. Polimanti^{1,2}, Z. Cheng^{1,2}, H. Zhou^{1,2}, Y. Nunez^{1,2}, S. Jain⁵, F. He⁵, R. Ursano⁶, R. Kessler⁹, J. Smoller^{9,10,11}, M. Stein^{3,4,5}, H. Kranzler^{6,7}, J. Gelernter^{1,2}. 1) Yale University, New Haven, CT; 2) Veteran Affairs Connecticut Healthcare Center, West Haven, CT; 3) University of California San Diego, CA; 4) VA San Diego Healthcare System, San Diego, CA; 5) University of California San Diego, La Jolla, CA; 6) University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; 7) Veterans Integrated Service Network 4 Mental Illness Research, Education and Clinical Center, Crescenz Veterans Affairs Medical Center, Philadelphia, PA; 8) Uniformed Services University of the Health Sciences, Bethesda, MD; 9) Harvard Medical School, Boston, MA; 10) Massachusetts General Hospital, Boston, MA; 11) Broad Institute of MIT and Harvard, Cambridge, MA.

Objective: In 2015, ~800,000 people died by suicide worldwide. For every death by suicide there are as many as 25 suicide attempts, which can result in serious injury even when not fatal. Despite this large impact on morbidity and mortality, the genetic influences on suicide attempt are poorly understood. **Method:** We performed a genome-wide association study (GWAS) of severity of suicide attempts to investigate genetic influences. A discovery GWAS was performed in Yale-Penn sample cohorts of European Americans (EAs, $n=2,439$) and African Americans (AAs, $n=3,881$). Replication was attempted in the completely independent Army-STARRS cohorts consisting of EAs ($n=9,382$), AAs ($n=1,488$), and Latinos ($n=2,963$). When combined for a trans-population meta-analysis, the final sample size included $n=20,153$ individuals. **Results:** In the Yale-Penn discovery sample, we found one genome-wide significant (GWS) signal in EAs near the gene *LDHB* (rs1677091, $p=1.07 \times 10^{-8}$) and three GWS associations in AAs: *ARNTL2* on chromosome 12 (rs683813, $p=2.07 \times 10^{-8}$), *FAH* on chromosome 15 (rs72740082, $p=2.36 \times 10^{-8}$), and on chromosome 18 (rs11876255, $p=4.61 \times 10^{-8}$). We conducted a limited replication analysis in the Army-STARRS cohorts. rs1677091 replicated in Latinos ($p=6.52 \times 10^{-3}$). A variant in LD with *FAH* rs72740082 (rs72740088; $r^2=0.68$) was replicated in AAs (STARRS AA $p=5.23 \times 10^{-3}$; AA meta, 1.51×10^{-9}). **Conclusions:** To our knowledge, this is the first GWAS of suicide attempt severity. We identified GWS associations near genes involved in anaerobic energy production (*LDHB*), circadian clock regulation (*ARNTL2*), and the catabolism of tyrosine (*FAH*). These findings provide evidence of genetic risk factors for suicide attempt severity, providing new information regarding the molecular mechanisms involved.

2484W

Severity of nausea and vomiting during pregnancy predicted by psychiatric and psychological traits. P.A. Lind¹, L. Colodro-Conde¹, B. Couvy-Duchesne^{1,2}, J. Painter¹, M. Wright², G. Montgomery², D. Nyholt^{1,3}, S.E. Medland¹. 1) QIMR Berghofer Medical Research Institute, Brisbane, Australia; 2) University of Queensland, Brisbane, Australia; 3) Queensland University of Technology, Brisbane, Australia.

Nausea and vomiting in pregnancy (NVP) affects ~70% of pregnant women to different degrees. Around 14% of pregnant women experience severe NVP and this progresses to hyperemesis gravidarum (HG) in around 1-3% which, without treatment, can be life threatening. Women who experience severe NVP and HG have higher rates of major depressive disorder (MDD) prior to and after the index pregnancy, as well as post-traumatic stress disorder PTSD as a consequence of severe NVP or HG. Herein we examined the extent to which individual differences in the severity of NVP can be predicted by a polygenic risk score (PRS) derived from the latest Psychiatry Genomics Consortium MDD (leave out QIMR; Wray et al 2018) and PTSD analyses (Bolton et al, 2014). We also tested to what extent the genetic risk for related traits (neuroticism and cortisol levels) could explain such differences. Our preliminary analysis using data on NVP were obtained from two studies undertaken at the QIMR (Australia), which are part of the NVP Genetics Consortium. The sample comprised 1,440 unrelated women with mean age at the time of the survey of 48 (range: 26-78). The severity of NVP was measured with a 5-point scale. Polygenic risk scores were calculated using the PLINK profile score method for clumped SNPs (thresholds of $p \leq 5 \times 10^{-8}$, 1×10^{-5} , 0.001, 0.01, 0.05, 0.1, 0.5 or 1). Linear regressions on the profile scores were performed, controlling for ancestry (using four principal components), age and age squared at survey time, cohort, and if it was a twin pregnancy. Around 19% women had severe NVP symptoms. Only the genetic risk for MDD predicted the severity of NVP (all PRS calculated from $p < 0.001$ to $p < 1$). The genetic risk for MDD as assessed by PGC2, which is independent of the time at which any depressive symptoms manifest, could predict NVP features. There is no evidence that the genetic risk for the other traits studies predict NVP.

2485T

Multi-omics analysis identifies mitochondrial pathways associated with anxiety-related behavior in mice and panic disorder patients. Z. Misiewicz^{1,2,3}, S. Iurato², N. Kuleskaya^{1,3}, L. Salminen^{1,3}, L. Rodrigues², G. Maccarrone², J. Martins², D. Czamara², M. Laine^{1,3}, E. Sokolowska^{1,3}, K. Trontti^{1,3}, I. Balcells^{1,3}, C. Rewerts², S. Saarnio^{1,3}, B. Novak², D. Park², A. Erhardt², C.W. Turck², I. Hovatta^{1,3}. 1) Molecular and Integrative Biosciences Research Program, University of Helsinki, Helsinki, Finland; 2) Max Planck Institute of Psychiatry, Munich, Germany; 3) Department of Psychology and Logopedics, University of Helsinki, Finland.

Anxiety disorders are common mental disorders. They are influenced by genetic and environmental factors and an interaction between them, which is largely undefined. Our aim was to identify biological pathways associated with anxiety through integrative analyses of omics data from a mouse model of chronic psychosocial stress and exposure-induced panic attacks in panic disorder (PD) patients. To induce an anxiety-like response, we subjected two inbred mouse strains, innately non-anxious C57BL/6NCRl (B6) and anxious DBA/2NCRl (D2), to psychosocial stress (chronic social defeat stress, CSDS). Following CSDS, we carried out the social avoidance (SA) behavioral test, to divide the mice into those susceptible and resilient to the behavioral effects of stress. The strains had distinct response to stress as 87% of D2, but only 30% of B6, showed social avoidance behavior. We then analyzed the bed nucleus of the stria terminalis (BNST), a part of the extended amygdala involved in stress response, for differences in gene (RNA sequencing) and protein (liquid chromatography-tandem mass spectrometry) expression between the stress-resilient, susceptible and control mice. The analyses showed downregulation of mitochondria-related pathways in the D2 (both on mRNA and protein levels) and upregulation in the B6 strain (on mRNA level). We also observed upregulation of cytochrome c (CYCS), a key element of the mitochondrial electron transport chain, in the B6 resilient and susceptible mice versus controls, which was validated by western blot ($P = 0.015$ and $P = 0.027$, respectively). We next conducted gene set enrichment analyses of the transcriptomic mouse CSDS and PD patients' blood data, collected directly and 24 h after exposure-induced panic attack. We found upregulation of mitochondria-related gene sets in the stressed B6 mice ($FDR < 0.05$), downregulation in the stressed D2 mice, and upregulation in the PD patients after exposure-induced panic attack ($FDR < 0.05$). To see if CSDS leads to changes in mitochondrial number, we carried out electron microscopy and observed that susceptible B6 mice had a larger number of pre-synaptic mitochondrial cross-sections than B6 resilient mice ($P < 0.05$). In conclusion, we found a global evolutionarily conserved response of mitochondrial pathways in anxiety-related behaviors. Identification of the underlying mechanisms will provide much needed insight into the molecular basis of anxiety, a critical step in developing targeted therapy.

2486F

Profile of the GCOP Study: A genome cohort on psychosocial traits in a dimensional perspective. T. Nishiyama¹, C. Wang², A. Hishida³, M. Watanabe¹, H. Nakagawa¹, S. Nakamura⁴, K. Katayama⁴, H. Narimatsu⁴, S. Suzuki¹. 1) Nagoya City University Graduate School of Medicine, Nagoya, Japan; 2) Department of Public Health, Aichi Medical Univ. School of Medicine, Aichi, Japan; 3) Department of Preventive Medicine, Nagoya Univ. Graduate School of Medicine, Nagoya, Japan; 4) Cancer Prevention and Control Division, Kanagawa Cancer Center Research Institute, Yokohama, Japan.

Mental disorders are the leading cause of years lived with disability, worldwide [Whiteford HA. 2013], and lead to higher morbidity and mortality rates. Conversely, many health conditions increase the risk for mental disorders. But, mechanisms for interactions between mental health conditions and other health conditions remain poorly understood [Prince M. 2007]. Therefore, we undertook a Genome Cohort on Psychosocial traits (GCOP) Study to clarify the mechanism, taking into account genetic vulnerability since every mental condition has a genetic contribution [Burmeister M. 2008]. The GCOP Study relies on a dimensional perspective that bridges between the healthy and the pathological mental traits [Kruger RF. 2014]. According to this perspective, we sampled individuals along these spectra in order to maximize precision of measurement and statistical power, as opposed to case-control designs. Recruitment of the GCOP Study was started in July, 2015 and would be continued until March, 2020. The Study collects blood, health checkup records and questionnaires for measuring psychosocial traits including personality, distress, autistic traits, attention-deficit/hyperactivity traits, obsessive-compulsive traits, psychoticism, nicotine/alcohol dependence, gambling addiction, sleep disturbance, morningness, social support, and social capital, as well as demographic and socioeconomic data, physical activity, and diet. This baseline survey was conducted in Gifu and Aichi prefectures, central Japan. 4,369 adult participants have been sampled to date and our target sample size is 40,000. After the baseline survey, GWAS will be conducted for the sample. Ethics committees of all participating centers approved the study, and all participants gave written informed consent. The GCOP Study has a dual structure. A small portion of the sample will be followed up annually by questionnaires for measuring psychosocial traits and health checkup records, while this follow-up is not conducted for the other sample. Furthermore, all the sample will be followed up for death, emigration, and cancer registration in Japan for 15 years. This unique large population-based cohort study enables us to elucidate mechanisms for interactions between mental health conditions and other health conditions. .

2487W

Genetic variations in ghrelin (*GHRL*) and its receptor (*GHSR*) genes associated with alcohol use disorder. N. Ramoz^{1,3}, P. Gorwood^{2,3}. 1) Center of Psychiatry and Neuroscience, INSERM U894, Paris, France; 2) Sainte-Anne Hospital, CMME, Paris, France; 3) University Paris Descartes, Paris, France.

It is clear that an addiction to a substance is the consequence of the combination of genetic factors (heritability of 0.5 for alcohol dependence) and environmental factors, such as exposure to alcohol, the level of school education or family and social situation. There is not a gene of alcoholism but there are vulnerability genes that increase the risk of developing alcohol dependence. Now, for addiction, we use the concept of "use disorder", and therefore for alcoholism or alcohol-dependence, we use the term of alcohol use disorder (AUD). The physio-pathobiology of AUD is likely involving the brain reward system and neurotransmitters, as well as, neuro-endocrine hormones and peptides acting in the physiology and regulation of brain and body homeostasis. Thus, the ghrelin system is implicated in the regulation of growth hormone (GH) secretion, food intake, energy expenditure and glucose homeostasis. There are several evidence confirming that the ghrelin system is involved in excessive alcohol consumption and AUD, from animal models to human evaluations. Thus, peripheral blood ghrelin concentrations are lower in AUD patients compared to healthy controls. In contrast, high ghrelin concentrations are observed in AUD patients who are abstinent. Furthermore, ghrelin levels may predict relapse to alcohol drinking in human. Finally, intravenous ghrelin administration compared to placebo in AUD patients induces the desire/craving to drink alcohol in response to alcohol cues, and increases alcohol self-administration. In our study, we search for identify mutations and polymorphisms (Single Nucleotide Variants SNVs or polymorphisms SNPs) specifically associated with alcohol dependence in candidate genes by sequencing. Our methodology was to perform high-throughput sequencing, in both directions with x70 reading depth, of the exons characterized by the ghrelin (*GHRL*) and its receptor genes (*GHSR*) and the ghrelin precursor, in 700 subjects of the COGA cohort, constituting 146 families, including 406 alcohol-dependent patients. Our sequencing results allowed us to identify several SNPs and SNVs. The specificity and the pathogenic impact in the alcohol dependence of each variant must be confirm and validate. Thus, the screening of de novo mutations in these families must be studied in other AUD cohorts and in the general population.

2488T

Genetic associations between psychiatric disorder risk and phenotypic factors in UK Biobank. R. Shafee^{1,2}, C.E. Carey^{2,3,4}, R.K. Walters^{2,3,4}, D. Palmer^{2,3,4}, L.E. Abbott^{2,3,4}, D.P. Howrigan^{2,3,4}, C. Churchhouse^{2,3,4}, B.M. Neale^{2,3,4}, E.B. Robinson^{2,3,4,5}. 1) Department of Genetics, Harvard Medical School, Boston, MA; 2) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA; 3) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, USA; 4) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA; 5) Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA.

Common polygenic variation strongly contributes to risk for psychiatric disorders including attention deficit/hyperactivity disorder (ADHD), autism spectrum disorder (ASD), and schizophrenia (SCZ). Every individual possesses some degree of polygenic risk for each of these conditions. Biobanks with large-scale phenotypic and genomic data, such as the UK Biobank (UKBB), can thus be leveraged to better understand the full spectrum of outcomes associated with that risk. Such analyses can clarify how genetic risk for psychiatric disorders varies in phenotypic presentation and highlight clinically useful associations. Beginning with more than 2500 traits measured in UKBB, we identified 350 traits with significant common-variant heritability in the white British UKBB subsample (N=333,554) using LD Score Regression (LDSR). To improve interpretability we next conducted factor analysis on these 350 traits and identified a 40-factor solution yielding cohesive and interpretable factors. The factors included a range of physical (e.g. pain, asthma), behavioral (e.g. neuroticism, sleep), and cognitive outcomes (e.g. intelligence/education). We performed a GWAS of each factor controlling for age, sex, and ancestry-informative principal components. Using disorder-specific GWAS results from recent consortium-level meta-analyses, we estimated the genetic correlation between each factor and ADHD, ASD, and SCZ via LDSR. All three psychiatric disorders were moderately genetically associated with poorer overall mental (e.g., life dissatisfaction factor: $rg=.30$ to $.39$) and physical (e.g., general health factor: $rg=-.18$ to $-.53$) health, as well as with certain behavioral factors (e.g., sleep issues: $rg=.22$ to $.56$). In contrast, divergent profiles of genetic correlation across disorders were observed for factors such as alcohol use (ADHD only $rg=.57$), weight (ADHD $rg=.37$, vs. SCZ $rg=-.12$), and wearing glasses (ADHD $rg=-.27$, vs. ASD $rg=.28$; all $p<1e-9$). It has been clearly established that the common polygenic influences on risk for neuropsychiatric disorders partially overlap. While polygenic risk for each disorder is defined by association to its clinical features (e.g. behavioral problems), we highlight how this polygenic risk can also differ substantially in its biomedical and cognitive correlates. These results suggest that the shared and unshared polygenic influences on risk for psychiatric disorders can be used to distinguish elements of the disorders' underlying biology.

2489F

Contribution of common and rare variants to bipolar disorder susceptibility in extended pedigrees from population isolates. J. Sul¹, S. Service¹, A. Huang^{1,2}, V. Ramensky^{1,3}, S. Hwang¹, T. Teshiba¹, Y. Park¹, A. Ori¹, Z. Zhang⁶, N. Mullins^{7,8}, L. Lohuis⁹, S. Fears¹, C. Araya⁹, X. Araya⁹, M. Spesny¹⁰, J. Escobar¹¹, J. Ospina-Duque¹², B. Kremer¹³, G. Bedoya¹⁴, A. Ruiz-Linares^{15,16}, R. Cantor¹, J. Molina¹⁷, G. Coppola¹, R. Ophoff^{18,19}, G. Macaya¹, C. Lopez-Jaramillo^{12,20}, V. Reus²¹, C. Bearden¹, C. Sabatti²², N. Freimer¹. 1) Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles, Los Angeles, CA 90095, USA; 2) Bioinformatics Interdepartmental Program, University of California, Los Angeles, Los Angeles, CA 90095, USA; 3) Moscow Institute of Physics and Technology, Dolgoprudny, Institutsky 9, Moscow Region 141701, Russian Federation; 4) Department of Computer Science, University of California, Los Angeles, Los Angeles, CA 90095, USA; 5) Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University California Los Angeles, Los Angeles, CA, USA; 6) Department of Genetics and Genomic Sciences, Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY 10029; 7) Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, De Crespigny Park, Denmark Hill, London, United Kingdom SE5 8AF; 8) Pamela Sklar Division of Psychiatric Genomics, Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; 9) Cell and Molecular Biology Research Center, Universidad de Costa Rica, San Pedro de Montes de Oca, San José, Costa Rica 11501; 10) Division of Pediatric Pulmonology, Hospital Nacional de Niños, San Jose, Costa Rica; 11) Department of Psychiatry and Family Medicine, Rutgers-Robert Wood Johnson Medical School, New Brunswick, NJ 08901; 12) Grupo de Investigación en Psiquiatría (Research Group in Psychiatry; GIPSI), Departamento de Psiquiatría Facultad de Medicina, Universidad de Antioquia, Medellín, Colombia 050011; 13) Department of Genetics, Evolution and Environment, University College London, London WC1E 6BT, United Kingdom; 14) Laboratory of Molecular Genetics, Institute of Biology, University of Antioquia, Medellín 050010, Colombia; 15) Ministry of Education Key Laboratory of Contemporary Anthropology and Collaborative Innovation Center of Genetics and Development, Fudan University, Shanghai 200438, China; 16) Aix Marseille Univ, CNRS, EFS, ADES, Marseille, France; 17) BioCiencias Lab, 01010 Guatemala, Guatemala; 18) Department of Human Genetics, University of California Los Angeles, Los Angeles, CA 90095, USA; 19) Department of Psychiatry, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands; 20) Mood Disorders Program, Hospital San Vicente Fundación, Medellín, Colombia 050011; 21) Department of Psychiatry and UCSF Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, CA 94143; 22) Department of Health Research and Policy, Division of Biostatistics, Stanford University, Stanford, CA 94305.

Current evidence from case/control studies indicates that genetic risk for psychiatric disorders derives primarily from numerous common variants, each with a small phenotypic impact. The literature describing apparent segregation of bipolar disorder (BP) in numerous multigenerational pedigrees suggests that, in such families, large-effect inherited variants might play a greater role. To evaluate this hypothesis, we conducted genetic analyses in 26 Colombian and Costa Rican pedigrees ascertained for BP1, the most severe and heritable form of BP. In these pedigrees, we performed microarray SNP genotyping of 856 individuals and high-coverage whole-genome sequencing (WGS) of 454 individuals, the largest BP pedigree sample sequenced to date to the best of our knowledge. In addition, we performed family-aware genotype imputation to nearly double the sample size of individuals with WGS data to 785 individuals. We analyzed these data to obtain high-quality genotypes for several categories of genetic variants, including single nucleotide variants (SNVs), copy number variants (CNVs), short tandem repeats (STRs), and *de novo* mutations. With this rich genetic information, we have investigated the roles of common and rare variants in risk of severe BP (BP1). We identified that the BP1 individuals, compared to their unaffected relatives, had higher polygenic risk scores estimated from SNPs associated with BP discovered in independent genome-wide association studies, and displayed a higher burden of rare deleterious SNVs and rare CNVs in a set of genes hypothesized previously to be relevant to BP1. Parametric and non-parametric linkage analyses using SNPs in microarray data and STRs detected from WGS identified 15 BP1 linkage peaks, encompassing about 100 genes. There was one locus (16q23) with the linkage evidence observed with both parametric and non-parametric analysis, but none of common coding SNVs in the locus had significant association p-values after multiple testing correction. We also developed a new statistical approach to test the segregation of rare variants, but we observed no significant segregation pattern for any particular rare coding SNVs and CNVs suggesting that the variants contributing to our linkage findings may lie outside of coding regions. These results suggest that even in extended pedigrees, genetic risk for BP appears to derive mainly from small to moderate effect rare and common variants.

2490W

Genetic markers of human evolution are enriched in schizophrenia. B.

Xiang, KZ. Liu, ML. Yu, XM. Liang, CH. Huang, W. Lei, J. Chen, J. Zhang, WY. He, T. Zhang. Affiliated of Southwest Medical University, Luzhou, Sichuan, China.

Aim: The human brain is the outcome of innumerable evolutionary processes. The systems genetics of schizophrenia could bear their signatures, which may be associated with the etiology of schizophrenia. **Methods:** On this basis, we analyzed schizophrenia (SCZ) using the GWAS summary statistics from the Psychiatric Genomics Consortium. Machine learning derived scores were used to investigate two natural-selection scenarios: complete selection (loci where a selected allele reached fixation) and incomplete selection (loci where a selected allele has not yet reached fixation). We chose these SNPs with the top-5% for incomplete and complete selection score, and based on the p value of these SNPs to obtain five subgroups: $p < 0.0001$, 0.001, 0.01, 0.05 or 0.1. **Results:** In the complete selection, we obtained 48 genes ($p < 0.0001$), 118 genes ($p < 0.001$), 289 genes ($p < 0.01$), 569 genes ($p < 0.05$), 775 genes ($p < 0.1$). In the incomplete selection, we detected 29 genes ($p < 0.0001$), 69 genes ($p < 0.001$), 213 genes ($p < 0.01$), 512 genes ($p < 0.05$), 776 genes ($p < 0.1$). The brain transcriptome data was obtained from BrainSpan, we found that 48 and 29 genes were enriched for the transcriptional co-expression profile in the prenatal dorsolateral prefrontal cortex (DFC), inferior parietal cortex (IPC), and ventrolateral prefrontal cortex (VFC), and three core genes (*GNA13*, *TBC1D19*, and *RNF111*) were identified to be involved in regulating early brain development. Functional enrichment also revealed that calcium ion binding, internalization of ErbB1, and IGF1 pathway. **Conclusions:** This study showed a marker of human evolution, which is in line with the hypothesis that the persistence of schizophrenia is related to the evolutionary process of becoming human, and demonstrated that the DFC, VFC and IPC may play a central role in neurodevelopment and SCZ risk.

2491T

Increased burden of rare protein-disrupting variants within neuron synaptic genes among 483 patients with bipolar 1 disorder. J. Xiaoming^{1,2},

M. Kvale², L. Shen², K.K. Thai², P.Y. Kwok², C. Schaefer², N. Risch^{2,3,4}. 1) Department of Neurology, University of California San Francisco, San Francisco, CA; 2) Institute for Human Genetics, University of California San Francisco, San Francisco, CA; 3) Kaiser Permanente Northern California Division of Research, Oakland, CA; 4) Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA.

Bipolar disorder (BD) is a severe and common mental disorder characterized by episodes of mania and depression, and affects approximately 5.7 million Americans. Genome-wide association studies have identified up to 30 loci that contribute to BD susceptibility, implicating genes that encode ion channels, neurotransmitter transporters, and synaptic components. However, whether rare variation contributes to BD pathogenesis remains unknown. We performed whole-exome sequencing (WES) in a multi-ethnic cohort of 483 unrelated individuals with bipolar 1 disorder (B1D) and 483 ancestry-matched controls. Populations examined include European (40%), African American (20%), Latino (20%), and East Asian (20%). Analysis of 966 exomes revealed 149,632 rare (gnomAD maximum population-specific minor allele frequency $< 1\%$) single nucleotide variants (SNVs) that disrupt a protein coding sequence or splice-site. We assigned variant pathogenicity according to ACMG criteria using the software InterVar, and identified 508 pathogenic (predominantly splice site and stop gain) and 819 likely pathogenic (predominantly missense) variants. We observed no exome-wide enrichment of rare synonymous or protein-altering SNVs in individuals with B1D compared to healthy controls. We then examined 3,230 genes categorized as loss-of-function (LOF) intolerant, which include genes disrupted in neurodevelopmental and psychiatric disorders, and found an increased burden of rare pathogenic SNVs within these genes in B1D patients (0.29 per individual) compared to controls (0.22) ($OR=1.5$, $P=0.03$). Given the episodic nature of BD, we then hypothesized that disease pathogenesis might be driven in part by dysregulated neuronal synaptic function. We examined 2,788 synaptic genes (whose RNAs undergo alternative splicing by RBFOX proteins), and observed a significant enrichment of rare pathogenic SNVs in B1D patients (0.24 per individual) compared to controls (0.16) ($OR=1.54$, $P=0.005$). The risk for B1D increased with the number of such variants, which ranged from 0 to 3 per individual. Importantly, this enrichment was exclusive to 1,222 synaptic genes that are considered LOF-intolerant ($OR=1.8$, $P=0.0002$ for all pathogenic and likely pathogenic SNVs), and was not present in 1,587 synaptic genes that are LOF-tolerant ($P=0.89$). Our results suggest that rare protein-disruptive variants, especially in genes that influence neuronal synaptic function, might contribute to susceptibility in bipolar 1 disorder.

2492F

Meta-analysis of GWAS summary data identifies novel pathways associated with alcohol use disorder. *H. Zhang, S.P. Chou.* National Institute on Alcohol Abuse and Alcoholism, Rockville, MD.

Alcohol use disorder is one of the most prevalent psychiatric disorders globally. Genetic component attributes to about half of the risk for alcoholism. Most genome-wide association studies (GWAS) mainly focus on single genetic variant or individual gene analysis, which may not have enough power to detect the multiple small independent association signals underlying the susceptibility loci. We conducted the meta-analysis of genetic pathways on alcohol dependence by using publicly available GWAS summary data from the Study of Addiction: Genetics and Environment (SAGE; dbGaP study Accession: phs000092; Illumina 1M beadchip array), which included 1,235 alcohol dependent cases and 1,433 unrelated, alcohol-exposed, non-alcohol dependent controls in the European ancestry (EA) population and 662 alcohol dependent cases and 499 controls in the African American (AA) population. The GCTA-fastBAT (a fast and flexible set-Based Association Test) software was used to perform association analysis of genetic pathways in either ethnic population separately, followed by meta-analysis of pathway association analysis in both these two ethnic groups using Fisher's method. A total of 33 genetic pathways yielded $P < 10^{-2}$ in the EA population, while 9 genetic pathways passed the same significant threshold ($P < 10^{-2}$) in the AA population. Notably, three biological pathways show genetic association ($P < 0.05$) in both discovery EA population and replication AA population: Reactome SHC mediated signaling pathway (discovery $P = 7.4 \times 10^{-3}$, replication $P = 0.04$, meta-analysis $P = 2.7 \times 10^{-3}$), KEGG PPAR Signaling pathway (discovery $P = 1.1 \times 10^{-2}$, replication $P = 0.02$, meta-analysis $P = 1.7 \times 10^{-3}$) and Biocarta CDC42RAC pathway (discovery $P = 4.3 \times 10^{-2}$, replication $P = 0.02$, meta-analysis $P = 8.1 \times 10^{-3}$). Our meta-analysis of genetic pathway on GWAS summary data uncovered several novel pathways involving in susceptibility to alcohol dependence. Our results indicated that pathway-level GWAS models is useful to identify and further validate small independent genetic signals for complex traits and diseases in human.

2493W

Depression state-related biomarkers detected by whole-blood RNA sequencing. *H. Lee¹, Y. Lee¹, C. Wu¹, I. Chen¹, Y. Lin¹, S. Liao¹, H. Chen², P. Kuo^{1,2}.* 1) Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan; 2) Research Center for Genes, Environment and Human Health, National Taiwan University, Taipei, Taiwan; 3) Department of Psychiatry & Center of Sleep Disorders, National Taiwan University Hospital, Taipei, Taiwan; 4) Department of Psychiatry, National Taiwan University Hospital, Taipei, Taiwan.

Recurrent depressive episode is characterized in patients with major depressive disorder (MDD) or bipolar disorder (BPD). To better understand the underlying biological mechanisms of depressive episode, we aimed to investigate transcriptome changes between acute episode versus remission status for depressive episode among patients with MDD or BPD using RNA sequencing. At the discovery phase, we recruited five clinically diagnosed patients with BPD or MDD, who were repeatedly measured for their severity of depressive symptoms using Hamilton Depression Rating Scale (HAMD) during episode (HAMD score ≥ 16), and follow-up for at least two-months till remission (HAMD score ≤ 7). Blood samples were drawn at the time of depression severity assessment. We applied RNA sequencing with appropriate quality control, normalization, and differential expression analyses to examine differentially expressed genes (DEGs) between the two statuses. Several DEGs were identified, including Small Integral Membrane Protein 10 Like 1 (*SMIM10L1*; log2FC = 1.38, $P = 3.80 \times 10^{-5}$), Maestro Heat Like Repeat Family Member 6 (*MROH6*; log2FC = 1.18, $P = 4.89 \times 10^{-4}$), and scribbled planar cell polarity protein (*SCRIB*; log2FC = 1.11, $P = 6.18 \times 10^{-4}$), though these genes were insignificant after FDR correction. Preliminary results were validated using quantitative real-time PCR and the replication study using an independent sample of additional five pairs of the patients are underway. Pathway analysis was performed to identify enriched pathways for DEGs, and we found that immune-related pathways stood out for their associations with depressive episode. Further studies with a bigger sample size are needed to examine the targets identified in our study in order to identifying reliable biomarkers for depression state.

2494T

Evidence for genetic contribution to the increased risk of type 2 diabetes in schizophrenia.

S. Hackinger¹, B. Prins², V. Mamakou^{3,4}, E. Zengini^{4,5}, E. Maroulis⁶, L. Brčić⁷, I. Serafetinides⁸, K. Lamnissou⁹, V. Kontaxakis¹⁰, G. Dedoussis¹¹, F. Gonidakis¹², A. Thanopoulou¹³, N. Tentolouris¹⁴, A. Tsezou¹⁵, E. Zeggini¹. 1) Human Genetics, Wellcome Trust Sanger Institute, Cambridge, Cambridgeshire, United Kingdom; 2) Strangeways Research Laboratory, University of Cambridge, 2 Worts' Causeway, Cambridge CB1 8RN, United Kingdom; 3) Medical School, National and Kapodistrian University of Athens, Athens 11528, Greece; 4) 5th Department, Dromokaiteio Psychiatric Hospital, Athens 124 61, Greece; 5) Department of Oncology and Metabolism, University of Sheffield, Beech Hill Road, Sheffield S10 2RX, United Kingdom; 6) William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, EC1M 6BQ UK; 7) Department of Medical Biology, University of Split, School of Medicine, Split 21000, Croatia; 8) Department of Gastroenterology, Gennimatas General Hospital, Athens 11527, Greece; 9) Department of Biology, National and Kapodistrian University of Athens, Panepistimioupolis, Ano Ilisia, Athens 15771, Greece; 10) Early Psychosis Unit, 1st Department of Psychiatry, Eginition Hospital, Medical School, National and Kapodistrian University of Athens, Athens 11527, Greece; 11) Department of Nutrition-Dietetics, Harokopio University, 17671 Athens, Greece; 12) 1st Department of Psychiatry, Eginition Hospital, Medical School, National and Kapodistrian University of Athens, Athens 11528, Greece; 13) Diabetes Centre, 2nd Department of Internal Medicine, Hippokratia General Hospital, Medical School, National and Kapodistrian University of Athens, Athens 11527, Greece; 14) 1st Department of Propaedeutic Internal Medicine, National and Kapodistrian University of Athens, Medical School, Laiko General Hospital, Athens 11527, Greece; 15) Faculty of Medicine, Department of Biology, University of Thessaly, 41500 Larissa, Greece This work was carried out at the Department for Human Genetics, Wellcome Trust Sanger Institute, Cambridge, United Kingdom.

Schizophrenia (SCZ) patients are 1.5-2 times more likely to develop type 2 diabetes (T2D) compared to the general population. Several explanations for this epidemiologic link have been proposed, including environmental factors, the use of antipsychotic medication, and/or shared genetic aetiology; however, the aetiology underlying this observed comorbidity remains poorly understood. Here, we investigate the presence and extent of a shared genetic background between SCZ and T2D using genome-wide approaches. We used summary statistics from two large-scale GWAS of SCZ and T2D from the PGC ($n_{\text{cases}}=34,241$; $n_{\text{controls}}=45,604$) and DIAGRAM ($n_{\text{cases}}=12,171$; $n_{\text{controls}}=56,862$) consortia, respectively, to perform genetic overlap analyses, including pathway analysis and a regional colocalisation test. We identify five genomic regions with evidence of colocalising SCZ and T2D signals, three of which contain known loci for both diseases. We also observe a significant excess of shared association signals between SCZ and T2D at nine out of ten investigated p-value thresholds. Finally, we identify 29 genes associated with both T2D and SCZ, several of which have been implicated in biological processes relevant to these disorders. Additionally, we performed a genome-wide association study (GWAS) and polygenic risk score analysis in a Greek sample collection (GOMAP) comprising three patient groups: SCZ only ($n=924$), T2D only ($n=822$), comorbid SCZ and T2D ($n=505$); samples from two separate Greek cohorts were used as population-based controls ($n=1,125$). We used the DIAGRAM and PGC data to construct risk scores in GOMAP. Risk scores for both disorders were significantly associated with comorbidity of SCZ and T2D compared to controls. Together our results show for the first time that patients with comorbid SCZ and T2D have a higher genetic predisposition to both disorders compared to controls, demonstrating that the observed comorbidity between SCZ and T2D is at least in part due to shared genetic mechanisms.

2495F

Regulatory Variants in *PDE4D* associate with psychotic and cognitive domains of major mental illness.

V. Sinha^{1,2,3}, L.-U. Vuotij^{1,2,3}, A. Ortega-Alonso^{1,2}, M. Tomiainen-Holm^{1,2}, S. Therman¹, A. Tuulio-Henriksson^{3,4}, P. Jylha⁵, J. Kaprio^{2,6}, I. Hovatta^{3,7}, E. Isometsa⁸, T. D. Cannon⁹, J. Lonnqvist^{1,5}, T. Paunio^{2,9}, J. Suvisaari¹, W. Hennah^{1,2,3}. 1) Department of Public Health Solutions, Mental Health Unit, National Institute for Health and Welfare, Helsinki, Finland; 2) Institute for Molecular Medicine FIMM, University of Helsinki, Helsinki, Finland; 3) Medicum, University of Helsinki, Finland; 4) The Social Insurance Institution, Research Unit, Finland; 5) Department of Psychiatry, University of Helsinki and Helsinki University Hospital, Helsinki, Finland; 6) Department of Public Health, University of Helsinki, Helsinki, Finland; 7) Molecular and Integrative Biosciences Research Program, University of Helsinki; 8) Department of Psychology, Yale University, New Haven, Connecticut, USA; 9) Department of Public Health Solutions, Genomics and Biomarkers Unit, National Institute for Health and Welfare, Helsinki, Finland.

BACKGROUND Studies of familial schizophrenia in the Finnish population have previously identified replicable associations for haplotypes and SNPs at the genomic loci of the phosphodiesterase 4D gene (*PDE4D*). However, causal variants have yet to be identified. We used a three-stage sequencing and genotyping study design in order to identify and replicate any potential functional variants at the *PDE4D* locus. **METHODS** The *PDE4D* genomic region was sequenced in 96 individuals from 20 families and identified 4,570 variants, of which 380 associated ($p<0.05$) with schizophrenia in these families. Based on bioinformatic functional prediction from location of the variant in exonic or regulatory regions, 14 variants were genotyped in 1,122 individuals from 301 families alongside 323 population controls (Stage 1). Three SNPs remained significant in their association to schizophrenia ($p<0.05$), and were thus further genotyped in an identically ascertained family cohort consisting of 1,696 individuals from 369 families, alongside an additional 794 unrelated population controls, and 2,733 individuals from other Finnish cohorts ascertained for major mental illnesses (Stage 2). **RESULTS** Two SNPs located in regions of predicted transcription factor binding sites were identified to significantly replicate in their association to schizophrenia: Stage 1, rs35278 OR=1.36, 95% CI 1.08-1.71; and rs165940 OR=1.40, 95% CI 1.11-1.77; Stage 2, rs35278 OR=1.18, 95% CI 1.04-1.34; and rs165940 OR=1.25, 95% CI 1.09-1.42. Further analysis revealed that rs165940 associates with a broad diagnosis of psychotic disorders ($p=0.025$, OR=1.18, 95% CI 1.07-1.30), and with an endophenotype of verbal learning and memory ($p=0.0078$, $\beta=-0.19$). Since both SNPs have potential regulatory functions, through their locations, we utilised the GTEx portal to establish their status as potential eQTLs. Only rs165940 (MAF=0.319) showed significant association with expression changes of *PDE4D* within the brain (cerebellum: $p\text{-value} = 0.04$, $\text{post-prob} = 0.9$ and $m\text{-value} = 1.4$). **CONCLUSIONS** Our findings support the role of *PDE4D* in psychiatric disorders, with replicable association in familial schizophrenia in Finland. Further characterisation suggests it plays a role in the psychotic and cognitive domains of these disorders. SNP rs165940, through its association and status as an eQTL, is the most likely principal variant of interest for being the functional mutation at this locus.

2496W

Non-Alzheimer's aging brain, behavior and biomarkers in down syndrome and mouse.

L. Dai¹, J. Schieving², J. Tippetts⁴, O. Abdullah³, M.C. Burbach³, A. VanHoek³, A. Ramirez³, J.O. Edgin⁵, M. Prigge⁵, J. Anderson⁷, J.R. Korenberg¹. 1) Pediatrics, Medical Genetics, University of Utah, Salt Lake City, UT; 2) Korenberg Research, University of Utah, Salt Lake City, UT; 3) Neurology/Korenberg Research, University of Utah, Salt Lake City, UT; 4) Bioengineering, University of Utah, Salt Lake City, UT; 5) Pediatrics, University of Utah, Salt Lake City, UT; 6) Psychology, University of Arizona, Tucson, AZ; 7) Radiology, University of Utah, Salt Lake City, UT.

A major challenge for therapeutics of aging is determining the mechanisms of normal aging and accelerated aging versus Alzheimer's disease (AD). Down syndrome (DS) is a major cause of genetic imbalance causing brain disease throughout the lifespan and provides an important genetic model to study aging and AD. Le Becker 1985 and Mann 1988 show basal ganglia calcification in both DS and AD patients, more severely with increased age in DS. The genetics, neurobiology and cognitive consequences of the pathology are unknown. To fill the gap, we used MRI in a young DS cohort ($n=22$; 21.5 ± 6.4 years), and showed lesions in the globus pallidus (GP) in 23% of full trisomy 21 subjects, showing early onset in DS. With respect to cognitive effects we found a significant correlation between the Intra-Extra Dimensional Set Shift test (IED) in young DS with lesions in the GP, suggesting a non-AD deficit related to the lesions. Further, the presence of a GP lesion in a 22yo individual, 22 yo with partial trisomy, suggested gene candidates between APP and MX1, whose overexpression determines the lesions. We noted no significant correlation between the lesions and the APOE genotyping results, suggesting that APOE does not contribute to the lesion in DS. Finally, we performed MRI and CT on the Ts65Dn mouse model of DS, aged (12-27 months) and found mild to severe lesions in the thalamus of all Ts65Dn mice, with minor lesions present only in aged litter mate wild type controls. Histologic calcium staining using the Von Kossa method defined the lesions as calcium bearing. This is the first report of calcification in the anterior nucleus of the thalamus of the Ts65Dn mouse, of interest because it is the target of the GPI, a region of the MRI holes in DS. Taken together, our results establish new genetic mechanisms for the cognitive deficits in DS at young ages, that could also affect the non-AD accelerated aging seen in DS. The prominent thalamic calcification in the Ts65Dn mouse provides an opportunity to study the genetic and neural mechanisms of accelerated aging and possibly its interaction with AD.

2497T

Calculating the effects of autism risk gene variants on dysfunction of five major biological processes identifies clinically-meaningful information.

O.J. Veatch¹, D.R. Mazzotti¹, J.S. Sutcliffe², R.T. Schultz³, T. Abel⁴, B. Tunc³, S.G. Assouline⁵, E.S. Brodtkin⁶, J.J. Michaelson⁷, T.K. Nickl-Jockschat⁸, Z.E. Warren⁹, B.A. Malow⁹, A.I. Pack¹. 1) Center for Sleep and Circadian Neurobiology, Department of Medicine, University of Pennsylvania, Philadelphia, PA; 2) Vanderbilt Genetics Institute, Department of Molecular Biology and Biophysics, Vanderbilt University Medical Center, Nashville, TN; 3) Center for Autism Research, Children's Hospital of Philadelphia, Philadelphia, PA; 4) Iowa Neuroscience Institute, University of Iowa, Iowa City, IA; 5) Belin-Blank Center for Gifted Education and Talent Development, University of Iowa, Iowa City, IA; 6) Department of Psychiatry, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA; 7) Department of Psychiatry, University of Iowa, Iowa City, IA; 8) Department of Pediatrics, Vanderbilt University Medical Center, Nashville, TN; 9) Department of Neurology, Vanderbilt University Medical Center, Nashville, TN.

Autism spectrum disorders (ASD) are neurodevelopmental conditions that encompass a wide-range and severity of symptoms. Quantitative ASD phenotypes are influenced by inherited common, rare, and *de novo* genetic variation. However, the details of how genetic variation contributes to variable symptomatology are still unclear. This creates a major challenge for the field in order to translate vast amounts of genetic data into clinically-useful information. To determine if current evidence could be useful to informing treatment of symptoms in ASD, we identified variants in known ASD risk genes that were predicted damaging to the encoded protein products using whole-exome sequencing available for 2,392 individuals with ASD whose data were included in the Simons Simplex Collection. We then developed an approach to calculate the likelihood of genetic dysfunction in five Gene Ontology-defined biological processes that were significantly enriched for ASD candidate genes. These processes included nervous system development, synaptic signaling, cognition, regulation of membrane potential, and chromosome organization. Unsupervised clustering was used to subgroup individuals based on scores reflecting the likelihood of genetic dysfunction in each enriched process. Dysfunction in genes related to cognition distinguished a distinct subgroup of individuals ($n=885$) with lower IQs ($pFDR=4.3 \times 10^{-3}$), reduced adaptive behavior ($pFDR=7.0 \times 10^{-3}$), and increased social deficits ($pFDR=4.3 \times 10^{-2}$) when compared to individuals with no evidence of dysfunction in these genes ($n=1,488$). Variation in the *PTGS2*, *ABCA7* and *SHANK3* genes was strongly associated with assignment to the subgroup defined by cognition gene dysfunction ($p < 1.0 \times 10^{-5}$). In particular, an inherited stop-gain variant in the pharmacogene, *PTGS2*, which encodes Cyclooxygenase-2 (COX2), was associated with having intellectual disability defined as an $IQ < 70$ ($OR=1.43$, $95\%CI=1.04$, 1.97 , $p=4.0 \times 10^{-2}$) and increased risk for irritable bowel syndrome ($OR=5.38$, $95\%CI=1.85$, 15.58 , $p=2.0 \times 10^{-3}$). We expect that screening for deleterious variants in the subset of ASD candidate genes involved in cognition may be useful for identifying genetic factors that are clinically informative and should be prioritized for functional follow-up. This has implications in designing more comprehensive genetic testing panels and may provide the basis for more informed treatment in ASD.

2498F

Deep phenotyping revealed differential global developmental milestone deficits in siblings with 15q11.2 microdeletion. *K.M.F Uddin^{1,2}, M.M Rahman³, B. Mohammed³, H. Akter³, N.J Dity², M.A Rahaman², M.R Islam², M.A Rahaman², M.A Baqu^{1,2}, K.A Islam³, M. Nasir¹, M. Uddin^{4,5}*. 1) Biochemistry, Holy Family Red Crescent Medical College, Dhaka, Bangladesh; 2) NeuroGen Technologies Ltd., Dhaka, Bangladesh; 3) Bangabandhu Sheikh Mujib Medical University, Dhaka, Bangladesh; 4) Mohammed Bin Rashid University of Medicine and Health Sciences, College of Medicine, Dubai, United Arab Emirates; 5) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Ontario, Canada.

Global developmental delay (DD) is a highly heterogeneous and complex condition that impacts at least ~3% of general population. Genetics contributes as a significant risk factor into the etiology of DD. Copy number variants (CNVs) are a genetic variant that has been shown to be major source of risk factor. We have analyzed a quartet family in Bangladesh where two siblings were impacted by DD. We have applied a custom array comparative hybridization (aCGH) designed to target the genomic regions that are outside of the common CNVs. The array comprised of 32,696 probes targeting regions impacted by rare deletions or duplications with a resolution of >100kb. Circular binary segmentation algorithm was used on the normalized $-\log_2$ values to detect CNVs. Here we describe the two cases of a non-consanguineous family, one is eleven years old boy and another is two years old boy with a confirmed diagnosis of DD with intellectual disability. aCGH analysis revealed a 0.460 Mb-microdeletion in both cases at chromosome 15q11.2. There was a history of maternal epilepsy that was reported at early age (eight years old) and had been treated on Tegretol drug. We have conducted deep phenotyping and clinical observation shows delay in speech, cognitive and physical developmental milestones, low IQ = 56, in the elder sibling. In contrast, the younger sibling's clinical presentations are significantly different that includes facial dysmorphism, microcephaly, delay in speech, cognitive and physical developmental milestones and low IQ. Our finding show, 15q11.2 microdeletion alone might not contribute to the entire phenotypic spectrum and the differential clinical presentation can be due to the differences in genetic background or a second penetrant variant. .

2499W

Mutation identified for of ADHD families in the US Hispanic population by whole exome sequencing. *P. Acevedo¹, F. Trevino¹, J. Peralta¹, K. Wang², J. Blangero¹, R. Crutchfield¹, B. Sur¹, C. Xu¹*. 1) University of Texas Rio Grande Valley, Brownsville, TX; 2) East Tennessee State University.

Introduction: Attention deficit hyperactivity disorder (ADHD) is a common psychiatric disorder that has a worldwide prevalence of 4%. The disabilities, morbidities, and economic burden inherent to ADHD represent a major public health problem and, consequently, an important area of focus in health research. Evidence from family, twin, and adoption studies on ADHD indicates a high degree of heritability. The recent advancement of molecular techniques has facilitated identification of several genetic variants associated with ADHD in predominately Caucasian populations. There has been limited research in the U.S Hispanics/Latino populations. Whole exome sequencing (WES), as a rapid and reliable means, has identified causative mutations for a number of diseases. Therefore, we applied WES to identify pathogenic mutations for ADHD from the US Latino population. Material and Methods: 29 subjects (16 affected with ADHD) from 9 families recruited from the U.S. Hispanic/Latino population from the Departments of Psychiatry/Neurology and Pediatrics at the Texas Tech University Health Sciences Center. WES was performed by the Illumina Nextera Rapid Capture Exome Enrichment kits, then bioinformatics predictions were conducted. Results: A total of 555,801 variants were observed in all 29 subjects. After filter, a number of pathogenic genes/mutations were identified in affected family members with ADHD, particularly one mutation (NM_000152.3:c.858+8G>A) in glucosidase alpha, acid (GAA) gene on 17q25.3 was discovered in seven affected subjects from three families with ADHD from the U.S. Hispanic population. Defects in this gene have been identified as the underlying cause of glycogen storage disease II, also known as Pompe's disease, autism and ADHD-like phenotypes. The consequence of this mutation may be involved in gene splicing. To identify high impact genes and biological pathways, further bioinformatics analysis, "hotspot" analysis, and literature-based searches are ongoing currently. Conclusion: This WES led to the discovery of additional ADHD genes in the U.S. Latino population; however, future validation and segregation analyses using a large sample are needed to confirm the current findings.

2500T

Genome-wide association of word reading: Overlap with risk genes

for neurodevelopmental disorders. K. Price^{1,2,3}, K.G. Wigg², Y. Feng², K. Blokland³, M. Wilkinson³, G. He⁴, E. Kerr^{5,6}, S. Guger⁶, M. Lovett^{6,6}, L. Strug^{4,7}, C.L. Barr^{1,2,3}. 1) Department of Physiology, University of Toronto, Toronto, Ontario, Canada; 2) Krembil Research Institute, University Health Network, Toronto, Ontario, Canada; 3) The Hospital for Sick Children, Toronto, Ontario, Canada; 4) Genetics & Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada; 5) Department of Paediatrics, University of Toronto, Toronto, Ontario, Canada; 6) Department of Psychology, The Hospital for Sick Children, Toronto, Ontario, Canada; 7) Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, Canada.

Specific reading disabilities (RD) represent a major health, social, and educational handicap. RD can cause significant academic impairment in children and subsequent occupational underachievement in adults. Comorbid psychiatric disorders are common in children with RD (e.g., mood disorders, anxiety, and attention-deficit/hyperactivity disorder), further impacting academic achievement and social development. RD is a neurobiological learning disability characterized by difficulties with accurate/fluent word recognition, poor spelling, and decoding abilities. Despite high heritability, there have been few published GWAS studies. We performed a GWAS for a measure of word reading on two samples, one a family based sample selected for reading difficulties from Toronto and the other a population based sample called the Philadelphia Neurodevelopmental Cohort (PNC). The final analysis was performed with >5 million single nucleotide polymorphisms (SNPs) on 624 children from the Toronto sample and 4629 individuals from the PNC. The results from the SNP based analyses did not identify genome-wide significant results, however results near significance, indicate overlap for genes previously identified in GWAS for educational attainment and risk factors for neurodevelopmental disorders, particularly autism spectrum disorder (ASD). Known or suspected ASD was an exclusion criteria for the Toronto sample, thus the overlap with autism suggests shared genetic risk for ASD, possibly for shared genetic contributions to language difficulties. Notable among the identified genes were *FOXP1* and *RBFOX1*. *FOXP1* is a transcription factor that forms heterodimers with *FOXP2*, the first gene shown to contribute to speech and language. *FOXP1* has been identified as contributing to speech delay, language, and cognitive impairment in addition to ASD. The *RBFOX* family of RNA splicing proteins have been previously implicated in ASD, language and Rolandic epilepsy, a benign epilepsy syndrome associated with a spectrum of learning, language, and reading problems. The findings implicating *FOXP1* and *RBFOX1* are particularly interesting because deregulation would predict global change in transcription and splicing in neural cells. Further, pathway analysis identified the panther pathway "axon guidance mediated by Netrin". The findings from pathway analysis support previous evidence from neuroanatomical studies for altered neuronal migration/axon guidance as a contributor to RD.

2501F

Comparison of two phenotypes to identify SNPs associated with poor reading skills.

H. Lancaster¹, V. Dinu², J. Li³. 1) Speech and Hearing Science, Arizona State University, Tempe, AZ; 2) Biomedical Informatics, Arizona State University, Tempe, AZ; 3) School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, Tempe.

Reading difficulties account for 85% of children receiving special education services. While the genetic basis of reading disorders is largely unknown, there is substantial evidence for an underlying genetic mechanism. Understanding the genetic mechanism of reading difficulties is crucial to developing earliest identification methods. The objective of this study was to compare two phenotypes for reading difficulties, a case-control and a continuous phenotype, to determine if there was convergence between the phenotypes and which SNPs were associated with reading difficulties. We used behavioral and genetic data provided by Avon Longitudinal Study of Parents and Children. There were ten measures of reading skills, including measures of single word reading, non-word reading, phonological awareness, spelling, reading fluency, and reading comprehension. A binary case-control phenotype was assigned if a child had three out of ten reading measures one standard deviation below average. There were 1,215 cases (58.76% male) and 6,856 controls (49.62% male). We derived a continuous reading skill score using confirmatory factor analysis. We had data for 3,933 children (51% male) for the reading skill score. Plink was used to perform logistic (case-control) and linear (continuous) regressions for genome wide association by chromosome. Results were compared between phenotypes. Preliminary results revealed that there 52 SNPs identified by both phenotypes before multiple correction. After multiple test correction, there were no SNPs in common between the two phenotypes and there were no significant SNPs for the binary case-control phenotype; however, there were ten SNPs that remained significant for the reading skill phenotype, six on chromosome 14 (rs74491386, rs45547933, rs77940109, rs78487170, rs78277203, rs192227868) and four on chromosome X (rs181384543, rs191271648, rs185462130, rs185003220). Five of the SNPs on chromosome 14 are intronic variants for PPP2R5C and the SNPs on chromosome X are intronic variants for PTCHD1-AS. Both of these genes are highly expressed in the brain, anterior cingulate cortex and cerebellum respectively. These preliminary analyses suggest that viewing reading ability as a continuous skill, similar to education attainment, is a fruitful line of investigation. Preliminary results suggest a complex biological mechanism plays a part in the genetics of poor reading.

2502W

Characterization of the genetic architecture of intellectual disability in northern Finland. L.M. Urpa¹, M.I. Kurki^{1,2,3}, E. Saarentaus¹, O. Pietiläinen^{2,4}, E. Hämäläinen¹, P. Gormley^{2,3}, M. Torniainen-Holm^{1,5}, V. Salomaa⁶, D. La^{2,3}, S. Kerminen¹, M. Pirinen^{1,6,7}, J. Suvisaari⁸, E. Rahikkala^{8,9,10}, R. Keski-Filppula^{8,9,10}, M. Rauhalä¹¹, S. Korpi-Heikkilä¹¹, J. Komulainen-Ebrahim¹², H. Helander¹², P. Vieira¹², J.S. Moilanen^{8,9,10}, J. Kärkkö¹¹, O. Kuusimä^{1,8,9,10}, M.J. Daly^{1,2,3,13,14}, A. Palotie^{1,2,3,13,14}. 1) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; 2) The Stanley Center for Psychiatric Research, The Broad Institute of MIT and Harvard, Cambridge, MA, USA; 3) Psychiatric & Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, USA; 4) Department of Stem Cell and Regenerative Biology, University of Harvard, Cambridge, Massachusetts, USA; 5) National Institute for Health and Welfare, Helsinki, Finland; 6) Department of Public Health, University of Helsinki, Helsinki, Finland; 7) Helsinki Institute for Information Technology HIIT and Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland; 8) PEDEGO Research Unit, University of Oulu, Oulu, Finland; 9) Medical Research Center, Oulu University Hospital, University of Oulu, Oulu, Finland; 10) Department of Clinical Genetics, Oulu University Hospital, Oulu, Finland; 11) Northern Ostrobothnia Hospital District, Center for Intellectual Disability Care, 90220 Oulu, Finland; 12) Department of Children and Adolescents, Oulu University Hospital, Medical Research Center Oulu, University of Oulu, Oulu, Finland; 13) Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA; 14) Department of Neurology, Massachusetts General Hospital, Boston, MA, USA.

Intellectual disability (ID) is a genetically and clinically heterogeneous disorder, making diagnosis and treatment of ID patients difficult. The Northern Finnish Intellectual Disability (NFID) study aims to identify inherited variation contributing to ID. Finland is known to have a small founder population that has undergone several bottlenecks and genetic drift, which is particularly accentuated in the late-settlement area in northern Finland. Simulation studies show that in such populations, variants which are acted on by negative selection can exist at higher frequencies than in older populations, and indeed we see an increase in the frequency of some loss-of-function (LoF) mutations in Finland compared to other European populations. Within Finland, ID is reported to have higher prevalence in the northern late-settlement areas, suggesting inherited genetic factors contribute to the risk of ID. These inherited genetic factors are also likely to be more important in the etiology of mild ID, enriched in our cohort, rather than in severe ID. Previous work by Kurki *et al.* (bioRxiv 332023) have found in the NFID cohort an enrichment of protein-truncating variants in LoF intolerant genes, and identified a homozygote variant in the *CRADD* gene associated with a specific syndrome with ID and pachygyria. This variant is 50 times higher in the Finnish population compared to non-Finnish European populations, demonstrating the benefits of utilizing population isolates in rare variant analysis of diseases under negative selection. We have also found that common variant load contributes to ID in the NFID cohort, and that the educational attainment (EDU), intelligence quotient (IQ) and schizophrenia polygenic risk scores (PRS) were associated with ID, explaining 0.9%, 0.5%, and 0.6% of the heritability on the liability scale, respectively. The EDU PRS also appears to explain more heritability in mild (2.2%) than severe ID (0.6%) subsets of patients. We present the extended analysis of the NFID cohort with genotype and exome sequence data from 742 ID patients and 653 family members. We will expand this analysis to the exome sequence of 938 total ID patients.

2503T

Polygenic risk scores distinguish speech and language subgroups in the Cleveland Family Study Cohort. P. Benček¹, R. Igo¹, H. Voss-Hoynes¹, B. Truitt¹, L. Freebairn², J. Tag², B. Lewis², C. Stein¹, S. Iyengar¹. 1) Department of Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH; 2) Department of Psychological Sciences, Case Western Reserve University, Cleveland, OH.

Speech and language are evolved human attributes whose neural basis resides in the Broca (speech and language output) and Wernicke (language comprehension) areas in the brain's left hemisphere along the Sylvian fissure. Genetic analysis of neurodevelopmental disorders (e.g., dyslexia, autism, schizophrenia, and various syndromes), and evolutionary comparisons with multiple species have contributed to our understanding of how children acquire speech and language naturally from infancy to adolescence. A distinctive insight into the neural basis of speech and language comes from the 4-10% of children without other diagnoses, predominantly affected with speech and language disorders; little of this is explained by FOXP2. Speech sound disorders (SSD) refer to difficulty in producing sounds or syllables, while language impairment (LI) represents an inability to comprehend or utilize language appropriately; SSD and LI are frequently comorbid. We and others have shown that children with SSD+LI, and children with childhood apraxia of speech (CAS), a severe form of SSD, have poorer performance on several endophenotypes than children with only SSD, arguing for a continuum of risk. We hypothesized that risk variants associated with particular endophenotypes accrue with increasing severity of speech and language deficits. We generated polygenic risk scores (~950 variants (range=[890,1024]) at $P \leq 0.001$) using effect estimates from genome-wide association studies (GWAS) of articulation, expressive and receptive language, vocabulary, reading, and spelling (range of N per GWAS=[298,432]). Risk scores were constructed with index variants showing the strongest association in a region, removing other variants in linkage disequilibrium ($r^2 > 0.5$). We used a linear mixed model to model the relationship between polygenic risk score and clinical group (no disorder, LI only, SSD only, SSD+LI, CAS in increasing order of severity), controlling for familial relationship (total N=686). Clinical subgroup was strongly associated with polygenic risk score for all endophenotypes ($P < 1 \times 10^{-6}$). Furthermore, we observed the greatest risk load in the most severely affected group, CAS, and the second greatest load was usually in children with SSD+LI. These results demonstrate that increasing genetic burden is associated with greater severity of SSD and LI, suggesting that children seen in clinical practice have significant heterogeneity and may need personalized treatment paradigms.

2504F

Whole exome sequencing and functional validation in zebrafish identify novel genes for mitochondrial disorders. P. Arumugam¹, A.K. Meena², S.U. Megha³, C. Venkatapathi⁴, G. Sandeep⁴, K. Thangaraj¹. 1) CSIR-Centre for Cellular and Molecular Biology, Hyderabad, India; 2) Department of Neurology, Nizam's Institute of Medical Sciences, Hyderabad, India; 3) Department of Pathology, Nizam's Institute of Medical Sciences, Hyderabad, India; 4) Department of Neurology, Bhimavaram Hospitals, Bhimavaram, India.

Mitochondrial disorders have emerged as a common cause of inherited disease. They are caused by mutations which reside in either nuclear or mitochondrial DNA (mtDNA). Their diagnosis and treatment remains challenging. The prevalence of mitochondrial disorders in India is more than the projected data available in the literature. In India, with a huge number of endogamous ethnic groups, the transmission rate of mitochondrial disorder is more. We used Sanger sequencing in 325 patients with mitochondrial disorders to identify pathogenic mutations in complete mtDNA and common nuclear genes. We have identified pathogenic mutations in mtDNA and common nuclear genes in 27% of the cases. However, the majority of cases (>70%) remain genetically undetermined. Therefore, we performed whole exome sequencing of 30 patients with mitochondrial disorders who do not have mutations in both mtDNA and common nuclear genes with phenotype of mitochondrial disorders, identified several novel genes involved in mitochondrial functions including *ALKBH1*, Cystathionine-Beta-Synthase (*CBS*) and Androgen Receptor (*AR*) cause severe autosomal recessive mitochondrial disorders, furthermore we have sequenced complete *ALKBH1*, *CBS* and *AR* genes in 275 patients with mitochondrial disorders, found biallelic *ALKBH1*, *CBS* and *AR* mutations in 14 patients cause severe recessive mitochondrial disorders. These mutations disrupted the function of the gene, showed complete co-segregation with the disease phenotype and absent in 650 ethnically matched control chromosomes. *ALKBH1* was recently discovered as a demethylase for mitochondrial tRNA, a new epigenetic modification. We found mutations in *ALKBH1* cause combined oxidative phosphorylation deficiency and severe multisystemic mitochondrial disorders. Knock-down of *alkbh1* in zebrafish using morpholino oligonucleotides (MOs) showed severe developmental abnormalities. RT-PCR analysis of zebrafish *alkbh1* expression in adult organs and during development showed that the *alkbh1* mRNA is highly enriched in brain, testis, intestine, muscle, heart and it is highly enriched in the early stage of zebrafish embryos. Interestingly, the knockdown of zebrafish *alkbh1* also causes developmental abnormalities via p53-dependent apoptosis. Our study shows the usefulness of exome sequencing and functional validation in zebrafish for identification of novel genetic causes of inherited mitochondrial disease.

2505W

Improved methods for molecular genetic understanding of mitochondrial diseases to facilitate clinical diagnostics. I. Barbosa¹, K. Thompson², O.V. Poole³, M. Champion⁴, R.D.S. Pitceathly⁵, C. Fratter⁶, J. Poulton⁶, R. McFarland⁷, R.W. Taylor², C. Deshpande², M.A. Simpson¹. 1) Faculty of Life Sciences and Medicine, School of Basic and Medical Biosciences, King's College London School of Medicine, London, UK; 2) Wellcome Centre for Mitochondrial Research, Newcastle University, Newcastle upon Tyne, UK; 3) MRC Centre for Neuromuscular Diseases, University College London Institute of Neurology and National Hospital for Neurology and Neurosurgery, London, UK; 4) Evelina London Children's Hospital, Guys and St Thomas' NHS Foundation Trust, London, UK; 5) Oxford Medical Genetics Laboratories, Oxford University Hospitals NHS Trust, Oxford, UK; 6) Nuffield Department of Women's and Reproductive Health, University of Oxford, The Women's Centre, Oxford; 7) Clinical Genetics Department, Guys and St Thomas' NHS Foundation Trust, London, UK.

Approximately 1300 genes in the nuclear genome are known to encode components of the mitochondrial proteome. To date pathogenic variants in only a quarter of these genes have been reported to be associated with clinical mitochondrial disease. This genetic heterogeneity and the extensive clinical variability, pose challenges to establish the molecular genetic basis of mitochondrial disease. To establish best practice protocols for detecting pathogenic variations in patients with mitochondrial disease, we have undertaken whole exome sequencing of patients with clinical and biochemical evidence of mitochondrial disease in whom previous genetic investigations excluded mitochondrial DNA rearrangements and mutations. We established an analytical protocol to efficiently assist the interpretation of the patients' exome variant profiles in the context of nuclear encoded mitochondrial genes. Overall, our strategy yielded a diagnosis in more than half of the patients studied. About 30% of patients were found to harbour pathogenic mutations in established nuclear encoded mitochondrial disease genes. In approximately 25% of patients, pathogenic variants were identified in genes with no previous reports of association with mitochondrial disease. In the latter group, functional studies and the identification of causative mutations in the same gene in unrelated individuals led to the confirmation of novel genetic causes of mitochondrial disease. These include mutations in three genes (*HTRA2*, *GFM2*, *RTN4IP1*) with known mitochondrial function for which there was no previous association with a clinical mitochondrial disorder. As a result of this study, families have been able to benefit from prenatal genetic testing. Here we present the newly identified mitochondrial disease genes from our study with evidence supporting causality, as well as novel mutations in previously reported genes. We also present the bioinformatics strategies employed to improve diagnosis, including phenotype driven ranking as the basis of our approach to success, and finally highlight limitations of some approaches.

2506T

Gene expression profiling of dorsal root ganglion in the mouse whole-body inducible *Clec16a* knockout model. M. Bakay, R. Pandey, H. Hain, B. Strenkowski, H. Hakonarson. Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA.

CLEC16A is genetically linked with twelve distinct autoimmune disorders including multiple sclerosis. We and others reported that *Clec16a* knockout (KO) and *Clec16a* mutant mice display a neurologic disease. Specifically, our whole-body inducible *Clec16a* KO mice exhibit a neuronal phenotype including tremors, impaired gait and dystonic postures with an increase in severity over time. However, despite strong evidence implicating CLEC16A in the nervous system its functional role in it is not determine. We hypothesized that an RNA-based expression signature might provide us clues to the mechanism of CLEC16A action in the nervous system. Towards this end, we have expression profiled the dorsal root ganglia (DRG) from four control and six *Clec16a* KO mice with Affymetrix Clariom D Assay. The Transcriptome Analysis Console (TAC 4.0) Software was used to analyze data and visualize results. We also used both, Ingenuity Pathway Analysis and iPathwayGuide solutions to perform functional analysis. Among the total 65,956 genes on the Mouse Clariom D Assay expression of 29,402 (45%) genes was detected in the DRGs. The array data distribution indicated that expression of most genes was not strongly changed. We used a 2-fold change in signal intensity in combination with $p\text{-value} < 0.05$ as a cut-off to consider the differential expression of a gene as significant. We identified 403 differentially expressed genes DEG (258 up- and 145 down-regulated) in DRG from *Clec16a* KO mice when compared to control. Among the down-regulated genes, 60% of the known genes were mitochondrial and metabolic genes. The majority of up-regulated genes are genes involved in signaling, including PI3K-mTOR, TGF-beta, MAPK and cytokine signaling. Diseases and conditions that are overrepresented in the set of differentially expressed genes were preferentially represented by metabolic, mitochondrial and neurodegenerative or nervous system diseases. Our results revealed that lack of *Clec16a* expression is resulting in downregulation of metabolic pathways, proteasome and oxidative phosphorylation; and in upregulation of several signaling pathways leading to altered immune response and increased apoptosis. This could suggest that drugs modulating cytokine-mediated proinflammatory responses may be effective in treating or preventing neurological autoimmune disorders such as multiple sclerosis in individuals with risk associated CLEC16A variants.

2507F

A frequent variant in the Japanese population is acting as a quasi-Mendelian allele in rare retinal ciliopathies. M. Quinodoz¹, K. Nikopoulos¹, K. Cisarova², H. Koskiniemi-Kuendig¹, N. Miyake², P. Farinelli³, M.I. Khan^{3,4}, A. Prunotto¹, M. Akiyama⁵, Y. Kamatani⁶, C. Terao⁵, F. Miya⁶, Y. Ikeda⁷, S. Ueno⁸, N. Fuse⁹, A. Murakami¹⁰, H. Terasaki⁸, K.-H. Sonoda⁷, T. Ishibashi⁷, M. Kubo¹¹, F.P.M. Cremers^{3,4}, N. Matsumoto², K.M. Nishiguchi^{12,13}, T. Nakazawa^{12,13}, C. Rivolta^{1,14}. 1) Department of Computational Biology, Unit of Medical Genetics, University of Lausanne, Lausanne, Switzerland; 2) Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan; 3) Department of Human Genetics, Radboud University Medical Center, Nijmegen, the Netherlands; 4) Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, The Netherlands; 5) Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 6) Department of Medical Science Mathematics, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan; 7) Department of Ophthalmology, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan; 8) Department of Ophthalmology, Nagoya University Graduate School of Medicine, Nagoya, Japan; 9) Department of Integrative Genomics, Tohoku Medical Megabank Organization, Sendai, Japan; 10) Department of Ophthalmology, Juntendo University School of Medicine, Tokyo, Japan; 11) RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 12) Department of Ophthalmology, Tohoku University Graduate School of Medicine, Sendai, Japan; 13) Department of Advanced Ophthalmic Medicine, Tohoku University Graduate School of Medicine, Sendai, Japan; 14) Department of Genetics and Genome Biology, University of Leicester, Leicester, United Kingdom.

Hereditary retinal degenerations (HRDs) are rare diseases leading to progressive visual impairment and blindness. They are generally transmitted as monogenic traits, obeying Mendel's laws of inheritance, and affect indiscriminately individuals from all ethnic groups. However, compared to other populations, Japanese display a unique landscape of HRD mutations. Following the genomic screening of 331 unrelated Japanese patients, we identified two alleles in the ciliary gene *RP1* that were enriched in patients but were also relatively prevalent in the general population. The first was determined by the insertion of an *Alu* element in the last exon of the gene (allele frequency in the Japanese population = 0.1%), while the other was a nonsense variant (p.Arg1933*) in the same exon. The *Alu* insertion which is causing a frameshift and a truncated protein (p.Tyr1352Alafs*9) was found homozygously in six patients and co-segregated with the disease in the families as a classical Mendelian recessive allele. However, this insertion was also found heterozygously in 10 patients. Six of them were in fact compound heterozygous for either of two other changes in *RP1*: a novel frameshift mutation (p.Cys1399Leufs*5, two individuals) and a nonsense variant (p.Arg1933*, four individuals). Both variants segregated within their respective families. This latter nonsense was found to be almost polymorphic in Japan (allele frequency = 0.6%), did not cause disease in homozygosis or heterozygosis, but was considerably enriched in HRD patients (allele frequency = 2.1%, i.e. a 3.5-fold enrichment; $p\text{-value} = 1.29 \times 10^{-4}$) and could rarely act as a Mendelian mutation, *in trans* with other pathogenic variants like the *Alu* insertion. This variant is therefore acting as a quasi-Mendelian allele. After having identified 10 patients carrying the nonsense variant heterozygously without any second hit, we looked at enrichment of rare functional or disruptive alleles affecting retinal ciliary proteins elsewhere in the genome. We found indeed an enrichment of these alleles in patients vs controls supporting the hypothesis of a ciliary mutational load previously proposed in ciliopathies. Our results suggest that rare conditions such as HRDs can be paradoxically determined by relatively common variants, following a quasi-Mendelian model that could fill the gap that currently exists between monogenic and complex diseases.

2508W

Genome-wide association analyses identify 139 loci associated with macular thickness in the UK Biobank cohort. X. Gao, H. Huang, H. Kim. Department of Ophthalmology and Visual Sciences, University of Illinois at Chicago, Chicago, IL.

The macula, located near the center of the retina in the human eye, is responsible for providing critical functions, such as central, sharp vision. Structural changes in the macula are associated with many ocular diseases, including macular degeneration and glaucoma. Although macular thickness is a highly heritable trait, no genome-wide association study (GWAS) has been performed. Here, we describe the first GWAS of macular thickness, which was measured by spectral-domain optical coherence tomography using close to 68,000 participants from the UK Biobank cohort. We identified 139 unique genetic regions associated with macular thickness at genome-wide significance ($P < 5 \times 10^{-8}$). The most significant loci were *LINC00461* ($P = 5.1 \times 10^{-120}$), *TSPAN10* ($P = 1.2 \times 10^{-118}$), *RDH5* ($P = 9.2 \times 10^{-105}$), and *SLC6A20* ($P = 1.4 \times 10^{-71}$). Results from gene expression demonstrated that these genes are highly expressed in the retina, further providing functional support. Other hits included previously reported age-related macular degeneration genes (*NPLOC4*, $P = 1.7 \times 10^{-103}$; *RAD51B*, $P = 9.1 \times 10^{-14}$) and known primary open-angle glaucoma genes (*SIX6*, $P = 4.4 \times 10^{-11}$; *CDKN2B-AS1*, $P = 4.6 \times 10^{-10}$). Through cross-phenotype analysis, these genetic loci also exhibited pleiotropic effects with myopia, neurodegenerative diseases (e.g. Parkinson's disease, schizophrenia, and Alzheimer's disease), cancer (e.g. breast, ovarian, and lung cancers), and metabolic traits (e.g. body mass index, waist circumference, and type 2 diabetes). Our findings provide the first insight into the genetic architecture of macular thickness and may further elucidate the pathogenesis of macular degeneration and glaucoma.

2509T

Combining information across multiple traits dramatically improves polygenic risk profiling for glaucoma. S. MacGregor¹, X. Han¹, J. An¹, J.S. Ong¹, M.H. Law¹, E. Souzeau², A. Qassim², D.A. Mackey^{3,4}, P. Gharahkhani¹, A.W. Hewitt^{3,5}, J.E. Craig². 1) QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia; 2) Department of Ophthalmology, Flinders University, Flinders Medical Centre, Bedford Park, Australia; 3) Menzies Institute for Medical Research, University of Tasmania, Australia; 4) Lions Eye Institute, Centre for Vision Sciences, University of Western Australia, Australia; 5) Centre for Eye Research Australia, University of Melbourne, Australia.

Primary open-angle glaucoma (POAG), one of the most heritable human diseases, is a leading cause of blindness worldwide. Whilst current treatments are generally effective in preventing disease progression, many patients present when irreversible damage has already occurred. Given the large genetic component to risk, gene based risk prediction may improve outcomes by identifying early those at highest risk, whilst reducing monitoring costs for those at low risk. It is now clear that glaucoma has a polygenic genetic architecture and hence very large sample sizes are required for accurate prediction. Current sample sizes in glaucoma genome-wide association studies are limited. Two key endophenotypes for glaucoma are intraocular pressure (IOP) and vertical cup disc ratio (VCDR) - both show significant genetic correlation with POAG. We built a genetic risk prediction model using 7,947 glaucoma cases and 119,318 controls from UK Biobank. To improve prediction accuracy, we combined glaucoma with data from the correlated endophenotypes using a multivariate approach (MTAG). Endophenotype data comprised 133,492 individuals with IOP from UK biobank and a previously published International Glaucoma Genetic Consortium data set, with a subset also measured for VCDR. We tested our genetic risk score in 1,734 advanced glaucoma and 2,938 controls from the Australia and New Zealand Registry of Advanced Glaucoma. Prediction accuracy increased dramatically when the endophenotype data were included. Based on the multivariate prediction approach, individuals in the top decile for the genetic risk score were at 15 fold increased risk of glaucoma relative to the bottom decile. Amongst the cases, individuals in the top decile were diagnosed 7 years earlier than those in the bottom decile.

2510F

Unique spectra of deafness-associated mutations in Mongolians provide insights into the genetic relationships between among Eurasian populations. Y.H. Lin^{1,2}, J. Erdenechuluun^{3,4}, Y.H. Lin^{2,5}, K. Ganbat^{4,6}, Z. Makhba⁶, C.Y. Tsai^{2,6}, W.C. Hsu², C.C. Wu^{2,7}, P.L. Chen^{2,7}. 1) Graduate Institute of Molecular Medicine, National Taiwan University College of Medicine, Taipei, Taiwan; 2) Dept of Otolaryngology, National Taiwan University Hospital, Taipei, Taiwan; 3) Dept of Otolaryngology, Mongolian National University of Medical Sciences, Ulaanbaatar, Mongolia; 4) The EMJJ Otolaryngology Hospital, Ulaanbaatar, Mongolia; 5) Graduate Institute of Medical Genomics and Proteomics, National Taiwan University College of Medicine, Taipei, Taiwan; 6) Dept of Otolaryngology, National Center for Maternal and Child Health, Ulaanbaatar, Mongolia; 7) Dept of Medical Genetics, National Taiwan University Hospital, Taipei, Taiwan.

More than 50% of idiopathic sensorineural hearing impairment (SNHI) cases have a genetic etiology. From the epidemiological perspective, mutations of three deafness genes: *GJB2*, *SLC26A4*, and *MT-RNR1*, are much more prevalent than those of other genes worldwide. However, mutation spectra of common deafness genes differ significantly across different populations. Here, we performed comprehensive genetic examination and haplotype analyses in 188 unrelated Mongolian families with idiopathic SNHI, and compared their mutation spectra and haplotypes to those of other European and Asian cohorts. We achieved genetic diagnoses in 18 (9.6%) of the 188 families, including 13 with bi-allelic *GJB2* mutations, three with bi-allelic *SLC26A4* mutations, and two with homoplasmic *MT-RNR1* m.1555A>G mutation. Interestingly, three *GJB2* mutations prevalent in other populations, including c.35delG in Caucasians, c.235delC in East Asians, and c.-23+1G>A in Southwest and South Asians, were simultaneously detected in Mongolian patients. Haplotype analyses further confirmed founder effects for each of the three mutations, indicating that each mutation derived from its ancestral origin independently. By demonstrating the unique spectra of deafness-associated mutations, our findings may have important clinical and scientific implications for refining the molecular diagnostics of SNHI in Mongolian patients, and for elucidating the genetic relationships among Eurasian populations.

2511W

Ocular disease mechanisms elucidated by genetic regulation of gene expression in human retinal pigment epithelium. B. Liu¹, M. Calton², N. Abell³, B. Balliu⁴, G. Benchorin², M. Gloudemans³, J. Hu⁴, D. Bok⁴, S. Montgomery⁵, D. Vollrath². 1) Department of Biology, Stanford University, Palo Alto, CA; 2) Department of Genetics, Stanford University, Palo Alto, CA; 3) Program in Biomedical Informatics, Stanford University, Palo Alto, CA; 4) Department of Ophthalmology, Jules Stein Eye Institute, UCLA, Los Angeles, CA; 5) Department of Pathology, Stanford University, Palo Alto, CA.

The importance of vision to humans has motivated the identification of hundreds of GWAS loci for complex ocular disorders such as age-related macular degeneration (AMD) and myopia, but the molecular mechanisms are unknown for most loci. The retinal pigment epithelium serves important roles in ocular development and retinal homeostasis. To identify candidate causal genes for complex eye phenotypes, we genotyped 23 unrelated human fetal retinal pigment epithelium (fRPE) samples, and profiled their transcriptomes under culture conditions that modulated cellular energy metabolism (glucose or galactose). Comparison with GTEx revealed 130 fRPE-enriched genes (>4 standard deviations), including known fRPE markers such as *RPE65* and *RGR*. Gene set enrichment analysis revealed *pigment granule* and *sensory perception of light stimulus* (FDR < 1x10⁻³) to be the top two pathways over-represented among fRPE-enriched genes. Comparison of transcriptomic profiles across glucose and galactose treatments identified 1924 differentially expressed genes (FDR < 1x10⁻²) primarily enriched in *cholesterol metabolism* (FDR < 1x10⁻⁴), a key pathway in AMD development. Genes associated with inherited retinal disorders and glaucoma had elevated expression in fRPE (P < 1.6x10⁻¹⁰), and such elevation was more pronounced in fRPE than in GTEx samples. Using stratified LD score regression, we found that variants near fRPE-enriched genes explain a larger fraction of risk than variants near GTEx tissue-enriched genes for both myopia and AMD. Together, these results suggest that fRPE can model both monogenic and polygenic ocular disorders better than cell types and tissues in public datasets. We next characterized the genetic regulation of gene expression and splicing. By combining total and allele-specific expression, we found 759 shared, 288 glucose-specific, and 201 galactose-specific eQTLs (FDR < 0.05), including known ocular disease genes such as *ABCA1* (glucose-specific) and *PRPF8* (galactose-specific). Comparison with GTEx identified three RPE-specific eQTLs (*TYR*, *CRX*, and *MFRP*). Fine-mapping of myopia and AMD GWAS loci with e/sQTL identified 8 candidate causal genes, 5 of which have not previously been linked to ocular diseases. Our study highlights the unique transcriptomic characteristics of fRPE compared to publicly available cell types and tissues and provides a novel resource to map the genetic regulation of expression and splicing in fRPE to monogenic and complex ocular disorders.

2512T

Multiple new genetic regulatory effects in blood discovered for age-related macular degeneration. A. V. Segre^{1,2,3}. 1) Ocular Genomics Institute, Massachusetts Eye and Ear, Boston, MA; 2) Department of Ophthalmology, Harvard Medical School, Boston, MA; 3) Broad Institute of Harvard and MIT, Cambridge, MA.

The majority of common variant associations found to date for Age-related Macular Degeneration (AMD), lie in noncoding regions, making it hard to decipher the causal mechanisms and genes in these regions. Furthermore, about half of the heritability of AMD has yet to be detected. Given the implication of the innate immune system in AMD pathogenesis, we tested whether variants associated with gene expression changes (expression quantitative trait loci or eQTLs) in whole blood or monocytes could help identify a proportion of the missing heritability of AMD and prioritize causal genes in known and new loci. We applied an extension of our eQTLEnrich method, that tests whether eQTLs in a given tissue are enriched for multiple modest (GWAS $p < 0.05$) to strong ($p < 5 \times 10^{-8}$) complex trait associations, correcting for various confounding factors, including eQTL tissue-specificity, to the largest publicly available GWAS meta-analysis of intermediate and advanced AMD (~23,000 cases, ~18,000 controls) and whole blood eQTLs (FDR < 5%) from the Genotype-Tissue Expression (GTEx) Project. We found significant enrichment of AMD associations amongst blood eQTLs ($P = 1 \times 10^{-5}$, Fold enrichment = 1.7), suggesting 140 new genetic associations with regulatory effects on AMD. To identify the true associations over noise, we applied our eGeneEnrich method to the target genes of eQTLs with AMD GWAS $p < 0.05$, and found multiple genes enriched in immune system processes, e.g., interferon signaling, the complement cascade, and MHC class I-mediated antigen presentation, as well as other biological processes, such as cell adhesion and lipid metabolism. Our analysis proposes new genes in the complement system with regulatory effects on AMD risk, whose genetic effects are smaller than previously reported associations for AMD in complement genes that harbor protein altering variants. Colocalization analysis helped prioritize high confidence causal genes for co-occurring AMD loci with eQTLs. By estimating the tissue-specificity of top ranked AMD-associated blood eQTLs compared to 47 non-blood GTEx tissues, and testing their effect on gene expression in normal and interferon-induced monocytes, we highlight regulatory effects that are likely to influence AMD risk via the immune system and inflammatory response. This work demonstrates that variation in transcriptional regulation substantially contributes to AMD risk, and proposes new target genes whose dysregulation in the blood may lead to AMD development.

2513F

Pervasive evidence of gene-environment interactions in genetic variants associated with refractive error. A. Pozarickij¹, J. Guggenheim¹, C. Williams². 1) Cardiff University, Cardiff, UK; 2) University of Bristol, Bristol, UK.

Refractive error is a complex trait that is estimated to affect 49.8% of the world population by 2050. It leads to an increased risk of developing myopic maculopathy, glaucoma and other ocular health problems. Susceptibility is determined both by genetic and environmental factors, but the evidence for their interplay is lacking. Here, we aimed to find out whether SNPs associated with refractive error are involved in gene-environment interaction (GxE). Our analysis included 72,985 individuals from UK Biobank cohort. To assess the pervasiveness of interactions, 149 genetic variants were selected. Conditional quantile regression models were fitted to detect changes in genetic variant effects across quantiles of the phenotype distribution. An unweighted polygenic risk score (PRS) was constructed and used to identify GxE with University attendance. We found that the majority (90%) of SNPs for refractive error showed evidence of non-constant effects. Typically, these variants had inverse-U shaped effect size profiles, with the strongest effects observed at the phenotype extremes, and effects closer to zero in emmetropes. Individuals who had a University degree had a more negative refractive error, on average, irrespective of their genetic susceptibility to myopia and their location in the refractive error distribution. The PRS x University attendance interaction also varied across percentiles, but with a different profile. This interaction effect was largest at the 30th percentile ($b_{\text{GxE}} = -0.38$ diopters, 95% CI -0.32 to -0.43, $p = 4.8 \times 10^{-41}$) and smallest at the 75th percentile ($b_{\text{GxE}} = +0.01$ diopters, 95% CI -0.02 to +0.03, $p = 0.68$). Thus, the PRS x University attendance interaction appeared to have negligible influence on the refractive error in some individuals (for example, those at the 75th percentile), and yet was associated with a difference in refractive error of more than 1.50 diopters for those with a -2 SD vs. +2 SD PRS located at the 30th percentile of the refractive error distribution. In contrast to current findings, our results indicate that GxE are common in refractive error development. The interaction effects typically have major impacts, such that effect sizes are often 2-fold greater for individuals in the phenotype extremes compared to those in the centre. Thus, the complex interplay between genetic and environmental factors uncovered in this work explains some of the missing heritability for refractive error.

2514W

Persistent developmental stuttering: Genetic factors and therapy

outcomes. C. Frigerio Domingues¹, Z. Gkalitsiou², A. Zezinka¹, E. Sainz¹, J. Gutierrez¹, C. Byrd³, R. Webster³, D. Drayna¹. 1) National Institute on Deafness and Other Communication Disorders, National Institutes of Health, Bethesda, MD - USA; 2) Lang Stuttering Institute and Moody College of Communication, University of Texas, Austin, TX - USA; 3) Hollins Communications Research Institute, Roanoke, VA - USA.

Stuttering is a common, genetically complex disorder characterized by involuntary disruptions in the flow of speech. The disorder has thus far been associated with variants in the *GNPTAB*, *GNPTG*, *NAGPA*, and *AP4E1* genes. These four genes are related to intracellular trafficking that is responsible for delivering endosomal cargoes to their correct location within the cell. We studied a set of patients with persistent developmental stuttering (50 mutation carriers and 47 non-mutation carriers) that all underwent a standardized intensive fluency shaping therapy program (IFSTP). Subjects' pre-therapy and post-therapy speech fluency records were analyzed to test whether therapy outcomes differed in those who carried a mutation in any of these known stuttering genes compared to individuals without such mutations. Speech fluency pre- and post-therapy was evaluated with observer-based quantitative stuttering dysfluency measures (percent words stuttered, PWS), and by subjects' self-reported measures of fluency, anxiety, and struggle behavior (perceptions of stuttering inventory, PSI). Both the subjects and speech pathologists who performed fluency ratings were blind to mutation status. The differences between pre-therapy and post-therapy fluency scores were taken as the measure of near-term therapy efficacy, and differences between groups were evaluated by t-tests. The IFSTP was effective for both groups, however, PSI scores showed significantly less improvement in individuals who carry a mutation (relative risk=1.75, odds ratio=2.92, $p=0.0058$). In addition, there was a trend toward poorer therapy outcomes in mutation carriers using PWS as scored by independent raters. We conclude that this fluency shaping therapy program produced significant improvements in speech fluency, however, mutation carriers achieved significantly less resolution of self-reported dysfluency, anxiety and struggle behavior following therapy, and that independent professional ratings of speech fluency displayed a trend toward poorer therapy outcomes in mutation carriers as well. Taken together, these results suggest that stuttering mutation carriers are more resistant to therapy compared to those without such mutations.

2515T

Three loci associated with risk of advanced age-related macular degeneration (AMD) also influence anti-VEGF treatment response.

O. Garcia Rodriguez¹, S.S. Pan¹, P. Whitehead¹, L.D. Adams¹, R.A. Laux², J.K. Welch¹, J.L. Kovach³, S.G. Schwartz³, A. Agarwal^{4,5}, J.L. Haines², M.A. Pericak-Vance¹, W.K. Scott¹. 1) John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL; 2) Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH; 3) Bascom Palmer Eye Institute, University of Miami Miller School of Medicine, Miami, FL; 4) Ophthalmology and Visual Sciences, Vanderbilt University Medical Center, Nashville, TN; 5) West Coast Retina, San Francisco, CA.

Response to anti-vascular endothelial growth factor(anti-VEGF) treatment for choroidal neovascularization(CNV) is variable. Identifying genetic factors that modulate treatment response could allow targeting of therapies to individuals most likely to respond positively. Fritsche et al. (2016) demonstrated that 52 common/rare single nucleotide polymorphisms(SNPs) in 34 loci account for 50% of the variation in the risk of developing AMD. Most genetic studies of treatment response have examined the association of these known genetic risk factors on treatment response, with few replicated results. We examined the association between 15,161 SNPs in these 34 loci and one-year change in visual acuity(VA) in 90 non-Hispanic white individuals with CNV(107 eyes) treated monthly for the first three months and as needed afterward. Samples were genotyped using the Illumina HumanCoreExome-24 v1.3 array and imputed against the 1000 Genomes Project Phase I reference panel. Association of one-year change in VA(measured as the change in logMAR) with SNP dosage was analyzed using linear mixed models, adjusting for age, sex, treatment, correlation of paired eyes, and population substructure (the top 3 principal components from EIGENSTRAT analysis). There were six variants in three AMD risk loci significantly associated with treatment response when correcting for the 34 loci tested($\alpha=1.5 \times 10^{-3}$). Each C allele at rs1951409($p=9.62 \times 10^{-4}$) and each A allele at rs199555630($p=6.73 \times 10^{-4}$) in *RAD51B* ($R^2=0.28$) were associated with worse VA one year after anti-VEGF treatment. Each T allele at rs117086110($p=5.21 \times 10^{-5}$) and each A allele at rs76904544($p=6.00 \times 10^{-5}$) in *RAD51B* ($R^2=0.94$) were associated with improved VA. There was no correlation between the SNPs in *RAD51B* positively and negatively associated with VA. Two other variants: rs2233739($p=3.51 \times 10^{-4}$) in *LIPC* and rs12519450 in *RAD51B* with VA. *PRLR/SPEF2*($p=1.28 \times 10^{-3}$) were associated with improved VA. Previous studies have not associated these six variants with anti-VEGF treatment response in CNV. *RAD51B* is involved in DNA repair and is associated with several cancers which may also be treated with anti-VEGF therapy. These results suggest that some biological mechanisms governing risk of AMD may also govern response to anti-VEGF treatment. These findings could allow anti-VEGF therapeutic regimens to be tailored to particular genetic profiles associated with positive or negative responses, improving outcomes and saving costs.

2516F

Identifying genes that underlie primary open-angle glaucoma using genetically determined gene expression. J. Hirbo¹, J. Bailey¹, P. Pawar², P. Evans¹, E. Gamazon¹, R. Tao¹, K. Joos¹, J. Wiggs², M. Brantley¹, N. Cox¹. 1) Vanderbilt University School of Medicine, Nashville, TN; 2) Harvard Medical School, Boston, MA, United States; 3) Case Western Reserve University School of Medicine, Cleveland, OH, United States.

Glaucoma is an eye disease characterized by progressive loss of the nerve connecting to the brain that leads to visual field loss, which if untreated will ultimately cause irreversible blindness. Glaucoma poses a large public health and economic burden that is estimated at nearly 5.8 billion annually. Increased intraocular pressure (IOP) is the most common risk factor for glaucoma, especially in Primary Open Angle Glaucoma (POAG). However, IOP-independent risk factors, such as vascular dysregulation, play an important part in the pathogenesis of glaucoma in some individuals who usually have a "normal" IOP termed normal-tension glaucoma. The only proven therapeutic approaches to treat POAG are for lowering IOP. One of the main limiting factors in developing effective therapies for POAG is lack of a clear understanding of the molecular mechanisms in disease etiology. There have been over twenty genome wide association studies (GWAS) on POAG, POAG risk factors and related endophenotypes that have cumulatively identified >200 potentially causal variants, but the biological mechanisms underlying these associations are not clear. All except one of these GWAS variants are noncoding, and might play a regulatory role. We hypothesized that the correlation of genetically determined component of gene expression with POAG can provide a powerful method to identify genes involved in the etiology of the POAG. We used two gene-based methods, PrediXcan and MetaXcan, that provide a framework for directly correlating imputed genetically determined gene expression with disease risk. We applied PrediXcan using models trained on 48 GTEx tissues to estimate individual level gene expression in a total of 5020 POAG cases and 36,330 controls of European descent. We applied MetaXcan on summary statistics from GWAS meta-analysis of IOP by Springelkamp et al., 2017. We determined the joint effects of gene expression variation imputed across the tissues using Multi-Tissue PrediXcan. We identified two novel genes, PGM5P2 and RAB38, that fall in regions with sub-genome wide significant SNPs in GWAS results. We also confirmed four genes, GAS7, TMCO1, SIX6 and CDKN2B that were previously suggested to underlie IOP and causal in Glaucoma because of being in vicinity to loci identified in GWAS. Our results points to novel pathophysiology that underlie POAG. If confirmed by experiment in model animals, these findings would offer novel gene targets for POAG therapies.

2517W

Gene-based study of ocular quantitative traits in 14,346 patients from Geisinger. N.A. Restrepo¹, C.R. Bauer¹, N.S. Josyula¹, E.D.K. Cha¹, S.A. Pendergrass¹. Biomedical and Translational Informatics, Geisinger, Rockville, MD.

Studies of vision impairment and blindness are rapidly accruing, elucidating how vision loss and blindness will burden economies and limit quality of life for affected patients. Genetics plays an integral role in ocular diseases such as age-related macular degeneration (AMD) and glaucoma which have been extensively validated, however the biological underpinnings are still vague and poorly characterized in certain conditions. Determining shared genetic risks is the first step towards investigating the genetic architecture of vision health. Gene- and pathway-based association analyses were performed to test the effects of rare variants in 1,385 autosomal genes extracted from the NHGRI-EBI GWAS catalog with 11 quantitative ocular traits. Targeted genes were selected to contain all genes associated ($p < 1 \times 10^{-5}$) with an ocular trait, ocular disease, or traits correlated with ocular health (i.e., HDL). Whole exome sequencing data was used from 44,020 European samples in Geisinger MyCode Community Health Initiative. The associations were performed on a subset of 14,346 patients who underwent an eye exam at Geisinger. A SKAT-based linear regression algorithm was implemented adjusting for age and sex. Biofilter and Biobin software packages were utilized to retrieve genomic locations and to perform association testing, respectively. Pathway analyses were conducted using GO Pathway and Reactome. We identified fifty-two significant gene associations ($p < 1 \times 10^{-5}$) all correlated to phenotypic traits related to corneal topography (K1, K2, and central corneal thickness). A number of these risk loci were previously found to contribute to the risk of genetically-correlated eye diseases including myopia, age-related nuclear cataracts, and corneal astigmatism. Of interest is the association of genes related to eye diseases not typically correlated with corneal topography traits, such as, AMD (i.e., *TXNB* and *COL4A1*) and glaucoma (i.e., *LMO7*). Pathway analysis (FDR $p < 0.05$) highlights an enrichment of top genes relevant to the structural complex of collagen trimers, collagen biosynthesis, assembly, and degradation. This study highlights the potential of gene discovery to elucidate shared genetic etiologies of eye diseases, previously not assumed to be related. This approach may identify cellular mechanisms involved in AMD and glaucoma risk. Future studies will explore interactions between genes and biological functions of ocular traits and their impact on ocular health. .

2518T

PMEL missense variants are associated with pigment dispersion syndrome and pigmentary glaucoma. M.A. Walter¹, A.A. Lahola-Chomiak¹, T. Footz¹, K. Nguyen-Phuoc¹, G.J. Neil², B.J. Fan³, K.F. Allen³, D.S. Greenfield⁴, R.K. Parrish⁵, K. Linkroum³, L.R. Pasquale³, R.M. Leonhardt⁶, R. Ritch⁷, J.E. Craig⁸, W.T. Allsion^{1,2}, O.J. Lehmann^{1,9}, J.L. Wiggs³. 1) Medical Genetics, University of Alberta, Edmonton, Alberta, Canada; 2) Department of Biological Sciences, University of Alberta, Edmonton AB, Canada; 3) Ocular Genomics Institute and Department of Ophthalmology, Massachusetts Eye and Ear Infirmary, Harvard Medical School, Boston, MA; 4) Department of Ophthalmology, Bascom Palmer Eye Institute, University of Miami Miller School of Medicine, Palm Beach Gardens, FL; 5) Anne Bates Leach Eye Hospital, Bascom Palmer Eye Institute, University of Miami Miller School of Medicine, Miami, FL; 6) Department of Immunobiology, Yale University School of Medicine, New Haven, CT; 7) Einhorn Clinical Research Center, New York Eye and Ear Infirmary of Mount Sinai, New York, NY; 8) Department of Ophthalmology, Flinders Medical Centre, Adelaide, South Australia; 9) Department of Ophthalmology, University of Alberta, Edmonton AB, Canada.

Pigmentary Glaucoma (PG) is a common glaucoma subtype that results from release of pigment from the iris, called Pigment Dispersion Syndrome (PDS), and its deposition throughout the anterior chamber of the eye. Although PG has a substantial heritable component, no causative genes have yet been identified. We used Whole Exome Sequencing (WES) of two independent pedigrees to identify 2 *PMEL* variants associated with heritable PDG/PG. *PMEL* (premelanosome protein) encodes a key component of the organelle essential for melanin synthesis, storage, and transport (melanosome). Targeted screening of *PMEL* in three independent cohorts (n= 394) identified seven additional PDS/PG-associated non-synonymous variants. Five of nine variants exhibited defective processing of the *PMEL* protein. In addition, analysis of PDS/PG-associated *PMEL* variants expressed in HeLa cells revealed structural changes to the melanosome indicating altered amyloid fibril formation in 5 of 9 variants. Knockdown of the homologous *pmel* in zebrafish caused profound pigmentation defects in the eye further supporting *PMEL*'s role in ocular pigmentation and function. Taken together, these data support a model in which missense *PMEL* variants represent dominant negative mutations that impair the ability of *PMEL* to form functional amyloid fibrils. While *PMEL* mutations have previously been shown to cause pigmentation and ocular defects in animals, this research is the first report of mutations in *PMEL* causing human eye disease.

2519F

Do carriers of "high penetrance" myocilin variants always get glaucoma? P. Gharahkhani¹, X. Han¹, E. Souzeau², J.S. Ong¹, J. An¹, O.M. Siggs³, K.P. Burdon^{2,3}, D.A. Mackey^{4,4}, A.W. Hewitt^{5,5}, J.E. Craig², S. MacGregor¹. 1) QIMR Berghofer Medical Research Institute, Brisbane, Australia; 2) Department of Ophthalmology, Flinders University, Flinders Medical Centre, Adelaide, Australia; 3) Menzies Institute for Medical Research, University of Tasmania, Hobart, Australia; 4) Centre for Ophthalmology and Visual Sciences, University of Western Australia, Lions Eye Institute, Perth, Australia; 5) Centre for Eye Research Australia, Royal Victorian Eye and Ear Hospital, University of Melbourne, Melbourne, Australia.

Glaucoma is the leading cause of irreversible blindness worldwide. Current treatments for glaucoma are generally effective in preventing disease progression if applied early. Given glaucoma has a strong genetic basis, gene-based prediction of risk is likely to assist with early diagnosis. The p.Gln368Ter variant in *MYOC* is a well-established variant for glaucoma with carriers reported to have a high risk of glaucoma. However, the clinical utility of screening this variant remains controversial because although the variant was initially reported to have a very high penetrance (70-100%) in family-based studies, recent population-based studies found much lower estimates (10-20%). We used 411337 white-British ancestry participants including 7997 glaucoma cases within UK Biobank (UKB) to examine penetrance and effect size of *MYOC* p.Gln368Ter for glaucoma and ocular hypertension (OHT). We used OHT as a proxy for glaucoma since a proportion of glaucoma cases remain undiagnosed due to the lack of comprehensive eye examination. We then compared these results with those from two Australian registry-based studies comprising 3071 glaucoma cases and 6750 historic controls. In UKB, the ORs (95% confidence interval, CI) of p.Gln368Ter for primary open angle glaucoma (POAG), glaucoma (including POAG, self-report, and unspecified glaucoma), OHT, and OHT or glaucoma were 6.60 (3.86, 10.48), 4.43 (3.49, 5.55), 3.65 (2.62, 4.99), and 4.28 (3.15, 5.75), respectively. The penetrance of p.Gln368Ter in glaucoma, OHT, and OHT or glaucoma was 7.55%, 24.3% and 30.84%, respectively. In registry-based studies, the OR (95% CI) of p.Gln368Ter for advanced glaucoma was 12.16 (6.34, 24.97) while the risk was lower for non-advanced glaucoma: OR=3.97 (1.55, 9.75). The penetrance was 56.10% and 69.51% for glaucoma and glaucoma/suspects, respectively. For age 65 and above, ~50% of *MYOC* p.Gln368Ter carriers had glaucoma or OHT in UKB and ~83% in registry-based studies. These results show that although the penetrance of p.Gln368Ter in general population is not as high as it has been reported in registry-based studies, it is much higher than those reported in previous population-based studies, probably due to ascertainment bias towards healthy individuals in those studies. Overall, our results supports clinical utility of genetic testing for p.Gln368Ter to help identify individuals who are at greater risk of developing glaucoma and direct them to early screening and appropriate management.

2520W

Dissecting corneal layers from keratoconus cornea towards differential gene expression analysis. M. Gajęcka^{1,2}, M. Udziela³, R. Malinowski⁴, M. Olszak⁴, J.P. Szaflik³, K. Jaskiewicz^{1,2}. 1) Poznan University of Medical Sciences, Poznan, Poland; 2) Institute of Human Genetics Polish Academy of Sciences, Poznan, Poland; 3) Department of Ophthalmology, Medical University of Warsaw, Warsaw, Poland; 4) Institute of Plant Genetics, Polish Academy of Sciences, Poznan, Poland.

Background: Keratoconus (KTCN) is a multifactorial eye disorder leading to a loss of vision acuity. The human cornea comprises of 5 layers, including three cellular: epithelium, stroma and endothelium. Precise pathogenesis of the disease, including significance of the individual corneal layers, remains unclear. Analyses conducted on the full-thickness corneas provide results from the heterogeneous mixture of cells, masking differential gene expression and phenotype correlations. **Aim:** The protocol selection and optimization for obtaining homogeneous cell populations from specific corneal layers, towards differential gene expression analysis. **Material and Methods:** The fresh-frozen cornea from a KTCN-patient, after a penetrating keratoplasty procedure, is cut on a cryotome. During a laser microdissection (LM) procedure the corneal epithelium, stroma and endothelium are dissected. After RNA extraction (Single Cell RNA Purification Kit, Norgen Biotek, Canada) assessment of RNA quality (RIN) and quantity is performed using Agilent RNA 6000 Pico Kit. RT-qPCR reactions with multiple primers are executed. **Results:** We examined the LM method for our purposes. Other considered technics (fluorescence activated cell sorting, FACS, microfluidics) require a dissociation step, during which the tissue architecture (extracellular matrix, cell-cell junctions) is disrupted. Collagenase treatment would be necessary for the connective tissue such as cornea, what significantly changes the gene expression. We successfully optimised the LM protocol, embracing tissue fixation, staining and setting of laser's parameters, visually documenting all steps. RNA quality and quantity was satisfactory towards RT-qPCR reactions. Surprisingly, the *Actinβ* (as a reference gene), chosen in the first place, was not evenly expressed in all analysed layers. Future research plans include differential expression analyses based on RNA-seq experiments. **Conclusions:** We assume that it is crucial to assess differential expression in the separate corneal layers. We consider that both, downstream and high-throughput studies using dissected corneal layers will contribute to the increase of knowledge about the KTCN pathogenesis.

2521T

Association between additive effects of genetic variants associated with primary open-angle glaucoma and age at diagnosis of glaucoma. F. Mabuchi¹, N. Mabuchi¹, Y. Sakurada¹, S. Yoneyama¹, K. Kashiwagi¹, H. Iijima¹, Z. Yamagata², M. Takamoto³, M. Aihara⁴, T. Iwata⁵, K. Kawase⁶, Y. Shiga⁷, K. Nishiguchi⁸, T. Nakazawa^{7,8,9}, M. Ozaki¹⁰, M. Araie¹¹ for the Japan Glaucoma Society Omics Group (JGS-OG). 1) Dept Ophthalmology, Univ Yamanashi, Chuo, Yamanashi, Japan; 2) Dept Health Sciences, Univ Yamanashi, Chuo, Yamanashi, Japan; 3) Dept Ophthalmology, Saitama Red Cross Hospital, Saitama, Saitama, Japan; 4) Dept Ophthalmology, Graduate School of Medicine, Univ Tokyo, Bunkyo-ku, Tokyo, Japan; 5) Division of Molecular and Cellular Biology, National Institute of Sensory Organs, National Hospital Organization Tokyo Medical Center, Meguro-ku, Tokyo, Japan; 6) Gifu University Hospital, Gifu, Gifu, Japan; 7) Dept Ophthalmology, Tohoku Univ Graduate School of Medicine, Sendai, Miyagi, Japan; 8) Dept Advanced Ophthalmic Medicine, Tohoku Univ Graduate School of Medicine, Sendai, Miyagi, Japan; 9) Dept of Retinal Disease Control, Tohoku Univ Graduate School of Medicine, Sendai, Miyagi, Japan; 10) Ozaki Eye Hospital, Hyuga, Miyazaki, Japan; 11) Kanto Central Hospital of the Mutual Aid Association of Public School Teachers, Setagaya-ku, Tokyo, Japan.

Purpose: Genetic variants associated with primary open-angle glaucoma (POAG) can be classified into two types. One is a genetic variant associated with intraocular pressure (IOP) elevation (IOP-related genetic variant), while the other is a genetic variant associated with a vulnerability of optic nerve independent of IOP (non-IOP-related genetic variant). This study was conducted to investigate the association between the additive effects of IOP-related or non-IOP-related genetic variants and age at diagnosis of glaucoma as an indicator of POAG progression. **Methods:** Eleven variants identified as IOP-related genetic variants, including rs1052990 (near gene: *CAV2*), rs11656696 (*GAS7*), rs59072263 (*GLCCI1/ICA1*), rs2472493 (*ABCA1*), rs58073046 (*ARHGEF12*), rs2286885 (*FAM125B*), rs6445055 (*FNDC3B*), rs8176743 (*ABO*), rs747782 (*PTPRJ*), rs4619890 (*AFAP1*), and rs2745572 (*FOXC1*), and 9 variants considered as non-IOP-related genetic variants, including rs3213787 (*SRBD1*), rs735860 (*ELOVL5*), rs1063192 (*CDKN2B*), rs1900004 (*ATOH7*), rs10483727 (*SIX6*), rs11969985 (*GMD5*), rs3785176 (*PMM2*), rs1192415 (*TGFBR3*), and rs35934224 (*TXNRD2*), were genotyped for 505 patients with POAG and 246 control subjects. The total number of risk alleles of the 11 IOP-related or 9 non-IOP-related genetic variants was calculated for each participant as genetic risk scores (GRSs). A logistic regression model was used to estimate the risk (odds ratio) of glaucoma for each risk allele of genetic variants, and the sum of the logarithmically converted odds ratios of 11 IOP-related or 9 non-IOP-related genetic variants was used as weighted GRSs (WGRSs). The association between these GRSs or WGRSs and age at diagnosis of glaucoma was evaluated. **Results:** There were significant relationships between the GRS (Beta = -0.16, P = 0.0005) or WGRS (Beta = -0.13, P = 0.0039) of 11 IOP-related genetic variants and the age at diagnosis of glaucoma by multiple linear regression analysis adjusted for gender. As the GRS or WGRS of IOP-related genetic variants increased, the age at diagnosis of glaucoma became younger. **Conclusions:** Additive effect of IOP-related, not non-IOP-related, genetic variants was associated with age at diagnosis of glaucoma in POAG, which indicates that IOP-related genetic variants (IOP elevation) rather than non-IOP-related genetic variants (vulnerability of optic nerve) may play an important role for POAG progression.

2522F

Opsin gene probes accurately predict red-green color vision deficiencies derived from a web-based Ishihara plate color perception test. *M.L. Multhaup, M.N. Johnson, J.F. Shelton, M.J. Strand, B.L. Koelsch, B. Alipanahi.* 23andMe, Mountain View, CA.

Red-green colorblindness is an umbrella term for several X-chromosome-linked color vision deficiencies highly prevalent among men. These conditions encompass a spectrum of severity, and many men with less severe disorders are unaware that they have color vision deficiencies at all. The vast majority of red-green colorblindness in humans is due to variation in the X-chromosome medium and long wave opsin genes, *OPN1MW* and *OPN1LW*. These genes are derived from a relatively recent gene duplication and share 98% similarity by nucleotide sequence. The nucleotides that are divergent between these genes are disrupted in red-green colorblindness, making the creation of colorblindness-informative gene-specific probes difficult and rendering this region uncovered by most standard genotyping arrays. 23andMe has created custom genotyping probes to detect variants in these two genes. In parallel, we have also collected self-reported colorblindness status from over 350,000 male research participants and developed a custom 8-plate Ishihara color perception test completed by more than 40,000 male research participants. We developed a scoring system for our Ishihara test that indicates both full and partial colorblindness and is highly correlated with self-reported colorblindness. We used raw probe intensity data to fit gradient tree boosting models predicting full and partial colorblindness. These models had area under receiver-operator curves of greater than 0.95 and 0.87, respectively, positive predictive values of 60% and negative predictive values of 99% in out-of-sample testing cohorts. Using raw probe intensity for prediction resulted in substantially higher prediction accuracy than either default genotype calls or a custom-calling scheme designed individually for each probe. We hope to use these models to identify 23andMe customers who may not know that they are partially red-green colorblind and should undergo further testing in order to facilitate positive quality of life interventions..

2523W

The role of regulatory variants in *FZD6*-related gene network in nonsyndromic cleft lip and palate. *L. Mailli¹, Q. Yuan², B.T. Chiquet^{2,3}, A. Letra³, G.T. Eisenhoffer⁴, J.T. Hecht^{1,2,3}*. 1) MDAnderson UTHealth Graduate School of Biomedical Sciences, Houston, TX; 2) McGovern Medical School, Houston, TX; 3) University of Texas School of Dentistry, Houston, TX; 4) University of Texas MD Anderson Cancer Center, Houston, TX.

Nonsyndromic cleft lip with or without cleft palate (NSCLP) is a common birth defect with a prevalence of 1 in 700 that annually affects approximately 4000 newborns in the United States. Despite improvement in treatments, NSCLP continues to pose major medical, psychosocial, and financial burdens. We, and others, have shown that perturbation of *FZD6* and related genes causes severe craniofacial anomalies: 1) knockdown or overexpression of *fzd6* causes craniofacial abnormalities in zebrafish, 2) loss of *lrp6* significantly reduces Wnt signaling in mutant mouse embryos causing orofacial clefts, 3) knockdown or mutant forms of *lrp5* cause abnormal craniofacial cartilage phenotypes and cleft palate in zebrafish, and 4) the facial region completely fails to form in mice deficient for *dkk1*. These genes act at the receptor level of the β -catenin mediated Wnt pathway. The goal of this study was to identify variation in regulatory elements upstream of *FZD6*, *LRP5*, *LRP6* and *DKK1* and their contributions to NSCLP. Twenty-one variants in these genes were assessed for allele-specific binding and 10 variants, 1 in *FZD6*, 3 in *LRP6*, 3 in *LRP5* and 3 in *DKK1*, showed differential DNA-protein binding patterns. To assess their effect on gene expression, these variants and their flanking regions were cloned upstream of luciferase promoter constructs. Five variants, rs115369160, rs7139600 and rs7136380 in *LRP6* and rs7069912 and rs114205486 in *DKK1*, had significant allele-specific effects on luciferase assays in 293T cells and are being assessed in other cell types to evaluate tissue specificity. Utilizing a stable zebrafish reporter line, β -catenin was evaluated in craniofacial development from 1-7 days post fertilization (dpf) and was highly active in the oral cavity and other facial regions between 3-5 dpf. Morpholino knockdown of *fzd6* in these reporter fish resulted in a reduction in β -catenin activity in the oral cavity at 3 dpf. *Lrp5*, *lpr6* and *dkk1b* are now being similarly assessed. The results of this study provide important information about the effects of regulatory variants in *FZD6* and related β -catenin/Wnt pathway genes and their role in craniofacial development. This sets the stage for subsequent evaluation of these noncoding variants in NSCLP families.

2524T

Minimizing facial dysmorphologies associated with Down syndrome by inhibition of DYRK1A related pathways. N. Martínez-Abadías¹, J.M. Starbuck², R. González³, X. Sevillano³, A. González³, J. Sharpe⁴, R. De La Torre⁴, M. Dierssen^{5,6}. 1) Multicellular Systems Biology, European Molecular Biology Lab (EMBL), Barcelona, Spain; 2) Department of Anthropology, University of Central Florida, Orlando, FL, USA; 3) Grup de Recerca en Tecnologies Mèdia (GTM), La Salle - Universitat Ramon Llull, Barcelona, Spain; 4) Integrative Pharmacology and Systems Neuroscience, IMIM-Hospital del Mar Medical Research Institute, Barcelona, Spain; 5) Cellular & Systems Neurobiology, Center for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Barcelona, Spain; 6) Universitat Pompeu Fabra (UPF), Barcelona, Spain.

Down syndrome affects 1 out of 700 live born individuals and is the most prevalent genetic cause of intellectual disability. It is caused by trisomy of chromosome 21, which leads to a complex pleiotropic phenotype that involves a recognizable set of facial traits. Recent studies have shown that the overall genetic imbalance caused by the 1.5-fold gene dosage increase associated with trisomy 21 can be modulated through inhibition of kinase activity by the green tea flavanol epigallocatechin-3-gallate (EGCG). Although all genes and signaling pathways are yet to be discovered, available evidence indicates that EGCG inhibits several DYRK1A related pathways, such as the *NFAT* (Nuclear factor of activated T-cells) signaling pathway and transcription factors *DCAF741* and *GLI142*, which are overexpressed in Down syndrome and play a critical role as regulators of brain, skeletal and facial development. Here we provide for the first time clinical evidence of the potential benefits of EGCG treatment to facial morphology in Down syndrome. We analyzed a cross-sectional sample of 287 children from 0 to 18 years old including euploid children, children with either non-disjunction or mosaic Down syndrome, as well as children with Down syndrome who had taken EGCG as dietary supplements. We assessed global and local facial shape differences between children with Down syndrome and euploid controls analyzing the 3D coordinates of 21 anatomical landmarks identified on 3D facial reconstructions using Geometric Morphometric methods. Our observational study revealed that EGCG treatment since early in development is associated with intermediate facial phenotypes and significant facial improvement scores. If taken during the three first years of life, EGCG supplementation reduced by 47% the number of facial differences of trisomic and euploid individuals, by normalizing length, width, and depth dimensions of the midface, with strongest effects on the philtrum and nasal regions. Our findings thus suggest a potential beneficial effect of EGCG on facial development in Down syndrome. In a broader context, our phenomic approach could be applied to other disorders associated with facial dysmorphologies to correlate facial phenotypes with different levels of genetic expression and to test the efficacy of potential therapeutic treatments.

2525F

Insights into the etiology of nonsyndromic orofacial clefting by WGS of Colombian trios. E.J. Leslie¹, L. Moreno-Urbe², P. Chopra¹, D.J. Cutler¹, E. Feingold³, C. Valencia-Ramírez⁴, C. Restrepo⁴, A.C. Lidral⁵, M. Arcos-Burgos⁶, E. Liao⁷, R.A. Cornell², M.L. Marazita². 1) Emory University, Atlanta, GA; 2) University of Iowa, Iowa City, IA; 3) University of Pittsburgh, Pittsburgh, PA; 4) Fundación Clínica Noel, Medellín, Colombia; 5) Private Practice, MI; 6) Australian National University, Canberra, Australia; 7) Massachusetts General Hospital; Harvard Medical School, Boston, MA.

Orofacial clefts (OFCs) represent the largest group of craniofacial malformations in humans with a worldwide prevalence of 1 in 1000 live births. OFCs, including cleft lip and palate (CLP), are regarded as complex traits with a strong genetic component influencing risk. Genetic studies including linkage, GWAS, and sequencing have identified a number of genetic risk factors, but have not been able to define a single genetic mechanism underlying this disorder. Because all types of variants (common, rare, de novo, etc.) appear to confer risk, the genetic etiology of OFCs remains difficult to fully elucidate. To comprehensively examine genetic risk factors for OFCs, we performed whole genome sequencing (WGS) on 276 case-parent trios with CLP from Colombia as part of the Gabriella Miller Kids First Pediatric Research Program. Genomic DNA was extracted from peripheral blood cells; paired end sequencing was performed at >30X coverage on the Illumina HiSeq X instrument and variant calling was performed using GATK. Our first analyses focused on de novo mutations (DNMs) which, despite multiple lines of evidence supporting them as causes of OFCs, have been poorly represented in sequencing studies. We identified 251 exonic DNMs across all genes; four genes (*CTNND1*, *KLB*, *HLA-DRB5*, and *MUC2*) had multiple DNMs, resulting a genome-wide statistical significance. To further explore the genes with only one DNM, we focused on genes belonging to the IRF6 regulatory network because the IRF6 transcription factor has been demonstrated to be a chief determinant of both syndromic and nonsyndromic forms of OFCs. We identified 238 putative IRF6 target genes by performing ChIP-seq for Irf6 in zebrafish embryos and cross-referencing with RNA-seq in wildtype and *irf6*^{-/-} zebrafish and mouse craniofacial gene expression databases. In the WGS data, we identified one de novo mutation in *IRF6* (p.Asn88Asp) and 7 additional nonsynonymous or loss-of-function mutations in IRF6 target genes (representing 3% of all DNMs). These included mutations in *CDH2* (p.Arg656Trp), *GRHL3* (p.Pro480Ser), and *TFAP2A* (p.Glu104*). Together, these analyses support growing evidence for de novo mutations in the etiology of OFCs. Further, it demonstrates the important role of IRF6 and provides rationale for further explorations of rare inherited variants and noncoding regulatory variants in and around these genes, and expansion of this approach to other OFC risk loci.

2526W

Structural variation in neural tube defects: Adding another layer to a complex genetic architecture. *P. Wolujewicz^{1,2}, V. Aguiar-Pulido¹, R.H. Finnell³, C.E. Mason^{1,2,4}, O. Elemento^{1,2,4,5}, M.E. Ross¹.*

1) Center for Neurogenetics, Weill Cornell Medicine, New York, NY; 2) Department of Physiology and Biophysics, Weill Cornell Medicine, New York, NY; 3) Center for Precision Environmental Health, Baylor College of Medicine, Houston, TX; 4) Institute for Precision Medicine, Weill Cornell Medicine, New York, NY; 5) Institute for Computational Biomedicine, Weill Cornell Medicine, New York, NY.

Neural tube defects (NTDs), especially spina bifida (SB), are common debilitating birth defects that involve multiple genes and environmental factors. While genetic heritability estimates for NTDs range from 60-70%, the genetic architecture has proven to be complex and mutations in more than 400 genes can produce NTDs in mice. Attempts to identify human mutations that confer risk for NTDs remains a challenge. While several studies have investigated the inherent genomic variation in NTDs (e.g. exploring SNVs from exomes as well as CNVs from array data), the interrogation of all types of structural variants remains largely unexplored and no study to date has analyzed structural variation in NTD patients via whole genome sequencing (WGS) approaches. Here, we conducted WGS on over 150 SB patients and over 150 population-matched controls to elucidate the contribution of structural variants in order to better understand the complex genetic nature of this malformation. To this end, we have developed an integrative approach for germline structural variant detection and annotation as well as a visualization framework to aid in variant validation and interpretation. Using our approach, we report on the burden of structural variants in NTD cases and genes that are enriched with this form of variation that are likely to be associated with the condition.

2527T

Utility of early tongue reduction surgery in patients with Beckwith-Wiedemann syndrome. *J.L. Cohen^{1,2}, B. Sajorda¹, E. Traxler¹, W. Frederic¹, K. Duffy^{1,2}, E. Hathaway^{1,2}, J.A. Taylor^{1,2}, C. Cielo^{1,2}, J.M. Kalish^{1,2}.*

1) The Children's Hospital of Philadelphia, Philadelphia, PA; 2) The Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA.

Macroglossia is a cardinal feature of Beckwith-Wiedemann syndrome (BWS) and has been associated with obstructive sleep apnea (OSA), speech and/or feeding difficulties, and may be treated with tongue reduction (TR). There are no guidelines advising standard technique or timing for surgery. Some experts recommend waiting until age 6-12 months to avoid possible tongue tissue regrowth. We sought to determine who would benefit from early tongue reduction (ETR), defined as less than 12 months old, or very early tongue reduction (VETR), defined as less than 6 months old. We hypothesized that ETR and VETR would benefit BWS patients who presented early in life with clinical indications for surgery. We conducted a retrospective review of the BWS registry and determined 28 patients underwent TR at any age during the past 4 years at our institution. We determined the BWS clinical scores and reviewed molecular testing on available tongue tissue. We examined polysomnography reports and clinical feeding, pulmonary, and surgery evaluations. Of the 28 TR subjects, 8 had VETR, 7 had ETR. Tongue methylation analysis showed concordance with the patients' blood analysis. The average BWS clinical score of subjects with VETR (11.3+/- 0.7) was greater than those who had ETR (8.6+/- 0.6) ($p < 0.05$). Indications for performing VETR included OSA (8/8), often accompanied by other respiratory support needs (5/8), or feeding difficulties (5/8). Of the 15 patients within the VETR and ETR groups, at least 3 were prenatally diagnosed with macroglossia. In subjects for whom polysomnography was performed, median obstructive apnea hypopnea index was 30.8 (VETR) and 16.8 (ETR) preoperatively. Six of the eight subjects that underwent VETR experienced resolution of OSA based on post-operative polysomnography or clinical assessment. The earliest TR case, at 7 days old, was diagnosed prenatally and demonstrated improved speech, feeding, and sleep outcomes prior to NICU discharge and at long-term 2-year follow-up. Based on this data, if a patient with BWS is prenatally diagnosed with macroglossia, or requires respiratory support in the neonatal period, we recommend evaluation for OSA followed by consideration of VETR. The data supports that VETR and ETR are safe and effective in treating the indications for surgery.

2528F

A population-based study of gene-alcohol interactions and orofacial cleft risk. A.L. Petrin¹, W. Lyu², L. DeRoo³, P.A. Romitti⁴, M.M. Jenkins⁵, R.G. Mungers⁶, J.C. Murray⁷, A.J. Wilcox⁸, R.T. Lie⁹, G.L. Wehby¹⁰, L.M. Moreno Uribe^{1,9}, *National Birth Defects Prevention Study*. 1) Dows Institute, College of Dentistry, the University of Iowa, Iowa City, IA, USA; 2) Department of Health Management and Policy, University of Iowa, Iowa City, Iowa, USA; 3) Department of Global Public Health and Primary Care, University of Bergen, Bergen, Norway; 4) Department of Epidemiology, College of Public Health, The University of Iowa, Iowa City, Iowa, USA; 5) Centers for Disease Control and Prevention, National Center on Birth Defects and Developmental Disabilities, Atlanta, Georgia, USA; 6) Department of Nutrition and Food Sciences, Utah State University, Logan, Utah, USA; 7) Department of Pediatrics, Carver College of Medicine, University of Iowa, Iowa City, Iowa, USA; 8) Epidemiology Branch, National Institute of Environmental Health Sciences/National Institutes of Health, Durham, North Carolina, USA; 9) Department of Orthodontics, College of Dentistry, the University of Iowa, Iowa City, IA, USA.

Prior studies have shown conflicting results for maternal prenatal alcohol consumption and the risk for orofacial clefts (OFCs), possibly hiding significant heterogeneity in risk by maternal and fetal alcohol metabolism genes. We examined interaction effects on isolated OFC risk between first-trimester maternal alcohol use (any alcohol and average drinks/sitting) and 7 alcohol metabolism genes/loci (ADH1B/ADH1C, ADH1C, ALDH1A1, CTBP2, MLLT3, OPRL1, SMC2) in both mothers and children. We combined individual-level data from 5 population-based studies—Iowa Case–Control Study, National Birth Defects Prevention Study, Norway Facial Clefts Study, Norwegian Mother and Child Cohort, and Utah Child and Family Health Study—to create the largest population-based case-control sample to date, consisting of 1,330 (1,743) case children (mothers) and 2,549 (3,417) control children (mothers). We evaluated interactions between each alcohol consumption measure and gene/locus using multivariable logistic regression, including fixed effects for study site, maternal age and education at delivery, pre-pregnancy body mass index, and first-trimester smoking and folic acid/multivitamin use. We observed significant interaction effects between any maternal first-trimester alcohol consumption and maternal ADH1B/ADH1C ($p < 0.05$) and ALDH1A1 ($p < 0.01$) genotypes. Cleft lip only (CLO) risk doubled for mothers homozygous for the minor allele of ADH1B/ADH1C rs1789891 (adjusted odds ratio (aOR)=2.36, 95% confidence interval (CI)=1.16–4.81), while cleft palate only (CPO) risk increased (aOR=1.54, 95% CI=1.06–2.23) for mothers homozygous for the major allele of ALDH1A1 rs2303317. Similar interactions were observed when mothers reported < 5 drinks per sitting. CLO risk was significantly higher when mothers reported repeated ≥ 5 drinks per sitting with the minor allele of maternal MLLT3 rs4621895 (aOR=3.86, 95% CI=1.65–9.12) and fetal OPRL1 rs6010718 (aOR=24.05, 95% CI=1.95–298.87); CPO risk also increased with the same fetal variant (aOR=11.47, 95% CI=1.16–112.17). These findings suggest meaningful heterogeneity in the risk of OFC subtypes associated with maternal alcohol consumption by variants in certain alcohol metabolism genes.

2529W

WES and large-scale resequencing in cleft lip with/without cleft palate patients provides further evidence for a causative role of CDH1 and TRMO. N. Ishorst^{1,2}, L. Henschel^{1,2}, F. Thieme^{1,2}, D. Driche³, S. Sivalingam^{1,2}, S.L. Mehrem^{1,2}, A.C. Fechtner^{1,2}, J. Engel^{1,2}, A. Heimbach^{1,2}, A. Hoischen^{4,5,6}, A. Rojas-Martinez⁷, K. Aldhorae⁸, B. Braumann⁹, M. Martini¹⁰, H. Reutter¹¹, S. Nowak¹, L. Götz^{12,13}, M. Knapp¹⁴, M.M. Nöthen^{1,2}, M. Nothnagel⁵, T. Becker¹⁵, K.U. Ludwig^{1,2}, E. Mangold¹. 1) Institute of Human Genetics, University of Bonn, School of Medicine & University Hospital Bonn, Bonn, Germany; 2) Department of Genomics, Life & Brain Center, University of Bonn, Bonn, Germany; 3) Cologne Center for Genomics, University of Cologne, Cologne, Germany; 4) Department of Human Genetics, Radboud University Medical Center, Nijmegen, The Netherlands; 5) Department of Internal Medicine, Radboud University Medical Center, Nijmegen, The Netherlands; 6) Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, Nijmegen, The Netherlands; 7) Tecnológico de Monterrey, School of Medicine, and Universidad Autónoma de Nuevo León, Centro de Investigación y Desarrollo en Ciencias de la Salud, Monterrey, Mexico; 8) Orthodontic Department, College of Dentistry, Thamar University, Thamar, Yemen; 9) Department of Orthodontics, University of Cologne, Cologne, Germany; 10) Department of Oral and Maxillo-Facial-Plastic Surgery, University of Bonn, Bonn, Germany; 11) Department of Neonatology, Children's Hospital, University of Bonn, Bonn, Germany; 12) Department of Orthodontics, University of Erlangen, Erlangen, Germany; 13) Department of Orthodontics, University of Bonn, Bonn, Germany; 14) Institute of Medical Biometry, Informatics and Epidemiology, University of Bonn, Bonn, Germany; 15) Institute for Community Medicine, University of Greifswald, Greifswald, Germany.

Non-syndromic cleft lip with or without cleft palate (nsCL/P) is a common congenital malformation and has a multifactorial etiology. The common risk variants, mostly identified by GWAS, explain about 25% of the genetic heritability. We hypothesize that a fraction of the remaining genetic liability is explained by rare dominant *de novo* mutations. In order to identify such rare *de novo* mutations, we performed whole exome sequencing (WES) in 50 trios, consisting of nsCL/P patients and their unaffected parents. This resulted in identification of 33 *de novo* mutations in 33 genes. Among these were *de novo* mutations in *CDH1* and *TRMO*, both already known to be associated with nsCL/P. To find further support for a contribution of these candidate genes to nsCL/P, and to strengthen our hypothesis of dominant *de novo* events adding to nsCL/P etiology, the candidate genes were subjected to a large-scale resequencing study with single-molecule molecular inversion probes (smMIPs). The independent resequencing cohort comprised nsCL/P patients and population-matched controls of Arabian, Mexican and European ancestry ($n_{\text{cases}}=1,013$, $n_{\text{controls}}=1,591$). The assay was designed using the MIPGEN pipeline. smMIPs for 32 genes were successfully designed. Libraries were sequenced on HiSeq2500 with 2x125bp read length. Raw reads were aligned with BWA and variants were called with UnifiedGenotyper. Downstream analysis included filtering for CADD ≥ 15 and MAF $\leq 0.1\%$. Only those variants not present in any of the controls were used for further analysis which included manual inspection of reads and segregation analysis. As a first step of our data analysis, we checked the genes *CDH1* and *TRMO* for deleterious variants. We identified 1 rare stop gain and 8 rare missense variants in *CDH1*, and 1 rare stop gain and 3 rare missense variants in *TRMO*. Notably, the average CADD score of these variants was > 25 . Of particular interest is the stop gain in *CDH1* (NM_004360:p.R598X) which leads to a loss of the cytosolic binding domains of the CDH1 binding partners CTNND1 and CTNNB1. This protein complex is part of the epithelial adhesion complex which was recently shown to lead to nsCL/P when disrupted in a mouse model. First results of our resequencing approach found further indications for a contribution of rare variants in *CDH1* and *TRMO* to nsCL/P development. Further analyses of our data are on the way, aiming at identification of rare *de novo*/co-segregating variants in novel candidate genes for nsCL/P.

2530T

Integrated clinical and omics approach to rare diseases: Novel genes and oligogenic inheritance in holoprosencephaly. A. Kim¹, C. Savary¹, C. Dubourg², W. Carré³, C. Mouden¹, H. Hamdi-Rozé¹, H. Guyodo¹, J. Le Douce¹, L. Pasquier³, E. Flori⁴, M. Gonzales⁵, C. Bénédicteau⁶, O. Boute⁷, T. Attié-Bitach⁸, J. Roume⁹, L. Goujon³, L. Akloul³, E. Watrin¹, V. Dupé¹, S. Odent³, M. de Tayrac^{1,2}, V. David^{1,2}. 1) Univ Rennes, CNRS, IGDR (Institut de génétique et développement de Rennes) - UMR 6290, F - 35000 Rennes, France; 2) Service de Génétique Moléculaire et Génomique, CHU, Rennes, France; 3) Service de Génétique Clinique, CHU, Rennes, France; 4) Strasbourg University Hospital, France; 5) Service de Génétique et Embryologie Médicales, Hôpital Armand Trousseau, Paris, France; 6) Service de Génétique, CHU, Nantes, France; 7) Service de Génétique, CHU, Lille, France; 8) Service d'Histologie-Embryologie-Cytogénétique, Hôpital Necker-Enfants-Malades, Université Paris Descartes, 149, rue de Sèvres, 75015, Paris, France; 9) Department of Clinical Genetics, Centre de Référence "AnDDI Rares", Poissy Hospital GHU PIFO, Poissy, France.

Holoprosencephaly (HPE) is a pathology of forebrain development characterized by high phenotypic and locus heterogeneity. The disease presents with various clinical manifestations at the cerebral or facial levels. Several genes have been implicated in HPE but its genetic basis remains unclear: different transmission patterns have been described including autosomal dominant, recessive and digenic inheritance. Conventional molecular testing approaches result in a very low diagnostic yield and most cases remain unsolved. In our study, we address the possibility that genetically unsolved cases of HPE present an oligogenic origin and result from combined inherited mutations in several genes. All patients initially underwent a routine diagnostic procedure (Whole Exome Sequencing and CGH-array analyses) following the American College of Medical Genetics and Genomics–Association (ACMG) guidelines. 26 unrelated families with no established diagnosis were considered eligible for the hypothesis of oligogenic inheritance and included in the study. Their exome data was reanalyzed using more permissive settings. To identify variants of clinical interest, we used an integrative approach combining variant- and gene-based prioritization involving clinical ontologies and analysis of gene co-expression networks. Clinical phenotyping and exploration of cross-species similarities with known mouse mutants were further performed on a family-by-family basis. Statistical validation was performed on 744 ancestrally matched healthy controls provided by the Genome of the Netherlands (GoNL) project. Variants of clinical interest were identified in 180 genes significantly associated with key pathways of forebrain development including Sonic Hedgehog (SHH) and Primary Cilia. Oligogenic events were observed in 10 families and involved both known and novel HPE genes including recurrently mutated *FAT1*, *NDST1*, *COL2A1* and *SCUBE2*. The incidence of oligogenic combinations was significantly higher in HPE patients compared to the GoNL control population (Fisher's exact test p -value = 3.237×10^{-14}). We also show that depending on the affected genes and pathways, patients present with particular secondary clinical phenotypes. This study reports novel HPE and HPE-related phenotype genes and supports oligogenicity as clinically relevant model in this disease. It also underlines that integrating clinical phenotyping in genetic studies will improve the identification of causal variants in rare disorders.

2531F

Screening for 22q11.2 deletions and duplications in a large biobank data-set at a tertiary medical center uncovers new and undiagnosed cases. J.J. Connolly¹, M.H. Harr¹, J.T. Glessner¹, F.D. Mentch¹, F. Mafra¹, T.B. Crowley^{1,2}, P.M.A. Sleiman^{1,2}, D.M. McDonald-McGinn^{1,2}, H. Hakonarson^{1,2,3}, Electronic Medical Records and Genomics (eMERGE) Outcomes Group. 1) Childrens Hospital of Philadelphia, Philadelphia, PA; 2) The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; 3) Divisions of Human Genetics, Childrens Hospital of Philadelphia, Philadelphia, PA, USA.

Background: The biorepository at the Center for Applied Genomics (CAG) at the CHOP contains samples from over 100,000 CHOP patients and over 20,000 parents. In addition to biosamples, the majority of patients have consented to electronic health record (EHR) review with regular updates and re-contact, facilitating longitudinal collection of health information. It has been recognized that there is variability in the deletion breakpoints across the 22q11.2 critical region leading to variable inclusion of critical genes, and consequently, some phenotype-genotype correlations have been established. The phenotype of 22q11.2 duplications is reported to be more variable than the reciprocal deletion, and therefore more difficult to diagnose clinically.

Methods: To further investigate the prevalence and variability of these syndromes we screened a dataset of over 30,000 genotyped subjects who had been randomly recruited to the CAG biorepository and approximately 10,500 genotyped subjects from the Electronic Medical Records and Genomics (eMERGE) consortium in search of copy number variation (CNV) across the 22q11.2 region. CNVs were classified according to breakpoints and size consistent with the regions of segmental duplications LCR22A through LCR22H. EHR review was completed for all CHOP patients to determine whether the diagnosis was previously known by the patient's healthcare providers. Participants with previously unknown CNV(s) underwent review of clinical features to identify findings consistent with 22q11.2 deletion or duplication syndrome.

Results: We identified 48 deletions and 34 duplications in 81 CHOP patients and 10 duplications in eMERGE participants. Of 81 patients, 37 patients did not have EHR documentation indicating the diagnosis was previously known. The undiagnosed cases were duplications (34) or atypical deletions (3). EHR review revealed fewer hospital encounters and fewer items on the problem list for those with duplications than those with deletions. **Conclusions:** These data suggest that even at tertiary medical centers with expertise in 22q copy number variation, a proportion of patients with 22q CNVs remain undiagnosed, in particular those patients who do not have a standard LCR22A-LCR22D deletion or those who carry a duplication. EHR review adds to the knowledge of the phenotypic spectrum of patients with 22q11.2 duplications. Data lend support to national efforts of newborn screening for this relatively common syndrome.

2532W

Development of CNV analysis method for newborn screening using Global Screening Array (GSA). *M.-J. Kim¹, J. Joo², K.C. Lee², H.J. Kwon², A. Goyal², A. Kurkciyan¹, A. Espiritu¹, M.-F. Wu¹, J.-S. Bae², S. Lee², M.S. Lee^{1,2}.* 1) Diagnostics, Inc. 5795 Kearny Villa Rd, San Diego, CA 92123 USA; 2) Eone-Diagnomics Genome Center (EDGC), Inc. #291 Harmony-ro, Yeon-su-gu, Incheon, 22014, South Korea.

Structural variation is the structural alteration in the genome including duplications, deletions, translocations, and inversions. Some copy number variations (CNVs) are clearly related to diseases such as Angelman, Cri-Du-Chat, Digeorge and Jacobsen syndromes. It is critical to detect these variations as early as possible to prevent the onset of symptoms or minimize/reduce disease impact. Genome-wide genotyping microarrays have been used to detect those CNVs, which is reliable and efficient method for large-scale analysis. Newborn screening test using genome-wide CNV analysis can detect chromosomal abnormalities of a variety of genetic disorders. With a high resolution, microarray-based CNV analysis is a useful method to identify microdeletions and microduplications as well as novel CNVs. The microarray has provided a genomic tool for research applications including disease risk profiling studies, pharmacogenomics research, wellness characterization, and complex disease discovery. The Infinium Global Screening Array (GSA, Illumina, USA) contains multi-ethnic genome-wide content, curated clinical research variants, and quality control (QC) markers for precision medicine research. In addition to the consortium-developed 646,824 SNP markers, we designed 34,136 additional markers including pharmacogenomics and Korean-specific contents, resulted in 680,960 markers in total. In this present work, we performed analytical validation of the GSA for CNV detection using Coriell control human genomic DNAs for the newborn screening.

2533T

Co-occurrence of recurrent copy number variations in a child with complex clinical phenotype. *B. Melegh^{1,2,3}, A. Till^{1,2,3}, M. Czako^{1,2,3}, A. Szabo^{1,2,3}, K. Hadzsiev^{1,2,3}.* 1) Department of Medical Genetics, University of Pecs, Pecs, Hungary; 2) Szentagothai Research Center, University of Pecs, Pecs, Hungary; 3) Clinical Center, Medical School, University of Pecs, Pecs, Hungary.

Two copy number variations observed on the same chromosome in a patient is quite rare. Here we report a patient, who beside of the 17p11.2 microduplication had an additional deletion of 17q12 region. The routine cytogenetic investigation, which was indicated at 8 months of age of the baby because of neonatal hypotonia, developmental delay, renal cysts, hearing loss and minor anomalies, resulted in normal karyotype. Because of the complex clinical symptoms array CGH analysis using Agilent Sureprint 8x60K oligo-array (ISCA v.2) was performed. The array CGH detected a 3,573 Mb recurrent microduplication of 17p11.2 and additionally a 1,638 Mb deletion affecting 17q12.; and the unexpected results provided explanation of the complex clinical features allowing for a more detailed genotype-phenotype analysis. Our patient has been diagnosed as having Potocki-Lupski syndrome, renal cysts, and diabetes syndrome at the same time, which is a novel phenotypic variant of this disease group.

2534F

Chromosome 22q11 microdeletion: Potential modifiers of the cardiovascular phenotype found in heart development genes. *B. Rebolledo-Jaramillo¹, G. Repetto¹, F. Benavides¹, H. Poggi¹, A. Cuiza¹, P. Alvarez².* 1) Center for Genetics and Genomics, Clin Alemana - Univ Desarrollo, Santiago, Chile; 2) Hospital Dr Roberto del Rio, Santiago, Chile.

Chromosome 22q11.2 microdeletion syndrome (22q11DS) is the most common microdeletion syndrome in humans, with an estimated frequency of 1/4000. The syndrome shows marked variable expressivity not associated with deletion size or parental origin, suggesting the presence of modifiers elsewhere in the genome. Approximately 50-60% of patients have congenital heart disease (CHD), so the aim of this study was to search for genetic modifiers of the cardiovascular phenotype through whole exome sequencing (WES). We included 60 patients with 22q11DS: 25 cases with conotruncal severe CHD and 35 controls. WES was performed at the Northwest Genome Center (NWGC), using Roche Nimblegen V2 Exome capture and 75 bp paired-end sequencing. Variant calling was done with BWA+GATK best practices, and 22 cases and 33 control remained after quality control. We conducted a weighted combination of burden and SKAT tests, on loss of function (LOF) and rare (MAF <0.001) predicted deleterious single-nucleotide variants (rdSNVs), defined as missense variants where at least two of six predictors SIFT, PolyPhen-2, MutationTaster, MutationAssessor, FATHMM and CADD predicted damaging impact (6,560 variants in total), and found 58 genes with p-value<0.05. Among them, a single gene, *NES*, matched gene pathways related to CHD. The *NES* gene has been described as a putative target of *WNT1* during heart development, and experiments in rodents have shown its role in heart and brain development. We then relaxed the rare variant criteria and analysed all predicted deleterious LOF and missense variants (23,363 variants in total). We found 304 genes with p-value<0.05, nine of which matched CHD related pathways: *MBP*, *HDAC10* and *NOTCH2* (neural crest differentiation), *SPHK1*, *HNF4A* and *S1PR3* (human stem cell pluripotency), *ADAM19* (mesodermal commitment), *BMP1* (cardiac progenitor differentiation), and *DOCK8* (factors involved in megakaryocytes development and platelet production). Neural crest cells are critical for the outflow tract septation, and histone modifier genes, such as *HDAC10*, has been previously associated with CHD. However, there not seem to be a single gene responsible for the cardiovascular phenotype observed in our cohort. In fact, the variants within the genes were rare, 1 to 4 alleles among the 22 cases. So, for the other cases, variants in genes in the remaining 22q11 region or elsewhere in the genome may be associated with the presence of CHD. Funding FONDECYT Chile 1130392, 1171014 and 3170280..

2535W

Novel candidate genes to nonsyndromic orofacial clefts revealed by exome and target gene sequencing. *L.A. Brito, S. Ezquina, G. Yamamoto, G. Hsia, S.G. Ferreira, M.R. Passos-Bueno.* Department of Genetics, University of Sao Paulo, Sao Paulo, Sao Paulo, Brazil.

Nonsyndromic cleft lip with or without cleft palate (NSCL/P) is a prevalent developmental disorder affecting ~1:1000 live-born children. Despite presenting a high heritability, which reaches 85% in Brazil, common variant can only explain a small fraction of, according to recent meta-analyses. On the other hand, the role played by rare, high-effect genetic variants in NSCL/P complex etiology has increasingly been focus of attention. Here, we used exome sequencing in familial cases of NSCL/P to unravel the relevance of rare variants to the disease, as well as the role of new candidate genes. We sequenced, in total, 32 affected individuals from 10 familial cases, segregating NSCL/P under autosomal dominant-like inheritance. After filtering for shared variants within families (assuming a major causal variant) and prioritizing rare (<.005) variants in functionally relevant genes, we identified, for two families, a pathogenic variant in the epithelial-cadherin gene (*CDH1*). For the remaining families, a list of 11 candidate variants in novel candidate genes was raised. Pathway analysis including all genes harboring rare variants in these families indicated an enrichment of Wnt signaling pathway genes (both canonical and non-canonical / Planar Cell Polarity pathway). To investigate the ethiological role of these pathways, sequencing of a panel containing 31 genes with known expression during lip and palate development, in 500 NSCL/P patients from familial cases, have also been performed.

2536T

Somatic mosaicism in the MAPK-ERK signaling pathway in sporadic cerebral arteriovenous malformations. S. Gao, D. Guo, M. Lawton, L. Pawlikowska, H. Kim. Center for Cerebrovascular Research, University of California San Francisco, 1001 Potrero Avenue, San Francisco, CA 94110.

Cerebral arteriovenous malformations (AVMs), with an estimated 0.01% frequency in the population, are a leading cause of hemorrhagic stroke in young adults. Most cerebral AVMs are sporadic, suggesting that somatic mutations might underlie these lesions. A recent study reported two recurrent somatic mutations in *KRAS*, a member of the MAPK-ERK signaling pathway, in cerebral AVM lesions at low allelic frequencies (0.9% to 4%) in 74% of patients studied. Given that somatic mutations in several genes of the MAPK-ERK pathway are associated with other vascular anomalies, we hypothesized that somatic mosaicism in other genes of the MAPK-ERK signaling pathway might contribute to the development of cerebral AVMs. **Methods:** We analyzed whole-exome sequencing data generated from paired AVM tissue and peripheral blood of 17 unrelated patients with sporadic cerebral AVMs. Somatic mutations were called using GATK Mutect2, intersected with a list of 267 genes belonging to the MAPK-ERK signaling pathway (KEGG database), and filtered to identify genes harboring somatic mutations in more than one patient and recurrent somatic mutations. We applied three filters: 1) at least three reads containing somatic mutations in AVM tissue; 2) no reads containing somatic mutations in blood; 3) allelic frequency of candidate somatic mutations greater than 3%. Two patients were excluded due to poor sequencing quality, as indicated by a greater number of variants identified compared to other patients. **Results:** Average sequencing read depth was 91X across the exome. We found 29 genes in the MAPK-ERK signaling pathway bearing candidate somatic mutations in at least two patients. Among these, we identified three recurrent somatic mutations in three genes (*CACNA2D4*, *RASGRF1*, and *KRAS*), including the previously reported *KRAS*^{p.G12V} mutation. *KRAS* mutations were validated using droplet digital PCR, with a high concordance of allelic frequency of the mutant allele (p.G12V) between whole-exome sequencing (4.80%) and droplet digital PCR (4.88%). **Conclusions:** We have replicated the finding of somatic, recurrent *KRAS* mutations in sporadic cerebral AVM. Potential somatic mosaicism was also observed in a subset of genes of the MAPK-ERK signaling pathway in patients with sporadic cerebral AVM.

2537F

Understanding the genetic contribution to heterotaxy syndrome. A. Sridhar, SM. Ware. Wells Center for Pediatric Research, Indiana University School of Medicine, Indianapolis, IN.

Heterotaxy (HTX) is a rare congenital anomaly syndrome accompanied by thoracic and/or abdominal situs abnormalities resulting from abnormal left-right patterning during embryonic development. HTX can occur as an autosomal dominant, recessive, or X-linked condition, but is most commonly multifactorial. Individuals with primary ciliary dyskinesia (PCD) or syndromic ciliopathies can exhibit a HTX phenotype, but the prevalence of these conditions amongst individuals with HTX is unknown. We aimed to identify the incidence of Mendelian disease amongst prospectively recruited individuals with HTX and further delineate genetic associations by defining variant burden. In order to investigate known Mendelian genetic cause we interrogated 285 whole exomes and evaluated 172 genes compiled from HTX, PCD and ciliopathy clinical panels. Forty-three subjects (15%) revealed likely pathogenic variants that were X-linked (2.1%, including multiple participants with Oculo-Facio-Cardio-Dental syndrome), autosomal recessive (8.8%, including 5.6% with PCD, 2.1% with syndromic ciliopathies and 1.1% with other genetic syndromes) or autosomal dominant (4.2%, including multiple individuals with variants in *CHD7*). Eight individuals (2.8%) had multigenic inheritance patterns. In this cohort, 37 individuals (13%) had a family history of HTX, of which only 6 (16%) had an identifiable genetic cause. In order to further define genetic associations, a variant burden analysis was performed using an expanded candidate list of 2293 genes associated with HTX, cilia, or congenital heart defects. Using a Wilcoxon Rank-Sum test we show an increased burden of likely pathogenic variants in European cases compared to controls from the 1000 genomes study (p-value = 0.03). A parallel analysis with 2293 autism genes revealed no difference in variant burden (p-value = 0.8) between cases and controls. In summary, this study demonstrates that in a prospectively recruited cohort, 15% have identifiable genetic causes, including critically ill infants with unsuspected PCD or syndromic ciliopathies. However, even in individuals with a known family history of HTX, the molecular basis was identified in a minority of participants, indicating an opportunity for novel gene discovery. Variant burden analyses indicate increased pathogenic variant burden in European cases compared to controls, suggesting that HTX may result from combinatorial interactions of multiple susceptibility alleles in many individuals.

2538W

An open-label, phase 1/2 study of miransertib (ARQ 092), an oral pan-AKT inhibitor, in patients (pts) with PIK3CA-related Overgrowth Spectrum (PROS): Preliminary results. A. Bartuli¹, P.S. Buonuono¹, C. Leon², I. Rana¹, R. Onesimo³, M. Macchiaiolo⁴, A. Diociaiuti⁵, S. Livadiotti⁶, N. Resta⁵, M. Ruggeri⁶, A.D. Pratico⁶, Y. Sheldon⁷, R. Savage⁸, M. Lamar⁸, J. Kazakin⁸, B. Schwartz⁸, D.M. Adams⁷, G. Zampino². 1) Rare Diseases and Medical Genetics Unit, Bambino Gesù Children's Hospital, IRCCS, Rome, Italy; 2) Center for Rare Diseases and Birth Defects, Institute of Pediatrics, Department of Woman and Child Health, Fondazione Policlinico Universitario A. Gemelli IRCCS, Università Cattolica del Sacro Cuore, Rome, Italy; 3) Dermatology Unit, Bambino Gesù Children's Hospital, IRCCS, Rome, Italy; 4) Clinical Trial Unit, Bambino Gesù Children's Hospital, IRCCS, Rome, Italy; 5) Division of Medical Genetics, Department of Biomedical Sciences and Human Oncology (DIMO), University of Bari "Aldo Moro", Bari, Italy; 6) Unit of Rare Diseases of the Nervous System in Childhood, Department of Clinical and Experimental Medicine, Section of Pediatrics and Child Neuropsychiatry, University of Catania, Catania, Italy; 7) Division of Hematology, Vascular Anomalies Center, Boston Children's Hospital, Boston, MA, USA; 8) ArQule, Inc., Burlington, MA, USA.

Dysregulation of PI3K/AKT signaling pathway has been implicated in a number of human cancers and overgrowth syndromes, including PROS. In contrast to cancer, characterized by multiple genetic aberrations, in PROS a single activating mutation has been identified and shown to be associated with overgrowth. ARQ 092 is an oral, allosteric, selective pan-AKT inhibitor that potently inhibits AKT1, 2 and 3 isoforms. Both biochemical and cellular studies showed that ARQ 092 inhibited AKT activity through binding to its active and inactive forms. In PROS-patient derived cells, ARQ 092 demonstrated higher antiproliferative activity with lower cytotoxicity as compared to other PI3K inhibitors. In the dose escalation cohort, pts received ARQ 092 QD at up to three dose levels (15-25-35 mg/m²). Based on the preliminary PK and safety data assessment, the highest biologically active and safe dose level was determined to be 25 mg/m². Patients enrolled in the expansion cohort receive ARQ 092 QD at up to two dose levels (15-25 mg/m²). To date, 11 pts (5 males, 6 females, median age 5.8 years [range 2.6-12.1]) with different phenotypes with confirmed PIK3CA mutations who satisfied all eligibility criteria were enrolled. All pts received ARQ 092 at the dose of 15 mg/m²; dose has been increased to 25 mg/m² in 2 pts and to 35 mg/m² in 1pt. All enrolled pts are ongoing and have been on treatment between 1 and 51 weeks. The primary objective is safety. Treatment-emergent adverse events (AE) were reported in 4 pts (36%) and were predominantly graded 1 or 2. Only one AE (skin mass) was assessed as drug-related. The most common AEs were fever (40%), rash and abdominal pain (each 20%). No pts were discontinued due to toxicity, intolerance or non-compliance. Preliminary efficacy is one of the secondary objectives. Some patients revealed clinical improvement of mobility and pain. In 5 pts who have been on treatment for at least 3 cycles, an improvement in Karnofsky/Lansky scale was noted. ARQ 092 has demonstrated preliminary activity and manageable toxicity in PROS pts. Further evaluation of ARQ 092 in this patient population is ongoing. Additional efficacy data, including volumetric imaging analyses; PK; pharmacodynamic; changes in Quality of Life, functional tests and performance status; and safety will be presented (NCT03094832).

2539T

An exome-wide association study identifies candidate susceptibility genes for congenital obstructive uropathy. D.F. Ahram¹, T.Y. Lim¹, R. Westland^{1,2}, P. Krithivasan¹, M. Bodria^{3,4}, F. Scolaris⁵, M. Saraga^{6,7}, G.M. Ghiggi⁸, A. Mitrović⁹, V. Tasic⁸, M. Gigante⁹, L. Gesualdo¹⁰, L. Allegri¹¹, K. Kiryluk¹, A.G. Gharavi¹, D.B. Goldstein¹¹, S. Sanna-Cherchi¹¹. 1) Division of Nephrology, Columbia University Medical Center, New York, USA; 2) Department of Pediatric Nephrology, VU University Medical Center, Amsterdam, The Netherlands; 3) Division of Nephrology, Dialysis, Transplantation, and Laboratory on Pathophysiology of Uremia, Istituto G. Gaslini, Genoa, Italy; 4) Department of Medicine and Surgery, University of Parma, Italy; 5) Cattedra di Nefrologia, Università di Brescia, Seconda Divisione di Nefrologia, Azienda Ospedaliera Spedali Civili di Brescia Presidio di Montichiari, Brescia, Italy; 6) Department of Pediatrics, University Hospital of Split, Split, Croatia; 7) School of Medicine, University of Split, 21000, Split, Croatia; 8) University Children's Hospital, Medical Faculty of Skopje, Skopje, Macedonia; 9) Department of Surgery and Medical Science, University of Foggia, Foggia, Italy; 10) Section of Nephrology, Department of Emergency and Organ Transplantation, University of Bari, Bari, Italy; 11) Institute for Genomic Medicine, Columbia University, New York, USA.

Introduction Congenital Obstructive Uropathy (COU) represents a common cause of kidney failure in children and adolescents. Monogenic forms of disease account for a small portion of cases, highlighting high genetic heterogeneity. Rare variants burden association tests represent promising approaches to identify novel disease genes. **Methods** We conducted whole exome sequencing (WES) in 292 unrelated patients with COU and 8,541 population controls. Following diagnostic variant annotation in genes associated with congenital anomalies of the kidney and urinary tract (CAKUT) using ACMG criteria, we searched for novel susceptibility genes using case-control exome-wide collapsing analyses. We tested gene-level dominant and recessive models using a Fisher-exact association test for ultra-rare functional variants. Qualifying variants for association were selected using computational algorithmic scores based on conservation and location in intolerant domains. We analyzed gene-set enrichment using 299 OMIM genes implicated in Mendelian and complex forms of CAKUT and compared to a control Olfactory gene set. Finally, we cross-annotated suggestive signals with results from a genome-wide association study (GWAS) for common variants composed of 398 cases and 8,197 controls. **Results** ACMG-based diagnostic variant prioritization revealed pathogenic mutations in 11 (3.7%) COU cases. Gene-level collapsing showed 20 genes with a $P < 5 \times 10^{-4}$ for dominant models ($MAF \leq 0.0005$) and 2 genes for recessive model ($MAF \leq 0.01$). Among these candidates, two were of particular interest. 1. *MAZ* (dominant missense model; $P = 3.58 \times 10^{-5}$) was recently shown to regulate urogenital development in a mouse model. 2. *FLNA* (recessive model; $P = 4.25 \times 10^{-5}$), a gene implicated in Otopalatodigital Spectrum Disorders, where hydronephrosis and megaureter have been reported. Cross-annotation with GWAS results showed supportive evidence for association of common variants to *MAZ* ($P = 7.06 \times 10^{-3} - 6.96 \times 10^{-4}$; joint P -value for common and rare variants = 4.6×10^{-7}), thus reinforcing candidacy for this gene in human COU. Gene set analysis showed enrichment for rare functional CAKUT variants ($\chi^2 = 9.12$; $P = 2.5 \times 10^{-3}$) compared to Olfactory control genes ($\chi^2 = 0.15$; $P = 0.70$) in cases versus controls. **Conclusions** This study presents an integrated approach to collectively address the contribution of rare and common variants in the etiology of COU. Preliminary results suggest rare and common variants in *MAZ* as novel risk factors for COU.

2540F

Rare inherited disorders that arise as a result of abrogating multiple systemic pathways. S. Mian, M. Samman, E. Faqeh, W. AlTuraif, M. Mudish, A. AlNawfal. King Fahad Medical City, Riyadh, Saudi Arabia.

Introduction The ability to introduce effective clinical management protocols for patients presenting with rare inherited disorders is predicated upon accurate delineation of loci with abrogated function. Exome sequencing facilitates rapid screening of variants that are pathogenic in nature. Non-random population dynamics can alter the frequency of inherited disorders and consanguinity is one such mechanism by which enrichment/selection can produce a bias in pathogenic allelic frequency. Evidence is presented of dual molecular diagnosis in a single patient from consanguineous parents. **Materials and Methods** IRB approval was obtained from KFMC ethics committee to undertake a retrospective analyses of clinical exome sequence data. DNA libraries were constructed using Agilent SureSelect v5. Sequencing was carried out to an average depth of coverage of 150x. Quality metrics were assessed using FASTQC v.0.11.5. Alignment, quality filtering and variant identification were undertaken using Lasergene v15.0 (DNASTAR, USA) and reference assemblies aligned against GRCh37.p13. QIAGEN Clinical Insight (Qiagen, USA) interpret software was used in variant analysis and interpretation. **Results** One patient from consanguineous parents presented with a neurodegenerative disorder including: failure to thrive, developmental regression, hypotonia and lactic acidosis. Clinical exome sequencing identified two known pathogenic homozygous variants in distinct genetic loci. The first variant impacted mitochondrial gene function while the second abrogated pathways linked to glycosylation. **Conclusions** In regions with high consanguinity, dual (and higher order) pathologies are likely. Molecular laboratories need to be vigilant that patients may present with more than one genetic disorder and consequently guide physicians on the management of these patients.

2541W

Ascertainment of APOE $\epsilon 4$ homozygotes without Alzheimer disease for investigation of protective factors: The ASPREE Healthy Ageing Biobank. A.J. Huq^{1,2}, M. Riaz², J. Ryan², I. Winship¹, E. Storey^{1,2}, R. Woods², J. McNeill², P. Lacaze^{1,2}, ASPREE Investigator Group. 1) Clinical Genomics, Royal Melbourne Hospital, Parkville, Victoria, Australia; 2) Department of Epidemiology and Preventive Medicine, School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia.

Alzheimer disease and related dementias are a significant public health concern globally, with an estimated 50 million individuals affected. Late onset Alzheimer disease (LOAD) is the predominant cause of dementia, yet characterization of the genetic causes of LOAD remain incomplete. The APOE $\epsilon 4$ allele is the strongest known genetic risk factor for LOAD, which in the homozygous state (estimated 2% of the population), confers an odds ratio of up to 15 for developing dementia, compared to the common APOE $\epsilon 3$ allele. Despite this elevated risk, some APOE $\epsilon 4$ homozygotes remain unaffected by Alzheimer disease to an advanced age, seeming to outlive their genetic risk. However, these individuals are difficult to ascertain and make the focus of research, due to their advanced age, rarity of genotype in the population and requirement of cognitive assessments to confirm the absence of cognitive decline. Here we describe the identification, and focused investigation, of 46 healthy elderly APOE $\epsilon 4$ homozygotes aged 75 years or older, with no diagnosis of dementia, cognitively screened to pass the Modified Mini-Mental State Examination (3MS) with a score >77 . A further 123 APOE $\epsilon 4/\epsilon 4$ cognitively normal participants aged between 70 and 75 are being followed up longitudinally. These individuals are currently enrolled in the ASPirin in Reducing Events in the Elderly (ASPREE) study and Healthy Ageing Biobank. Consented DNA samples are currently being sequenced, to accompany the battery of cognitive testing data conducted over the past five years. Some individuals in this group are aged >90 years, and score above average in all cognitive domains. We will compare this cohort with a group of early-onset Alzheimer disease cases (also $\epsilon 4$ homozygotes, >20 currently recruited) in an extreme-phenotyping study design. We conducted a systematic review of comparable studies, and found almost all were limited by sample size, and/or merged $\epsilon 4$ heterozygotes with $\epsilon 4$ homozygotes in study groups, lacking the adequate controls. Here, we describe the genomic screening of over 10,000 biobanked elderly individuals required to identify these 169 APOE $\epsilon 4$ homozygotes. We are generating polygenic risk scores in these individuals, to compare with those who develop dementia at a younger age, as well as other approaches to identify modifying factors. This cohort has the unique potential to identify possible protective factors in LOAD that may off-set APOE $\epsilon 4$ risk.

2542T

Serum metabolites associate with cognition phenotypes among Bogalusa Heart Study participants. M. Shi¹, J. He¹, C. Li², K. Yaffe³, S. Li¹, L.A. Bazzano¹, J. Kinchen⁴, P. Stuchlik¹, X. Mi¹, J.L. Nierenberg¹, X. Gu^{1,5}, T.N. Kelly¹. 1) Tulane University, New Orleans, LA; 2) University of Georgia, Athens, GA; 3) University of California, San Francisco, CA; 4) Metabolon, Inc., Durham, NC; 5) Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China.

Background Few studies have explored the relationship between the human metabolome and cognition using untargeted metabolomic profiling. Using this agnostic approach, we aimed to identify novel metabolites and metabolite modules associated with cognition in a cross-sectional analysis of Bogalusa Heart Study (BHS) participants. **Methods** Untargeted ultrahigh performance liquid chromatography-tandem mass spectroscopy was used to detect and quantify serum metabolites of 1,177 BHS participants (773 whites and 404 African-Americans). Phenotypes were collected by trained technicians using a battery of 8 cognitive tests and a global cognition score was calculated based on the crude scores from these tests. The association of single metabolites with cognition was assessed using multiple linear regression models to adjust for important covariables. Weighted correlation network analysis (WGCNA) was utilized to identify modules of co-abundant metabolites and examine their covariable adjusted correlations with cognition. All analyses were conducted according to race and in the entire population, using Bonferroni-corrected alpha-thresholds to determine statistical significance. **Results** Nine metabolites, including two with unknown biochemical identities, achieved Bonferroni-corrected significance, $P < 4.16 \times 10^{-5}$, in at least one of the three population groups. Among the seven named metabolites, four were from lipid sub-pathways that included fatty acid metabolism (9-hydroxystearate and 12, 13-Di-HOME), plasmalogen [1-(1-enyl-palmitoyl)-2-palmitoleoyl-GPC (P-16:0/16:1)] and primary bile acid metabolism (glyco-alpha-muricholate). The maltose, N-acetyl-isoputrescine and 7-methylguanine metabolites from the glycogen metabolism, polyamine metabolism and purine metabolism sub-pathways respectively were also discovered. In addition, there were three metabolite modules which achieved Bonferroni-corrected significance, $P < 5.56 \times 10^{-3}$, in whites and one module was also significant in the whole population. These three modules were largely comprised of metabolites from the lipids pathway. **Conclusion** The current study identified novel associations of 7 metabolites with known biochemical structures and 3 metabolite modules with cognition, providing new insights into the physiological mechanisms regulating cognitive function. Prospective studies are warranted to articulate their etiologic relevance in the development of cognitive impairment.

2543F

The construction of risk prediction models for dementia with supervised principal component analysis using miRNA expression data. D. Shigemizu^{1,2,3,4}, S. Akiyama¹, Y. Asanomi¹, K.A. Borojevich¹, A. Sharma^{3,4,5,6}, T. Tsunoda^{2,3,4}, K. Matsukuma⁷, M. Ichikawa⁷, H. Sudo⁷, S. Takizawa⁷, T. Sakurai^{8,9}, K. Ozaki^{1,10}, S. Niida¹. 1) Medical Genome Center, National Center for Geriatrics and Gerontology, Obu, Japan; 2) Department of Medical Science Mathematics, Medical Research Institute, Tokyo Medical and Dental University (TMDU), Tokyo, Japan; 3) Laboratory for Medical Science Mathematics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 4) CREST, JST, Japan; 5) School of Engineering & Physics, University of the South Pacific, Fiji; 6) Institute for Integrated and Intelligent Systems, Griffith University, Australia; 7) Toray Industries, Inc. Kanagawa, Japan; 8) The Center for Comprehensive Care and Research on Memory Disorders, National Center for Geriatrics and Gerontology, Aichi, Japan; 9) Department of Cognitive and Behavioral Science, Nagoya University Graduate School of Medicine, Nagoya, Japan; 10) Laboratory for Cardiovascular Diseases, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan.

With an increasingly aging global human population, the number of people with dementia is rapidly increasing. Alzheimer's disease (AD) is the most common subtype of dementia, followed by Vascular Dementia (VaD), and Dementia with Lewy Bodies (DLB). Although detection of these dementia, especially in the early stages, remains a substantial challenge due to the lack of specific biomarkers, microRNAs (miRNAs) have recently received a lot of attention as the novel biomarkers. Here, by means of serum miRNA expression of 2,181 Japanese individuals, we comprehensively investigated potential miRNA biomarkers. On a basis of a supervised PCA logistic regression method, we detected optimal biomarkers and constructed risk prediction models according to subtype of dementia. The final risk prediction model achieved a high accuracy of 0.873 on a validation cohort in AD, when using 78 miRNAs: Accuracy=0.836 with 86 miRNAs in VaD; Accuracy=0.825 with 110 miRNAs in DLB. We further predicted candidate target genes, of interest of pharmaceutical investigation, from the miRNAs using miRDB. Co-expression network analysis of the target genes revealed key functional modules/genes in dementia. This is also the first report applying miRNA-based risk prediction models to a dementia prospective cohort. Our study demonstrates our models to be effective in prospective disease risk prediction, and with further improvement, such as integrative analyses of genomic and/or transcriptomic data, may contribute to practical clinical use in dementia.

2544W

Understanding relationships among APOE4, BMI, Alzheimer's disease, and longevity, using Mendelian Randomization approach. *S. Ukrantseva, H. Duan, K. Arbeev, I. Zhbannikov, O. Bagley, A. Yashkin, E. Stallard, A. Kulminski, I. Akushevich, A. Yashin.* Duke University, Durham, 27708 NC, USA.

The APOE e4 allele is a major genetic risk factor for Alzheimer's disease (AD), and it has also been linked to BMI/weight loss in AD and to reduced longevity. However, the mechanism connecting the APOE e4 and AD (or longevity) through the aging-changes in BMI is unclear. In this study, we investigated the effects of carrying the APOE e4 on the behavior of individual age-trajectories of BMI, and the impact of an earlier and faster decline in BMI on AD risk and total survival in Framingham Study participants. We also employed Mendelian Randomization (MR) approach to help clarify causal connections between the earlier/faster decline in BMI and AD, as well as longevity, using the APOE e4 as an instrumental variable. We found that the APOE e4 significantly affected individual age-trajectories of BMI, so that it was associated with both earlier and faster declines in BMI in females, regardless of AD history. We also found that the probability of survival and staying free of AD at ages 75+ and 80+ depends on the age at which BMI starts to decline. Females with an early decline in BMI (in their 60s) had an increased risk of AD and all-cause mortality later in life. Next, in MR study, we found that the age at maximum BMI (after which it starts to decline) was a significant mediator of the APOE e4 effects on both AD and longevity in females. The rate of the decline in BMI was also causally linked to female longevity. These results suggest that APOE e4 may influence onset of AD in females through the earlier decline in BMI, and it may influence female longevity through both the earlier and faster declines in the BMI. One potential mechanism connecting the earlier decline in BMI and AD might involve female hormonal changes around menopause. This research was supported by the NIH grants P01AG043352 and R01AG046860.

2545T

Exploiting population cohorts to interrogate genomic and transcriptomic factors associated with exceptional blood aging. *E. Bader^{1,2}, M.J. Favé¹, V. Bruat¹, E. Gbeha¹, P. Awadalla^{1,2}.* 1) Computational Biology, Ontario Institute for Cancer Research Toronto, Ontario, Canada; 2) Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada.

Hematopoiesis is a tightly controlled, dynamic process that continuously replenishes the cellular populations that circulate in peripheral blood. While the negative effects of the aging process on hematopoiesis are well documented, the underlying factors that maintain healthy hematopoiesis in old age remain unknown. To investigate mechanisms that modulate the effects of the aging process on blood we exploit Canada's largest population cohort, the Canadian Partnership for Tomorrow Project (CPTP). Over 320 000 Canadians are enrolled in the CPTP and data available for these participants include RNA-seq from PBMCs (n=997), cytokine profiles (n=8975), genotypes (n=11 230), complete blood counts (CBC; n=43 128), and detailed health and lifestyle questionnaires. First, we characterize the transcriptional landscape of individuals relative to their age and CBCs. Differential gene expression analyses reveal that individuals with similar CBCs have few differentially expressed genes (DEGs), irrespective of age, while individuals similar in age with extremely different CBCs have many DEGs. We classify DEGs into genes that are involved in the normal aging process, genes that prevent blood aging, and genes that advance blood aging; GO biological processes enriched in these gene groups include cell adhesion (e.g. *FOXP3*), erythrocyte development (e.g. *SOX6*), and protein ubiquitination (e.g. *SPRTN*), respectively. Notably, *SPRTN* mutations are associated with genomic instability and have been observed in individuals with progeroid syndrome who present with advanced aging phenotypes. We also observe that the regulatory effects of some SNPs (e.g. rs992472166 in *VTI1B*) are dependent on an individual's age and CBCs by mapping expression quantitative trait loci (eQTLs). Interestingly, *VTI1B* has a role in autophagy, a process that, when induced, has been shown to increase longevity. Finally, to investigate the phenotypic effects of the DEGs and eQTL SNPs we test for associations with blood biochemistry variables, physical measures, and disease outcomes. Collectively, these results uncover genomic and transcriptomic factors associated with the blood aging process, which provide insight into the molecular mechanisms that regulate this process, and can be used to assess the extent of blood aging in an individual and potential health risks.

2546F

The Canadian longitudinal study on aging: Genome-wide genetic data on 9,900 participants available for your research! V. Forgetta¹, C. Darnmond-Zwaig², A. Belisle², R. Li², C. Balion³, C. Wolfson⁴, G. Lettre⁵, G. Pare³, A. Paterson⁶, C. Verschoor⁷, M. Lathrop², P. Raina⁷, J.B. Richards^{1,8,9}, J. Ragousis^{2,10}, *Canadian Longitudinal Study on Aging/Étude longitudinale canadienne sur le vieillissement*. 1) Centre for Clinical Epidemiology, Lady Davis Institute, Montreal, Quebec, Canada; 2) McGill University and Genome Quebec Innovation Centre, Department of Human Genetics, McGill University, Montreal, QC, Canada; 3) Hamilton Regional Laboratory Medicine Program, McMaster University, St. Joseph's Hospital St. Luke's Wing, Hamilton, ON, Canada; 4) Department of Epidemiology and Biostatistics and Occupational Health, McGill University, Montreal, QC, Canada; 5) Montreal Heart Institute and Université de Montréal, Montreal, Quebec, Canada; 6) The Hospital for Sick Children Research Institute, The Hospital for Sick Children, Toronto, ON, Canada; 7) Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, ON, Canada; 8) Department of Twin Research and Genetic Epidemiology, King's College London, London, UK; 9) Department of Human Genetics, McGill University, Montréal, Québec, Canada; 10) Department of Bioengineering, McGill University, Montreal, QC, Canada.

Genetic and environmental factors contribute to maintaining health and in the development of disease and disability as people age. The Canadian Longitudinal Study on Aging (CLSA) is a national long-term study that will follow 51,338 men and women with the aim to provide information on the changing biological, medical, psychological, social, lifestyle and economic aspects of participants' lives. Here we present the availability and quality assessment of genetic data for 9,900 participants. CLSA participants provided information through in-home interviews and additional in-depth information gathered through physical examinations and biological specimen collection (blood and urine) at one of 11 data collection sites across Canada. We have collected physical measures including cognition, vision (acuity, retinal camera, tonometry), audiometry, anthropometrics, blood pressure, electrocardiogram, carotid ultrasound, spirometry, performance tests, and bone mineral density and body composition by DXA on 30,097 participants. Whole genome genotyping was conducted on 9,900 participants using the Affymetrix Axiom array followed by imputation to the Haplotype Reference Consortium. Quality assessment of the genetic data included both marker- and sample-based tests, as well as analysis of sex-chromosome abnormalities, population structure, and familial relatedness. The genetic data release comprises both directly genotyped data for 794,409 genetic markers and whole-genome imputed data for >39.2 million genetic variants. Approximately 90% of the cohort is of Northern and Western European ancestry. The genotype imputation is of similar quality to UK Biobank, with 83% of the variants having an information quality score of >0.3. We detected 33 subjects with discordance between genetic vs self-reported sex and used copy number profiling to identify or confirm 19 participants with sex chromosome abnormalities, as well as 2 cases of Charcot-Marie-Tooth disease. We demonstrate the utility of this resource through genome-wide association studies on height, lipid traits, and bone mineral density. A future release will provide genetic data for an additional 20,000 participants. First follow-up of participants is due to end in mid-2018 with data release in Spring 2019. Qualified researchers from any country can easily access this genomic and phenotypic data release via the CLSA Data Access Portal at <https://www.clsa-elcv.ca/data-access>.

2547W

10-year follow-up of the super-seniors study: Compression of morbidity and genetic factors in survivors. L. Tindale^{1,2}, D. Salema¹, A. Brooks-Wilson^{1,2}. 1) Canada's Michael Smith Genome Sciences Centre, BC Cancer Research Centre, Vancouver, BC, Canada; 2) Biomedical Physiology and Kinesiology, Simon Fraser University, Burnaby, BC, Canada.

Background: Super-Seniors are healthy, long-lived individuals who were recruited at age 85 years or older with no history of cancer, cardiovascular disease, diabetes, dementia, or major pulmonary disease. In a 10-year follow-up, we aimed to determine whether surviving Super-Seniors showed compression of morbidity; and to test whether the allele frequencies of longevity-associated variants in *APOE* and *FOXO3* were more extreme in such long-term survivors. **Methods:** Super-Seniors who survived and were contactable were re-interviewed 10 years after initial characterization. Health and lifestyle were characterized via questionnaire. Geriatric tests including the Timed Up and Go, Geriatric Depression Scale, Instrumental Activities of Daily Living and the Mini-Mental State Exam were administered, and data were compared to results from on average 10 years earlier. As well, genotype and allele frequencies for SNPs rs7412 and rs429358 in *APOE*, and rs2802292 in *FOXO3* were compared to the frequencies in the original collection of Super-Seniors and mid-life controls. **Results:** Of the 480 Super-Seniors recruited from 2004-2007, 13 were alive, contactable, and consented to re-interview (mean age = 100.1 ± 3.3). Eight of these 13 participants (62%) still met Super-Senior health criteria. Diseases that occurred in late life were cardiovascular (5 of 13; 38%) and lung disease (1 of 13; 8%). The surviving group of centenarians had a higher frequency of *APOE* and *FOXO3* longevity-associated variants even when compared to the original long-lived Super-Senior cohort. **Conclusion:** Although physical and mental decline occurred in the decade between interviews, the majority of Super-Seniors re-interviewed still met the original health criteria. These observations are consistent with reports of compression of morbidity at extreme ages, particularly in centenarians. The increased frequency of longevity-associated variants in this small group of survivors is consistent with studies that reported a larger role for genetics at older ages.

2548T

Longitudinal study of gene expression and regulation during a critical period of human aging. B. Balliu¹, M. Durrant¹, O. de Goede¹, N. Abell¹, X. Li¹, B. Liu¹, M.J. Gloudemans¹, N. Cook², K.S. Smith¹, S. Jaiswal¹, C. Sabatti¹, L. Lind², E. Ingelsson¹, S.B. Montgomery¹. 1) Stanford University School of Medicine, Stanford, CA; 2) Uppsala University, Uppsala, Sweden.

Molecular and cellular changes are intrinsic to aging and age-related diseases. Several studies have investigated the combined effects of age and genetics on gene expression; however, there has been no large long-term, longitudinal characterization of these molecular changes within individuals. We performed RNA sequencing in whole-blood from the same 65 individuals at age 70 and 80 (130 samples), for which SNP array data was also available. Then, we examined how gene expression and genetic regulation were altered over this critical 10-year aging period. We discovered 1,291 genes whose expression level was significantly associated with age (FDR<5%). Enrichment analysis of the differentially expressed genes showed significant enrichment in age-related pathways, e.g. metabolism of proteins, adaptive immune system, and regulation of DNA replication and apoptosis. Moreover, we observed a depletion in the number of genetically-regulated genes at age 80, relative to age 70 (Binomial proportion test; $p\text{-value} = 3.3 \times 10^{-3}$). Genes with the largest loss of genetic regulation with age are involved in DNA repair and genome stability, e.g. *POLD3*; as well as in signal transduction pathways, e.g. *GNAS*. In addition, we observed that overall allelic imbalance increased slightly within individuals with age (Wilcoxon signed rank test; $p\text{-value} = 1.5 \times 10^{-3}$). We found four genes with individual-level differences in allele-specific expression (ASE) with age, e.g. *GNAS*; a known marker of clonal hematopoiesis that also showed a large loss of genetic regulation with age. Further, we found seven genes with population-level differences in ASE, e.g. *NEAT1*, which plays an important role in the innate immune response. Lastly, we identified 580 genes with significant patterns of age-specific alternative splicing, e.g. *SFPQ*, implicated in stress responses and double-strand break repair, and *SETX*, which has been associated with amyotrophic lateral sclerosis, an age-related disease with aberrant splicing patterns. In summary, we have performed the first long-term longitudinal study of human gene expression and genetic regulation of expression that highlights transcriptome dynamics late in life. Together, our findings highlight that age is associated with both population- and individual-level changes in gene expression, ASE, and alternative splicing and that the genetic architecture of a subset of genes is unstable and is characterized by a loss of genetic control at the population level.

2549F

Whole-genome sequence variants associated dyspnea among current and former smokers from the COPDGene study. M.N. McDonald¹, H. Tiwari², M. Flickinger³, V. Srinivasasainagendra⁴, P. Lakshman Kumar⁵, R. Casaburi⁶, E. Regan⁷, R.G. Barr⁸, A.W. Manichaikul⁹, M.E. Hall¹, E. Oelsner¹, M.H. Cho⁸, C.A. Vaz Fragoso⁹ for the COPDGene and NHLBI TOPMed Investigators. 1) University of Alabama at Birmingham, Birmingham, AL; 2) University of Michigan, Ann Arbor, MI; 3) Harbor-UCLA Medical Center, Torrance, CA; 4) National Jewish Hospital, Denver, CO; 5) Columbia University, New York, NY; 6) University of Virginia School of Medicine, Charlottesville, VA; 7) University of Mississippi Medical Center, Jackson, MS; 8) Harvard Medical School, Brigham and Women's Hospital, Boston, MA; 9) Yale University School of Medicine, New Haven, CT.

Dyspnea, also known as shortness of breath, is associated with poor health related quality of life and is a strong predictor of mortality. It is reported commonly in older individuals including those suffering from chronic obstructive pulmonary disease (COPD) and congestive heart failure (CHF) among other conditions. We hypothesize the identification of genomic regions associated with dyspnea may enable new diagnostic and treatment strategies for managing dyspnea across a wide range of disorders. Dyspnea was dichotomized based on whether a participant reported a modified Medical Research Council (mMRC) dyspnea score greater or equal to 2. We assessed whether both single nucleotide variants (SNVs) and gene-based regions were associated with dyspnea in 8321 current or former smokers aged 40 years or older from the COPDGene Study using whole-genome sequencing (WGS) as part of the Trans-Omics for Precision Medicine (TOPMed) program. SNVs, with minor allele frequencies (MAF) > 1%, were tested for association with dyspnea using EPACTS adjusting for age, sex, pack-years of smoking, lung function and principal components to control for population substructure. Gene-based tests were performed using SNVs with MAF < 5% using SKAT-O adjusting for the same covariates. Analysis of WGS data was performed using the University of Michigan's Encore pipeline. No SNV or gene-based region was associated at a level that achieved statistical significance. The top SNV associated with dyspnea was rs73882506 (MAF=4.4%, OR: 1.1, $P = 1.5 \times 10^{-7}$) intergenic between the genes *SPATA16* and *ECT2*. The top region associated in the gene-based testing was *PLCXD1* on the X chromosome ($P = 6.8 \times 10^{-6}$). Future efforts will be directed at expanding the analysis of dyspnea to additional TOPMed studies including the Framingham Heart Study, Jackson Heart Study, Multi-Ethnic Study of Atherosclerosis and ECLIPSE among others in addition to performing COPD and CHF stratified analyses.

2550W

Quantifying biological age using multiple modalities. *H. Tang^{1,2}, A. Berna^{1,2}, E.T. Cirulli^{1,2}, C. Leon Swisher^{1,2}.* 1) Human Longevity, Inc., San Diego, CA; 2) These authors contributed equally to this work.

Chronologically, we all age at the same rate, but the rates at which our bodies appear to age vary greatly from person to person. Multiple modalities can be used to measure changes that occur in our bodies, corresponding to different aspects of aging. For example, genomic DNA in cells gradually acquires damage over time via side effects of normal metabolic and molecular processes such as oxidative stress and cell division. Damage can accumulate at specific genomic locations and affect all cells, effectively acting as a genomic clock. Damage can also randomly spread throughout the genome of each cell, resulting in a state of somatic mosaicism that is thought to be an important cause for aging as well as other diseases. Here, we use whole-genome sequence data, untargeted metabolomics, and whole body and brain MRI scans to quantify individuals' biological ages. We find strong correlations between predicted biological age and actual chronological age, validating our models. Furthermore, we identify differences in the health characteristics of individuals whose chronological and biological ages are discordant. Our results indicate that measurements from multiple modalities can be harnessed to more precisely quantify the health and biological age of an individual.

2551T

No evidence of an association between mitochondrial DNA and accidental falls. *E. Ryu, G.D. Jenkins, P.Y. Takahashi.* Mayo Clinic, Rochester, MN.

Background Accidental falls are a major public health concern in the aging population. Mitochondrial dysfunction is potentially linked with accelerated aging and other factors associated with accidental falls like comorbidities. In this study, we assessed whether mitochondrial DNA (mtDNA) is associated with risk of accidental falls. **Methods** We used existing whole genome sequencing data from 953 Mayo Clinic Biobank participants. Accidental falls were defined as having at least one ICD-9 code for fall-related injuries treated at the emergency department (ED) between the time of blood collection and September 2017. We considered mtDNA copy number, single nucleotide variants (SNVs), and degree of heteroplasmy as factors related to mtDNA. MtDNA copy number was defined as the relative quantity of mtDNA compared to nuclear DNA, and the degree of heteroplasmy was defined as the proportion of non-reference alleles for a given SNV. Cox models were used to assess whether mtDNA-related variables were associated with risk of accidental falls during the follow-up period. The models were adjusted for age, sex, and comorbidities defined by the number of chronic conditions recommended by the US-DHHS. **Results** In the study cohort, 112 subjects (12%) experienced accidental falls requiring ED visits, with the median follow-up of 6.7 years (25th- 75th percentiles: 5.5 to 7.4 years). The study subjects had the median age of 61 year, 50% of females, and had the median of 3 chronic conditions at blood collection. Adjusting for age, sex, and the number of chronic conditions, mtDNA copy number was not associated with the risk of accidental falls (HR=0.67; 95% CI=0.29 – 1.56). Neither SNVs nor the degree of heteroplasmy was significantly associated with accidental falls after correcting for multiple testing. **Conclusions** Although the link between mitochondrial dysfunction and a wide range of age-related health outcomes has been widely supported in the literature, our study suggests that mtDNA may not play an important role in accidental falls that occur in the relatively short period of follow-up. However, a follow-up study with a larger sample size is required because the current study may lack statistical power.

2552F

The PhenX Toolkit: An update on a resource for geriatric research. L.A. Kilpatrick¹, M. Sano², S. Cummings³, P.L. De Jager⁴, C. Depp⁵, J. Guralnik⁶, T. Harris⁷, D. Houston⁸, C. Kawas⁹, D.P. Kiel¹⁰, R. Mayeux⁴, L. Farrer¹¹, M. Phillips¹, E.M. Ramos¹², W. Rossi¹, T. Hendershot¹, S. Hwang¹, D. Maiese¹, A. Riley¹, H. Pan¹, W. Huggins¹, C.M. Hamilton¹. 1) RTI International, Research Triangle Park, NC; 2) Icahn School of Medicine at Mount Sinai, New York, NY; 3) California Pacific Medical Center, San Francisco, CA; 4) Columbia University, New York, NY; 5) University of California, San Diego, San Diego, CA; 6) University of Maryland School of Medicine, Baltimore, MD; 7) National Institute on Aging, Bethesda, MD; 8) Wake Forest School of Medicine, Winston-Salem, NC; 9) University of California, Irvine, Irvine, CA; 10) Institute for Aging Research, Hebrew SeniorLife and Harvard Medical School, Boston, MA; 11) Boston University, Boston, MA; 12) National Human Genome Research Institute, Bethesda, MD.

In 2017, a Working Group (WG) of geriatrics experts convened to identify standard measures of phenotypes and exposures recommended for human genome, biomedical, and epidemiological research in older adults. As the Geriatrics WG finishes, we have started a Pediatric Development domain that also builds on existing Toolkit (TK) content. The Geriatrics WG followed a consensus-based process identifying 18 new measures and the information that supports their use (i.e., protocol [method used to collect data], how to administer the protocol) to the PhenX Toolkit. These measures, specifically for use in older populations, utilize protocols from the Health and Retirement Study (HRS), the Health Aging and Body Composition (ABC) Study, and other well-established sources. Along with presenting details of Geriatrics measures in the TK, we will share some gaps identified in available research instruments. The measures and protocols selected by the Geriatrics WG are available on the web-based PhenX Toolkit website (<https://www.phenxtoolkit.org/>). These new measures complement other measures for older adults available in the PhenX Toolkit. The PhenX Toolkit has been funded by the National Human Genome Research Institute (NHGRI) since 2007 and is a catalog of well-established measures covering 25 research domains with over 500 standard measures. This resource makes broadly validated, well-established measures available for biomedical researchers to improve statistical power and effects via cross-study analyses and helps facilitate study replication. Additionally, measures are valuable for researchers who are not experts in working with the older adults who may want to include the population in their study design. We will also present the new design and navigation of the PhenX Toolkit, updates on new content for Hemophilia and Health Equities, and collaborative efforts with the National Institutes of Health (NIH) Common Data Element (CDE) Initiative and the database of Genotypes and Phenotypes (dbGaP). The breadth of measures available in the TK is helpful when designing or expanding a study. The Geriatrics domain and measures in the PhenX Toolkit are another set of tools for the research community to use in its work to support research with older adult participants. *Funding provided by a Genomic Resource Grant (U41HG007050) from NHGRI, with co-funding from the National Institute on Drug Abuse (NIDA) and the Office of the Director, National Institutes of Health (OD).*

2553W

Varying effect of APOE alleles on extreme longevity in European ethnicities. A. Gurinovich¹, H. Bae², S.L. Andersen³, A. Puca^{4,5}, G. Atzmon⁶, N. Barzila⁶, T.T. Perls³, P. Sebastiani⁷. 1) Bioinformatics Program, Boston University, Boston, MA; 2) College of Public Health and Human Sciences, Oregon State University, Corvallis, OR; 3) Boston University School of Medicine, Boston, MA; 4) Department of Medicine and Surgery, University of Salerno, Fisciano, SA, Italy; 5) Cardiovascular Research Unit, IRCCS MultiMedica, Sesto San Giovanni, MI, Italy; 6) Albert Einstein College of Medicine, Bronx, NY; 7) Department of Biostatistics, Boston University School of Public Health, Boston, MA.

Apolipoprotein E (*APOE*) is a well-studied gene with multiple effects on aging and longevity. The gene has three common polymorphisms, E2, E3 and E4, whose frequencies vary by ethnicity and are associated with different levels of blood cholesterol. *APOE* E4 is a known risk factor for Alzheimer's and has a deleterious effect on extreme longevity, while E2 is associated with healthy aging and longevity. We analyzed ethnic specific effects of *APOE* alleles on extreme human longevity using genetic data of about 9,000 individuals from four studies of extreme longevity: the New England Centenarian Study, the Southern Italian Centenarian Study, the Longevity Gene Project, and the Long Life Family Study. The aggregated data comprised several European ethnicities and included 2144 cases of extreme longevity defined as individuals who lived past the 1 percentile survival age from the 1900 birth year cohort (i.e. age > 96 for males, and >100 years for females). For the analysis we used our new method PopCluster that combines logistic regression modeling, principal component analysis of genome-wide genetic data, hierarchical clustering, and a novel recursive bottom-up tree parsing procedure to automatically discover subsets of individuals in which the effects of a variant are statistically different. This analysis identified ethnically different clusters in which the effect of *APOE* E2 and E4 alleles on extreme longevity changed substantially. For example, PopCluster discovered a large group of 1089 subjects enriched of Southern Italians with weaker protective effect of *APOE* E2 and weaker damaging effect of *APOE* E4 on extreme longevity compared to other European ethnicities. Another example is a large cluster (977 subjects) enriched of subjects of Ashkenazi Jewish descent with strong protective effect of *APOE* E2 and significant detrimental effect of *APOE* E4 on extreme longevity. These results suggest possible interaction of this gene with nutrition habits or other environmental factors.

2554T

Comparison of multiple omics aging clocks. *E. Macdonald-Dunlop¹, P.K. Joshi¹, J.F. Wilson^{1,2}*. 1) Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, Scotland, United Kingdom; 2) MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, United Kingdom.

Background: Biological age, a measure of deterioration and aging that is distinct from chronological age has been found to predict disease and mortality (Fischer et al., 2014). Since the first published measure of biological age: Hannum's epigenetic clock, closely followed by Horvath's epigenetic clock, aging clocks have been built using telomere length (Zhang et al., 2014), facial morphology (Chen et al., 2015), metabolomics (Fischer et al., 2014), glycomics (Kristic et al., 2014) and proteomics (Enroth et al., 2015). However, comparison between clocks has so far been limited. Fortunately, the ORCADES study is a deeply annotated cross-sectional cohort with approximately 1,000 individuals, where multiple omics assays such as circulating plasma proteins, lipids, glycans and metabolomics are available, enabling consistent comparisons in a common dataset. **Aims:** (i) To replicate and potentially improve published aging clocks, and partition proteomic aging across protein panels (ii) for the first time to compare their accuracy in measuring age in a single dataset (iii) for the first time to analyse the effect of giving up smoking on the evolution of biological clock age. **Methods:** Elastic net regression were used to create multiple omics clocks in a training set and their accuracy assessed in a testing set and an additional independent cohort, in univariate and bivariate analyses. **Results:** We replicate existing clocks, (with some improvement of our proteomics clock over Enroth et al's, $r^2=0.83$) in predicting chronological age. We find our cardiovascular protein clock ($r^2=0.89$) outperforms our inflammation protein clock ($r^2=0.67$). We also find that our protein clock ($r^2=0.89$) outperforms our glycan clock ($r^2=0.45$). In bivariate analysis we find large but not complete overlap between the two clocks. Finally, we expand on Enroth et al's finding by showing that there is also a significant difference of average biological age between current smokers and ex-smokers and the half-life of the effect of having formerly been a smoker on biological age is 3.5 years. **Conclusions:** Proteins more accurately predict chronological age than glycans and can track the benefits of giving up smoking. Multi-omics clocks appear to track both overlapping and distinct aspects of aging and offer the prospect of identifying distinct aging pathways.

2555F

Exome variants associated with healthy aging: Findings from the Long Life Family Study (LLFS). *B. Vardarajan¹, B. Thyagarajan², A.I. Yashin³, J.M. Zmuda⁴, T. Perls⁵, M. Province⁶, J.H. Lee¹*. 1) Columbia University, New York, NY; 2) University of Minnesota, Minneapolis, MN, USA; 3) Duke University, Durham, NC, USA; 4) University of Pittsburgh, Pittsburgh, PA, USA; 5) Boston University, Boston, MA, USA; 6) Washington University, St. Louis, MO, USA.

Background: Exceptional longevity (EL) is a complex trait that is likely the result of an interaction between environmental, behavioral, genetic and stochastic factors. However, recent evidence from the Long Life Family Study (LLFS) suggests that the EL phenotype may also have strong genetic determinants. **Methods:** We selected 420 unrelated individuals from the LLFS families by computing a summary disease score (SDS) by summing across phenotypes in eight health domains - heart disease, stroke, lung disease, hypertension, diabetes, peripheral artery disease, cancer and dementia. Whole exome sequencing (WES) was conducted on 420 individuals (HiSeq 2500 sequencer, Roche exome capture) and rare coding variants were analyzed. We used independent WES data in 894 healthy elderly individuals from the Washington Heights and Inwood Community Aging Project (WHICAP) to compare LLFS variants in variant and gene-based tests. In addition, we replicated variants in 100 centenarians from Chinese Longitudinal Healthy Longevity Survey cohort. **Results:** We identified 13 potentially deleterious missense mutations that were found in at least four LLFS samples and absent in the ExAC database. Interestingly we found two damaging mutations in the *REST* gene and one in *SALL1* that segregated in elderly LLFS individuals. One variant in *REST* was also found in four centenarians in the CLHLS cohort. Variants in *REST* (putative master negative regulator of neurogenesis) confer susceptibility to cognitive aging and has been implicated in stress resistance in aging and Alzheimer's disease. Comparing single variants in LLFS with WHICAP, we found a frameshift mutation in *NOTCH2* ($p=1.9e-07$) that is more frequent in LLFS elderly individuals compared to the general WHICAP population. **Conclusions and Relevance:** We identified variants that confer susceptibility to EL in the LLFS cohort. We will present further evaluation of the genes discovered in independent long-living cohorts. Identifying the genetic determinants of EL could also lead to the discovery of genes that affect the rate of aging and predisposition for or against age-related diseases. Such discoveries would be helpful in the prediction, prevention and treatment of age-related diseases as well as predictors of and strategies for healthy aging.

2556W

Decreased burden of pathogenic variants in a healthy elderly cohort.

A. Telenti, S. Topol, E. Topol, A. Torkamani, J. di Iulio. The Scripps Research Institute, La Jolla, CA.

Background. Healthy aging is a distinct phenotype from extreme longevity and has been shown to be genetically driven. Here, we investigate whether the decreased burden of deleterious genetic variants contributes to healthy aging in the Wellderly cohort. **Methods.** The healthy aging individuals (Wellderly) cohort includes 1,152 individuals > 80 years old, without any chronic disease, and of non-finish European ancestry. As control group, we included non-finish European individuals from the gnomAD consortium (NFE, N=7,507), though to represent a younger representation of the general population. Pathogenic and benign variants were obtained from ClinVar. We focused on (i) cancer inherited genes and (ii) all genes annotated with pathogenic variants. **Results.** To confirm the experimental design, we first assessed the depletion of rs429358, marker of *APOE*E4* associated with early onset Alzheimer. Allelic frequency (af) in the Wellderly population vs controls was af=0.09 versus 0.14, $p=1.9e-10$. In particular, rates of homozygosity decreased to af=0.006 from 0.02, $p=0.0009$. As control, we compared the cumulative frequency of benign variants in 21 cancer inherited genes to confirm that the af were similar in both cohorts (mean and median ratio (af NFE/ af wellderly) were 1.016 and 0.985. To prevent biases due to the use of different pipeline to call variants in the 2 cohorts, we applied a *double ratio*: ((af pathogenic variants in Wellderly / af benign variants in Wellderly) / (af pathogenic variants in NFE / af benign variants in NFE)). We identified a depletion of inherited cancer variants in the Wellderly for 15 out of 21 genes. To assess whether the results extended to non-inherited cancer related genes, we investigated genes that were annotated with at least 1 pathogenic and 1 benign variant in ClinVar and variable in at least one individual in the cohort (N=877). 60% of those genes showed a *double ratio* below 1 (=depletion). The strongest depletion was found in the set of most essential genes ($pLI>0.9$, N=67, median *double ratio* of 0). **Conclusion.** There is a decreased burden of pathogenic variants in the Wellderly cohort. This approach supports analyses of the population impact of pathogenic variants on life span and the modelling of the health benefits of genetic screening.

2557T

Associations between a polygenic score for leukocyte telomere length and telomere length measures from multiple human tissues. M.B.

Chernoff¹, K. Demanelis¹, L. Tong¹, F. Jasmine¹, M.G. Kibriya¹, J. Shinkle¹, J. Doherty², B. Pierce^{1,2}, the GTEx Project. 1) The Department of Public Health Sciences, The University of Chicago, Chicago, IL; 2) The Department of Human Genetics, The University of Chicago, IL; 3) The Huntsman Cancer Institute and Department of Population Health Sciences, University of Utah, Salt Lake City, UT.

Background: Telomere length regulates the replicative capacity of cells and may contribute to the risk of developing cardiovascular disease and multiple cancer types. Previous genome-wide association studies (GWAS) have identified single nucleotide polymorphisms (SNPs) associated with leukocyte telomere length (TL), many of which reside in or near telomere-associated genes. However, it is unknown if these SNPs similarly influence TL in other human tissues. Here, we evaluate associations between known TL-associated SNPs and relative TL (RTL) measurements in 12 different tissue types from the Genotype-Tissue Expression (GTEx) Project. **Methods:** We used a Luminex-based method to measure RTL (i.e. telomere repeat abundance) for >4,500 GTEx tissue samples, restricting our analysis to tissues with >100 individuals with genotype data: whole blood, esophageal mucosa, pancreas, lung, stomach, colon, thyroid, testis, cerebellum, kidney, prostate, and ovary (ordered by descending sample size). The largest sample was whole blood with 444 individuals included. We used GTEx genotype data to calculate a genetic score for TL using 8 TL-associated SNPs identified from GWAS of leukocyte TL. We used linear models to assess the association of these SNPs and the SNP score with RTL in each tissue. We also built a linear mixed model for RTL, treating tissue as a random effect and SNP score as a fixed effect. **Results:** While individual SNPs did not show significant associations with RTL in any tissue, we observed a significant, positive association ($p<0.05$) between the SNP score and RTL measured in whole blood and the cerebellum. Associations between the SNP score and TL were positive in 9 out of 12 tissues ($p=0.05$ based on a binomial test), showing that SNP scores reflecting long TL tend to associate with tissue-specific TL measurements in the expected direction. In the mixed model including all tissues, the SNP score was positively, but not significantly, associated with RTL ($p=0.1$), explaining only 0.3% of RTL variation. **Conclusion:** Our results suggest that a SNP score for leukocyte TL may also predict TL in other human tissues. The small number of samples for each tissue limited our ability to detect statistically significant associations, thus we are currently measuring RTL for additional GTEx samples. We are also exploring the use of multivariate association methods that borrow information across tissues to increase the power to detect SNPs affecting TL in multiple tissues.

2558F

Genetically predicted telomere length and aging-related physical and cognitive traits among the UK Biobank participants. K. Demanelis, L. Tong, B.L. Pierce. Public Health Sciences, The University of Chicago, Chicago, IL.

Telomere length (TL) shortens in most human cell types and is a potential biomarker and cause of aging. However, the causal relationship between TL and physical and cognitive age-related decline has not been extensively examined. Using a Mendelian randomization (MR) approach, we estimated genetically predicted TL (GP-TL) and assessed its association with over 50 aging-related traits in 337,545 UK Biobank (UKB) participants age 39-72 years. We manually curated aging-related traits among the cognitive function, physical measures, and self-reported questionnaire variables across 11 health domains. The effect estimates from 9 TL-associated SNPs were extracted from the ENGAGE consortium GWAS of leukocyte TL. We estimated GP-TL in base pairs (bp) by constructing a weighted linear combination of the TL-associated SNPs. Multivariate regression models assessed the relationship between GP-TL and each aging-related trait, adjusting for age, sex, and 10 genotyping principal components. We obtained MR estimates from the two-sample inverse weighted (IVW) meta-analysis approach. Under MR assumptions, the MR estimate was interpreted as the effect corresponding to one standard deviation (SD) increase in TL. We identified 12 age-related traits associated with GP-TL with $p < 0.05$; however, only 5 traits reached Bonferroni-corrected significance ($p < 10^{-3}$). Pulmonary traits, forced expiratory volume (FEV1) and vital capacity (FVC), increased by 29 ($p=3 \times 10^{-4}$) and 43 ($p=1 \times 10^{-5}$) mL per 1,000 bp GP-TL. Half of the UKB participants were hypertensive, and unexpectedly, we observed that pulse pressure (PP), systolic (SBP) and diastolic (DBP) blood pressure increased by 1.2 ($p=6 \times 10^{-14}$), 1.9 ($p=5 \times 10^{-16}$) and 0.6 ($p=9 \times 10^{-6}$) mmHg per 1,000 bp GP-TL. From IVW MR analyses, TL increased PP, SBP, and DBP by 1.5 ($p=2 \times 10^{-3}$), 2.3 ($p=4 \times 10^{-6}$) and 0.8 ($p=1 \times 10^{-3}$) mmHg per SD, respectively, while FEV1 and FVC increased by 35 ($p=1 \times 10^{-2}$) and 52 ($p=4 \times 10^{-3}$) mL per SD in TL. Other two-sample MR methods yielded similar results; however, the MR-Egger intercept test differed from 0 for SBP ($p=5 \times 10^{-4}$), suggesting possible bias in its IVW MR-estimate. Under MR assumptions, these findings suggest that longer TL favorably increased pulmonary function while it adversely increased blood pressure. Further research is necessary to understand the relationship among GP-TL, measured TL, and blood pressure, to assess potential pleiotropy, and to evaluate GP-TL in cardiovascular and pulmonary age-related decline risk.

2559W

Novel genetic loci identified for telomere length leveraging 50,000 whole genome sequences in the trans-omics for precision medicine (TOPMed) project. M.A. Taub¹, K. Iyer², J. Weinstock³, A.R. Keramati⁴, J. Lane⁵, T. Blackwell⁵, L.R. Yanek⁴, N. Pankratz⁵, G. Abecasis⁵, R.A. Mathias⁴ on behalf of the NHLBI TOPMed Consortium. 1) Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD; 2) Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD; 3) Center for Statistical Genetics, University of Michigan School of Public Health, Ann Arbor, MI; 4) The GeneSTAR Research Program, Johns Hopkins School of Medicine, Baltimore, MD; 5) Department of Laboratory Medicine and Pathology, University of Minnesota School of Medicine, Minneapolis, MN.

Prior GWAS have identified 11 loci harboring common genetic determinants of telomere length (TL); several of these loci have been implicated in human disease. These studies have relied on genotype array/imputation data and TL generated through qPCR or Southern Blot. Large-scale whole genome sequencing (WGS) gives the opportunity to analyze TL data on large numbers of subjects with extensive phenotyping to (1) generate estimates of TL; (2) expand our knowledge of genetic determinants of TL; and (3) add to our understanding of the role these variants may play in disease etiology. WGS data were generated within the NHLBI Trans-Omics for Precision Medicine (TOPMed) project, and TL was estimated on 93,219 samples using TelSeq software. We performed single variant association tests using mega-analysis, combining data from 27 TOPMed studies on a subset of 49,899 subjects within the current genotype data freeze in TOPMed (~86 million variants after filtering sites with minor allele count below 5). We used a two-stage approach, initially screening variants with a standard linear model adjusting for sex and study, followed by a linear mixed-model also adjusting for population structure and relatedness on the subset of variants with $p < 0.01$ from the first-stage analysis (conducted on the Analysis Commons, <http://analysiscommons.com/>). We discovered four novel TL loci with common variants (MAF > 1%), including three with strong biological interest: (1) *TERF1* ($p=4.8 \times 10^{-9}$), encoding a telomeric repeat-binding factor; (2) *RFWD3* ($p=2.7 \times 10^{-9}$), encoding a protein involved in DNA damage repair; and (3) *TINF2* ($p=3.5 \times 10^{-10}$), a component of the shelterin complex, which protects against telomere shortening; and (4) *LINC01429* ($p=3.0 \times 10^{-8}$). There were 47 rare variants that achieved genome-wide significance for TL. We replicated 6 common loci previously associated with TL: *NAF1* ($p=4.2 \times 10^{-11}$), *OBFC1* ($p=6.4 \times 10^{-16}$), *TERC* ($p=4.9 \times 10^{-20}$), *TERT* ($p=1.3 \times 10^{-26}$), *ZNF208/ZNF676* ($p=1.5 \times 10^{-8}$), and *RTEL1* ($p=7.8 \times 10^{-11}$). Our TOPMed Working Group is examining methods to correct for batch effects in these TL data. In the near future, we will include an additional ~45,000 TOPMed subjects with WGS that will improve our ability to (1) fine-map the common variant loci; (2) identify novel loci through gene-based approaches; and (3) given known differences in TL by population, maximize our power to identify genetic determinants within groups.

2560T

The curious parental age and genetic effects on newborn telomere

length. W. Wong¹, B. Solomon², X. Liu¹, W. Zhu¹, J. Niederhuber^{1,3}. 1) Inova Translational Medicine Institute, Falls Church, VA; 2) GeneDx, Gaithersburg, MD; 3) Department of Public Health Sciences, University of Virginia, Charlottesville, VA.

Telomeres play a vital role in maintaining genomic integrity and chromosomal stability in humans. Previous studies have shown that the child's leukocyte telomere length (TL) is correlated with the father's age at conception and that the mechanism for elongation of telomere lengths in the sperm is not via the more common telomerase reverse transcriptase (TERT) pathway, but via the alternative lengthening of telomeres (ALT) pathway. The ALT pathway is characterized by elevation of TL heterogeneity whereas the TERT minimizes that. Most of these studies suffer from small sample sizes and incomplete family members, as well as studying as an adult, where measurement is a combination of initial TL and the TL attrition during aging. In this study, we estimate the TLs using the software telseq and telomerecat in 1,235 whole genome sequenced (~40X) newborns and their parents. This cohort is part of the "Childhood Longitudinal Study" at the Inova Translational Medicine Institute. Using multiple linear regression, we find that the TL of the newborn child is positively correlated with both parents' TLs ($P < 10^{-14}$), and is negatively correlated with gestational age ($P = 1.26 \times 10^{-8}$) male gender ($P = 9.57 \times 10^{-5}$) and paternal age at conception, whereas maternal age does not appear to be correlated with newborn's TL. While we did not detect a significant change in the slope among older fathers ($P = 0.95$, father's age > 33.49, median father's age), there is a significant difference in the variance in newborn's TL from older vs. younger fathers ($P = 0.003$). On the contrast, there is no difference in variance in the TLs in either mothers or fathers when divided into two groups by their median age ($P > 0.1$). Further, using the public GWAS data, we use 11 loci that are significantly associated with leukocyte TLs to calculate the polygenic risk scores (PRS) for TLs for each individual. These loci are mostly located near genes involved in TL maintenance and regulation. We show that only PRS from mother is significantly associated newborn's TL ($P = 0.08$) after accounting for both parents' TLs. Interestingly, when we separate the PRS into transmitted and untransmitted alleles from the parents, only the transmitted PRS from the mother is significantly associated with newborn's TL ($P = 0.08$). Taken together, these observations suggest female-based transgenerational telomere erosion, whereas the elongation of TLs in spermatozoa may be of a different mechanism from that in the leukocyte.

2561F

APOE is associated with declining BMI in older adults. C.R. Bauer¹,

M.T. Oetjens², G. Chittoor², E.D.K. Cha², A.E. Justice², S.A. Pendergrass¹. 1) Geisinger, Rockville, MD; 2) Geisinger, Danville, PA.

To better inform health policies, medical decisions, and our understanding of the pathogenesis of obesity, a refined understanding of the dynamic nature of body mass is needed. Specifically, how and why obesity changes across the life course and what health risks are associated with specific trajectories. Elevated body mass index (BMI) is a well-established risk factor for cardiometabolic diseases and excess body weight is a leading contributor to mortality in the developed world. Additionally, low BMI and weight loss can also be risk factors, especially in the elderly. However, few studies to date have considered the dynamic nature, the ultimate impact on health, and the underlying genetic contributors to the variability of BMI across the life course. To address these gaps in our understanding, we set out to identify the genetic determinants of longitudinal BMI changes using electronic health record (EHR) data in Geisinger's MyCode Community Health Initiative that is linked to genetic data as a part of the DiscovEHR collaboration. Our sample included 69,215 adults of European descent with at least 5 measurements of weight (median = 46) spanning at least 5 years (median = 13.8 years) and whole genome array data imputed to the Haplotype Reference Consortium (HRC) reference panel. For each patient, we modeled BMI as a linear function of age (BMI slope) and performed associations with the BMI slope, assuming an additive genetic model, and adjusting for age at baseline, sex, mean BMI across encounters, and principal components to control for population structure. We identified a variant in *APOE* (rs429358, MAF = 0.13) that was significantly associated with declining BMI in adults over the age of 50 ($p = 4 \times 10^{-9}$, $\beta = -0.03$ kg/m²/year). *APOE* variants are associated with a variety of conditions that are generally comorbid with obesity such as atherosclerosis as well as neurodegenerative diseases such as Alzheimer's. However, we find that age-related decline in BMI linked to *APOE* is not attenuated by conditioning on these variables and our results may help to explain conflicting reports that obesity can be both protective or a risk factor for Alzheimer's disease, depending on age. Our future directions include further investigations of genetic variants influencing weight change and elucidating the extent to which changes in adiposity at various life stages influences downstream disease outcomes.

2562W

Interaction analyses of genes and environmental exposures contributing to childhood wheeze in the Canadian Healthy Infant Longitudinal Development (CHILD) study. J. Choi¹, C. Dharma², A. Lamri³, A. Ambalavanan¹, D. Lefebvre³, S. Turvey⁴, P. Mandhane⁵, A. Becker⁶, M. Azad⁶, T. Moraes^{7,8}, M. Sears^{9,10}, P. Subbarao^{7,8,9}, Q.L. Duan^{1,2}. 1) Department of Biomedical and Molecular Sciences, Queen's University, Kingston, ON; 2) School of Computing, Queen's University, Kingston, ON; 3) Department of Clinical Epidemiology and Biostatistics, McMaster University, ON; 4) Division of Allergy and Immunology, Department of Pediatrics, University of British Columbia, BC; 5) Division of Pediatric Respiratory Medicine, University of Alberta, AB; 6) Department of Pediatrics and Child Health, University of Manitoba, MB; 7) The Hospital for Sick Children, Toronto, ON; 8) Department of Paediatrics, University of Toronto, Toronto, ON; 9) Division of Respiriology, Department of Medicine, McMaster University, Hamilton, ON; 10) Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, ON.

The World Health Organization (WHO) identified asthma, a chronic inflammatory disease of the airways affecting over 235 million individuals worldwide, as the most prevalent non-communicable disease (NCD) among children. Earlier studies estimated that genetic factors contribute to 35%-95% of asthma heritability, of which only a small portion may be explained by known loci. We hypothesize that the missing heritability lies in part within interactions among genes as well as between genes with environmental exposures. Using genomics data from the Canadian Healthy Infant Longitudinal Development study (CHILD; N=3455), we investigate the effects of common and rare genetic variants as well as their interactions with modifiable exposures on risk of asthma and asthma related outcomes throughout early childhood. We obtained genomic data using the Illumina HumanCoreExome BeadChip. After quality control and imputations, 22 million variants from 2830 children were included for analysis of recurrent wheeze reported from age 2-5, which is strongly correlated with asthma outcomes. Our exploratory genome-wide association study (GWAS) identified loci on chromosome 17q12, which has been repeatedly linked with childhood-onset asthma. Moreover, a weight-based association analysis of common (Genetic risk score analysis; GRS) and rare variants (SNP-set kernel association test; SKAT), identified sets of genetic variants significantly associated with wheeze, some of which are novel. In addition, gene-environment interaction analysis identified numerous variations correlated with increased prevalence of wheeze only in children who were exposed to prenatal smoking. Finally, epistatic interactions among genes were studied by constructing a network of inter-correlated variants via hierarchical clustering, a conventional machine learning method for grouping elements based on similarity. This analysis identified 8 clusters of genetic variants associated ($p < 0.05$) with childhood wheeze in our birth cohort. Our preliminary results indicate that both genes and environmental exposures contribute and interact to predispose to asthma and that these association signals may be detected in early childhood, prior to clinical diagnosis of asthma. On-going analyses include additional asthma-related outcomes such as longitudinal lung function and positive skin prick tests to allergens as well as examining more environmental variables such as pet ownership and nutrition.

2563T

Using whole genome sequence data from the Barbados asthma genetics study (BAGS) to elucidate the role of HLA in allergic disease. A. Shetty¹, M. Daya¹, S. Chavan¹, D. Kim², N.M. Rafaels¹, T.M. Brunetti¹, M.P. Boorgula¹, M. Campbell¹, P. Maul³, T. Maul³, A.R. Greenidge⁴, D. Walcott⁴, A. Franklin¹, R.C. Landis⁵, H. Watson⁴, M.A. Taub⁵, I. Ruczinski⁶, S.L. Salzberg⁵, T.H. Beaty⁷, R.A. Mathias¹, K. Barnes¹. 1) Department of Medicine, University of Colorado, Aurora, CO, USA; 2) University of Texas Southwestern Medical Centre, Dallas, TX, USA; 3) University of West Indies at Cavehill, Barbados, WI; 4) Faculty of Medical Sciences, The University of the West Indies, Queen Elizabeth Hospital, Bridgetown, St. Michael, Barbados; 5) Department of Biostatistics, Bloomberg School of Public Health, JHU, Baltimore, MD, USA; 6) Edmund Cohen Laboratory for Vascular Research, The George Alleyne Chronic Disease Research Centre, The University of the West Indies at Cave Hill, Barbados, W.I.; 7) Department of Epidemiology, Bloomberg School of Public Health, JHU, Baltimore, MD, USA.

RATIONALE: The human leukocyte antigen (HLA) complex encodes genes of the major histocompatibility complex and plays a central role in regulating the immune system. In this study, we use whole genome sequence (WGS) data from the Barbados Asthma Genetics Study (BAGS) generated through the NHLBI-supported Trans-Omics in Precision Medicine (TOPMed) program to call HLA genotypes and explore relationships between HLA genotypes and atopy. **METHODS:** HISAT-genotype is a novel software program for calling HLA genotypes by using a graph-based approach to align sequences from WGS data against HLA reference sequences from the IPD-IMGT/HLA Database. We used HISAT-genotype to perform HLA genotype calling on 998 samples from BAGS passing QC criteria. Carrier status for an HLA-allele (digit 1 & 2) was defined as having a called probability ≥ 0.4 . Carrier status was tested for association with a number of atopy phenotypes: any cockroach skin prick test (SPT, $n=266$), *Der p* SPT ($n=266$), *Blomia Tropicalis* specific IgE ($n=281$), *Der p* 2 specific IgE ($n=316$), phadiatop measures ($n=795$), and measures of total serum IgE (tIgE) in asthmatics ($n=410$) and non-asthmatics ($n=459$). Association analysis was performed using mixed effect models including relevant covariates (asthma, age, sex, tIgE batch, principal components). For each phenotype, p-values were corrected using the Benjamini & Hochberg method to control the false discovery rate (FDR). **RESULTS:** HLA A, B, C, DQA1, DQB1 had high call rates (92.0%, 96.0%, 98.3% and 99.6% respectively) and low Mendelian error rates (0.65%, 1.74%, 1.00% and 0.48% respectively), but HLA DRB1 had a relatively low call rate (79.8%) and high rates of Mendelian errors (35.6%). HLA DRB1 was therefore excluded from our association analysis. The HLA DQB1 06:09 allele was associated with increased tIgE in asthmatics (FDR $p=0.077$, uncorrected $p=0.0008$, carrier frequency=3.1%), but the association was not observed in non-asthmatics (uncorrected $p=0.67$). No other alleles had FDR-corrected p-values < 0.1 . **FUTURE WORK:** We will use HIBAG, an HLA imputation algorithm that uses SNP data to impute HLA genotypes, to call DRB1 alleles and to call HLA genotypes in additional subjects from this study with SNP array data (but no sequence data). In addition, we are in the process of measuring *Blomia tropicalis* specific IgE, a common allergen in the tropics, in all subjects included in this study.

2564F

Analysis of deep whole genome sequencing data from African American asthmatics. P.M.A. Sleiman, L. Tian, J. Glessner, X. Chang, Y. Liu, F. Mentch, H. Hakonarson. Children's Hospital of Philadelphia, Philadelphia, PA.

Asthma is one of the most common chronic diseases of childhood. Asthma prevalence varies significantly among racial/ethnic groups, African-Americans are disproportionately affected with 1.6 times the level of current asthma compared with white children and 7.6 times higher death rate, 2.6 times higher ED visits, 3 times higher hospitalizations, yet despite the increased incidence and adverse outcomes the genetic causes of asthma in African Americans remain understudied. We have generated deep (> 30X) whole genome sequencing on 4,400 African American asthmatics and controls as part of the TOPMed program. Detailed phenotypic and clinical measures were collected on all participants including spirometry, bronchodilator response, skin prick testing and IgE measurements. We geocoded the cohort to extrapolate environmental and socioeconomic traits including a deprivation index and two air pollution measures, daily spatial surfaces of O_3 (ppb) and $PM_{2.5}$ ($\mu g/m^3$). We present the results of WGS-derived burden and structural variant analyses as well as gene by environment analysis against the air pollution measures.

2565W

Gene expression responses at the 17q12-21 asthma locus to microbial exposures in nasal epithelial cells. M.M. Soliai^{1,2}, A. Kato³, C.T. Stanhope², J.E. Norton³, K.A. Naughton², A.I. Klinger², R.C. Kern⁴, B.K. Tan⁵, R.P. Schleimer³, D.L. Nicolae^{1,2,5,6}, J.M. Pinto⁷, C. Ober^{1,2}. 1) Committee on Genetics, Genomics and Systems Biology, University of Chicago, Chicago, IL; 2) Department of Human Genetics, University of Chicago, Chicago, IL; 3) Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL; 4) Otolaryngology-Head and Neck Surgery, Northwestern University Feinberg School of Medicine, Chicago, IL; 5) Department of Medicine, University of Chicago, Chicago, IL; 6) Department of Statistics, University of Chicago, Chicago, IL; 7) Department of Surgery, University of Chicago, Chicago, IL.

Genome-wide association studies (GWAS) of asthma identified the 17q12-21 (17q) locus as one of the most significant and most replicated genetic risks for childhood-onset asthma. We and others reported that genetic variants at this locus are associated with asthma risk only in children with rhinovirus (RV)-associated wheezing illness and with asthma protection in children exposed to barn animals in early life. SNPs at this locus are eQTLs for 2 genes, *ORMDL3* and *GSDMB*, and expression of both increased in response to RV in peripheral blood mononuclear cells (PBMCs). These data suggested that genetic risk at this important locus is environment-specific. To further explore these observations, we used a nasal epithelial cell (NEC) cell model of genotype by microbe interactions to study the effects of microbial exposures on 17q genes in individuals with and without asthma. We cultured NECs from 111 adults in the presence of RV-A16 or heat-killed *Staphylococcus aureus* (SA), and corresponding vehicles (RVveh and SAVeh) for 48 and 24 hours, respectively. RNA was collected from NECs for RNA-seq (50 bp, single-end). In contrast to studies in PBMCs, *ORMDL3* expression decreased in the presence of RV or SA compared to vehicle ($p < 10^{-10}$). Moreover, *ORMDL3* expression was higher in NECs from 42 asthmatics compared to 53 non-asthmatics in the vehicle-treated samples ($p < 10^{-2}$); but similar after exposure to RV or SA ($p > 0.4$). *GSDMB* expression was similar in all 4 treatments ($p > 0.10$). Finally, 17q SNPs were strong eQTLs for *ORMDL3*, but not *GSDMB*, in vehicle and treated samples (rs2290400: $p < 3 \times 10^{-6}$ and < 0.030 , respectively), with the asthma risk allele associated with increased *ORMDL3* expression. These results indicate that in cultured upper airway epithelial cells, *ORMDL3*, but not *GSDMB*, responds to microbes and is associated with genotype. *ORMDL3* expression decreased in response to both viral and bacterial exposures and was higher in asthmatics compared to non-asthmatics in untreated samples. These results suggest that higher constitutive expression of *ORMDL3* in airway epithelial cells, especially in individuals with the high-risk genotype, contributes to risk for childhood-onset asthma at the 17q21 locus. Supported by U19 AI106683; MMS was supported by T32 GM007197.

2566T

A large-scale genome-wide meta-analysis of gene-by-sex interactions on time-to-asthma onset identified sex-specific risk loci. R. Veil¹, R. Granelle², M. Ege³, C. Laprise⁴, M.B. Freidin⁵, M. Imboden⁶, A. James⁷, D. Jarvis⁸, E.K. Khushnudinova⁹, P. Margartte-Jeannin¹⁰, L.O. Ogorodova¹⁰, C. Samowski¹¹, M.-H. Dizier¹, A.J. Henderson², A.S. Karunas⁹, E. von Mutius³, N.M. Probst-Hensch⁶, B. Leynaert¹¹, F. Demeais¹, E. Bouzigon¹. 1) Inserm, UMR-946, Genetic Variation and Human Diseases Unit; Université Paris Diderot, Sorbonne Paris Cité, Institut Universitaire d'Hématologie, Paris, France; 2) Population Health Sciences, Bristol Medical School, University of Bristol, UK; 3) Dr von Hauner Children's Hospital, Ludwig Maximilian University; Comprehensive Pneumology Center Munich (CPC-M), German Center for Lung Research, Munich, Germany; 4) Département des sciences fondamentales, Université du Québec à Chicoutimi, Saguenay, QC, Canada; 5) Research Institute of Medical Genetics, Tomsk NRC, Tomsk, Russia; 6) Swiss Tropical and Public Health Institute; University of Basel, Basel, Switzerland; 7) Busselton Population Medical Research Institute, Department of Pulmonary Physiology and Sleep Medicine, Sir Charles Gairdner Hospital, Nedlands, Western Australia; School of Population Health, University of Western Australia, Crawley, Western Australia; 8) Respiratory Epidemiology, Occupational Medicine and Public Health, National Heart and Lung Institute, Imperial College; MRC-PHE Centre for Environment & Health, London, UK; 9) Institute of Biochemistry and Genetics - Subdivision of the Ufa Federal Research Center of the Russian Academy of Science; Bashkir State University, Department of Genetics and Fundamental Medicine, Ufa, Russian Federation; 10) Siberian State Medical University, Tomsk, Russia; 11) Inserm, Unit 700, Team of Epidemiology; Université Paris-Diderot Paris 7, Paris, France.

Asthma is a complex disease with sex-specific differences in prevalence, clinical and biological features. Asthma is more prevalent in males during childhood, while it becomes more frequent in females in adolescence and adulthood. The mechanisms behind these sex-specific differences are not well understood and may involve hormonal changes together with differential genetic predisposition. To identify genetic variants interacting with sex that influence time-to-asthma onset (TAO), we conducted a large-scale meta-analysis of nine gene-environment-wide interaction studies (GEWIS) of TAO (totaling 7,104 men and 6,970 females of European ancestry) by using survival techniques applied to pediatric and adult asthmatic and non-asthmatic subjects. We detected eight independent loci showing SNP×Sex interaction at the 10^{-5} level. Among these loci, three were sex-specific TAO risk loci. The most significant association with TAO was female-specific in an intergenic region at 5q32 ($P_{\text{female}} = 9.1 \times 10^{-8}$ versus $P_{\text{male}} = 0.56$). The other two associations were male-specific: within *SORCS2* intron 2 at 4q16 ($P_{\text{male}} = 1.3 \times 10^{-7}$ versus $P_{\text{female}} = 0.15$) and within *DGKB* intron 1 at 7p21 ($P_{\text{male}} = 3.9 \times 10^{-7}$ versus $P_{\text{female}} = 0.23$). The remaining five loci showed effects in opposite directions between males and females: 1p12 (within *GDAP2*, $P_{\text{male}} = 7.3 \times 10^{-4}$ versus $P_{\text{female}} = 4.6 \times 10^{-3}$), 2q32 (near *NCKAP1*, $P_{\text{male}} = 3.7 \times 10^{-3}$ versus $P_{\text{female}} = 3.7 \times 10^{-4}$), 3q26 (within *NLGN1*, $P_{\text{male}} = 7.3 \times 10^{-4}$ versus $P_{\text{female}} = 5.8 \times 10^{-5}$), 4p14 (in an intergenic region between *PCDH7* and *ARAP2*, $P_{\text{male}} = 6.2 \times 10^{-3}$ versus $P_{\text{female}} = 4.9 \times 10^{-5}$) and 7q34 (within *TBXAS1*, $P_{\text{male}} = 3.4 \times 10^{-4}$ versus $P_{\text{female}} = 4.9 \times 10^{-3}$). None of these loci had been previously associated with asthma phenotypes. Functional annotations indicated colocalization of these genetic variants with DNA regulatory elements in fibroblasts, lung or blood. By testing gene-by-sex interactions, we identified novel loci influencing asthma risk in a sex-specific manner. Most of the candidate genes in these loci are involved in inflammatory process and immune cell regulation. Further replication of these findings are ongoing.

2567F

A cell model of gene-environment interactions at the 17q21 asthma locus. C.W. Washington¹, J. Huffman¹, C. Tisler², Y.P. Liu², J.E. Gern², D.J. Jackson², R.F. Lemanske², J.A. Smith², C. Ober¹. 1) Department of Human Genetics, University of Chicago, Chicago, IL; 2) Division of Allergy, Immunology and Rheumatology, Department of Pediatrics, University of Wisconsin School of Medicine and Public Health, Madison, WI, USA.

Asthma is a common complex disease characterized by lower airway inflammation, resulting in airflow obstruction and clinical symptoms of wheezing, coughing, and breathlessness. However, the biological basis of asthma is largely unknown and diagnosis is often based on these clinical features. Heritability estimates of asthma span from 35% to 70%, but risk of asthma is also affected by a host of environmental factors. One particularly important exposure is rhinovirus (RV) infection, which is associated with wheezing illness in early life and asthma exacerbations throughout life. Asthma GWAS revealed a novel locus on chromosome 17q21 that is the most replicated and most significant asthma locus identified to date. Surprisingly, risk associated with this locus is confined to children with RV-associated wheezing illness in the first 3 years of life. We further explored this genotype-exposure interaction in a cell model to uncover genetic variation that influences transcriptional response to RV. In a previous study, we demonstrated that CD19+ B cells showed the highest fold-change of *ORMDL3* in response to RV infection of peripheral blood mononuclear cells (PBMCs) and the strongest genotype-specific responses. Here, we focused our studies on these cells from 120 children participants in the Childhood Origins of ASThma (COAST) prospective birth cohort. PBMCs were treated with either RV-A16 or vehicle. After 30 hours, CD19+ B cells were isolated and RNA was extracted for gene expression studies. Paired analysis of RV- and vehicle-treated samples revealed a strong treatment effect on expression of genes at the 17q locus with *ORMDL3* and *GSDMB* exhibiting an increased fold-change of 0.69 and 0.61, respectively ($p < 5 \times 10^{-4}$). Genotype for rs7216389, the lead SNP in the first asthma GWAS, was an eQTL for *ORMDL3* and *GSDMB* in both RV- and vehicle-treated samples, but there was neither a genotype effect on response nor genotype by RV interaction effects on response. Moreover, there were no differences in expression among children stratified by asthma at age 6 or by wheezing in the first 3 years of life. These data suggest that the previously observed interaction between rs7216389 genotype and RV-wheezing may be developmentally specific and/or other tissues (e.g., airway cells) may be the relevant site of interaction. This work was supported by P01 HL070831. CW was supported by T32 GM007197.

2568W

De novo indels within introns contribute to ASD incidence. A. Munoz¹, B. Yamrom¹, Y.H. Lee¹, P. Andrews¹, S. Marks¹, K.T. Lin¹, Z. Wang¹, A.R. Krainer¹, R.B. Darnell^{2,3,4}, M. Wigler^{1,2}, I. Iossifov^{1,2}. 1) Quantitative Biology, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY; 2) New York Genome Center, New York, NY; 3) Laboratory of Molecular Neuro-oncology, Rockefeller University, New York, NY; 4) Howard Hughes Medical Institute, Rockefeller University, New York, NY.

Over the past several years, analysis of whole-exome sequencing and microarray hybridization data from collections of families of children with autism like the Simons Simplex Collection (SSC) has increased our understanding of the genetic architecture of autism. Contributions from de novo (DN) mutation, rare and common variants have been established, but most progress has been made in the study of DN variants. DN germ line likely gene-disrupting (LGD), missense and copy number variants have been estimated to jointly contribute to approximately 30% of the autism in SSC [1]. The observed DN mutations in the ~5,000 affected children enabled the identification of lists of few hundred genes that with high confidence are involved in autism's etiology. For example, half of the 546 genes targeted by published DN LGD mutations (LGD targets) are expected to be true autism genes. We explored the contribution of DN noncoding mutation by focusing on the introns of the LGD targets. We analyzed whole genome data generated from the father, mother, a child with autism and an unaffected child of 1,886 families from the SSC. The 150bp paired-end dataset was generated at the New York Genome Center with average depth of 30x. About 80,000 de novo intronic substitutions (DIS) and 8,000 de novo intronic indels (DII) were identified. There was no significant difference between the numbers of all DIS and all DII between the affected and unaffected children. But in the introns of the 546 LGD target genes, we identified 273 DII in the 1,886 affected while we found only 218 in the 1,886 unaffected children. The difference of 55 events is significantly larger than 0 (p-value: 0.006). If we restrict to the half of the LGD targets that are more intolerant to damaging variants in human population and are expected to be even more enriched for autism genes [2], the difference raises from 55 to 59 and the significance increases (p-value: 0.0003). The increase of rates of DIIs in affected individuals was specific for the high quality autism candidate genes and was not observed in control gene lists. We did not observe a similarly convincing increase of the rates of DIS due to the larger background rates of the DIS. From the observed increase in DIIs rates in children with autism, we estimated that DN intronic variants might contribute ~15% of the autistic children in simplex families. 1. Iossifov et al., Nature, 2014 2. Iossifov et al., PNAS 2015.

2569T

Genome-wide linkage study meta-analysis of male sexual orientation.

A.R. Sanders^{1,2}, G.W. Beecham³, S. Guo³, J.A. Badner⁴, S. Bocklandt⁵, B.S. Mustanski⁶, D.H. Hamer⁶, E.R. Martin³. 1) Center for Psychiatric Genetics, NorthShore University HealthSystem, Evanston, IL; 2) Department of Psychiatry and Behavioral Neuroscience, University of Chicago, Chicago, IL; 3) Department of Human Genetics, University of Miami, Miami, FL; 4) Department of Psychiatry, Rush University Medical Center, Chicago, IL; 5) Laboratory of Biochemistry, National Cancer Institute, National Institutes of Health, Bethesda, MD; 6) Department of Medical Social Sciences, Northwestern University, Chicago, IL.

Male sexual orientation is a scientifically and socially important trait shown by family and twin studies to be influenced by environmental and complex genetic factors. Individual genome-wide linkage studies (GWLS) have been conducted, however, jointly analyzing the available GWLS datasets should maximize linkage information on the trait. Two main datasets account for >90% of the published GWLS concordant sibling pairs on the trait and are jointly analyzed here: Sanders (409 concordant sibling pairs in 384 families with 45,387 SNPs) and Hamer (155 concordant sibling pairs in 145 families with 408 STRPs). We conducted multipoint linkage analyses with Merlin on the datasets separately since they were genotyped differently. To integrate, we found the genetic positions of the respective markers in the Rutgers Map v.3 (hg19 build), and then used the non-parametric S-pairs and grid 1cM options to run multipoint linkage on both data sets, followed by combining LOD scores at each grid position across the marker sets. We continue to find the strongest linkage support at pericentromeric chromosome 8 and chromosome Xq28. We note that the pericentromeric chromosome 8 linkage peak contains a SNP with evidence from GWAS for the trait. Besides reporting these findings, we will report on incorporating the remaining published GWLS dataset (on 55 families) by using the multiple-scan probability approach on published summary statistics. This knowledge will maximize the positional information from GWLS of currently available family resources and help prioritize findings from GWAS and other approaches. Although increasing evidence highlights genetic contributions to male sexual orientation, our current understanding of contributory loci is still limited, consistent with the complexity of the trait. Further increasing genetic knowledge about male sexual orientation, especially via large GWAS, should help advance our understanding of the biology of this important trait. This work was supported NIH grants R01HD041563 and R21HD080410.

2570F

Contribution of transposable elements to disease and complex trait heritability. F. Hormozdiari^{1,2}, O. Weissbrod^{1,2}, C. Ju³, B. Geijn^{1,2}, M. Hujoei^{1,2}, S. Gazal^{1,2}, F. Hormozdiari^{4,5}, A. Price^{1,2}. 1) T.H. Chan School of Public Health, Harvard University, Boston, MA; 2) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; 3) Department of Computer Science, University of California, Los Angeles, California, USA; 4) Genome Center MIND Institute University of California, Davis, California, USA; 5) Department of Biochemistry and Molecular Medicine, School of Medicine University of California, Davis, California, USA.

Transposable elements (TEs), also called "repeat elements", comprise roughly half of the human genome. TEs were once dismissed as "junk DNA", but are now believed to be widely involved in gene regulation (Chuong et al. 2017 Nat Rev Genet). However, the contribution of TEs to disease and complex trait heritability has yet to be quantified. Here, we applied stratified LD score regression with the baseline-LD model (Gazal et al. 2017 Nat Genet) to 41 independent diseases and complex traits (average $N=320K$) to estimate the heritability explained by each of 856 TE annotations, including broad TE classes as well as specific TE families. We define enrichment as $(\% \text{ of heritability})/(\% \text{ of SNPs})$; a value less than 1 implies depleted heritability. We reached 5 main conclusions. First, the set of all TEs in the genome (54% of SNPs) was significantly depleted for trait heritability: enrichment (meta-analyzed across traits) of 0.67 ± 0.03 . The depletion varied across major TE classes, with enrichment estimates ranging from -0.02 ± 0.08 for LTR (9.8% of SNPs; $P < 10^{-12}$ for depletion; consistent with heritability close to 0) to 0.86 ± 0.10 for SINEs (16% of SNPs; $P = 0.08$ for depletion). Second, although most TE families were depleted for heritability, we identified 24 TE families that were significantly enriched ($P < 0.05/856$, corrected for multiple-testing), including 6.58 ± 0.42 enrichment for AmnSINE1. Third, most depletions were fully explained by the correlations between TE annotations and the annotations of the baseline-LD model (includes a broad set of coding, conserved, regulatory and LD-related annotations), providing a strong validation for the informativeness of the baseline-LD model; in particular, most TE families were depleted for regulatory function. In addition, the mean values of 3 measures of negative selection (nucleotide diversity, predicted allele age and McVicker B statistic) within these TE annotations mirror our findings. Fourth, older repeats explained more disease heritability than younger repeats, consistent with their stronger overlap with functional annotations. Finally, we identified 3 TE families that explained significantly more heritability for brain-related traits than for other traits ($P < 0.05/856$ for difference), including enrichment estimates of 0.89 ± 0.22 for brain-related traits vs. -0.11 ± 0.12 for other traits for *ERV1* (5.9% of SNPs). In conclusion, we show that TE contribute depleted but substantial heritability to diseases and complex traits.

2571W

The enhanced GWAS catalog: Catalysing discovery with the new REST API and improved web interface. D. Suveges¹, J.A.L. MacArthur¹, A. Bunie-lo¹, M. Cerezo¹, J. Hayhurst¹, P. Hall², L. Harris¹, L.A. Hindorff¹, H. Junkins¹, C. Malagone¹, A. McMahon¹, J. Morales¹, E. Sollis¹, O. Vrousseau¹, P.L. Whetzel¹, T. Burdett¹, F. Cunningham¹, P. Flicek¹, H. Parkinson¹. 1) European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD; 2) Division of Genomic Medicine, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA.

Over the past decade, thanks to the development of affordable large-scale genotyping technologies, genome-wide association studies (GWAS) have been providing invaluable insights into the genomic basis of various traits and disorders. The mission of the NHGRI-EBI GWAS Catalog (www.ebi.ac.uk/gwas) is to provide a highly curated, comprehensive collection of all GWAS publications and the discovered associations. Since 2008 the GWAS Catalog has collected and made publicly available more than 69,000 associations from more than 3,300 publications. To fulfill our aim of providing the best possible service for the scientific community, we continue to develop our resource to ensure users can efficiently extract association and study related meta-data satisfying their needs. Here we introduce our RESTful API that provides programmatic access to the GWAS Catalog data, enabling users to retrieve data from our database in a high-throughput manner. The RESTful API also enables integration of the GWAS Catalog into custom applications and services, so the returned data is always up to date without the need of constant maintenance of local datasets. To advance the accessibility of GWAS Catalog data we have enhanced the Catalog's web interface to better support common user queries. Our new publication, trait and variant-specific pages are designed to provide structured GWAS Catalog data and visualisations, along with links related to external data. For example, our newly developed publication page lists all GWAS analyses carried out in the paper and provides direct links to all discovered associations, while the trait page contains all publications, studies and associations reporting a specific trait. The variant page provides the details of the discovered associations and an insight into the genomic context of the variant, including LD information in the 1000 Genomes populations, and other already known phenotypes in the vicinity. These developments will vastly improve the utility of the GWAS Catalog to the scientific community, extending the use of Catalog data in investigations to identify causal variants, understand disease mechanisms, and establish targets for novel therapies..

2572T

Evaluation of genetic factors for morphine requirement prediction. J. Li^{1,2}, Z. Wei³, J. Zhang³, H. Hakonarson¹, S.D. Cook-Sather¹. 1) Children's Hospital of Philadelphia, Philadelphia, PA, USA; 2) Tianjin Medical University, Tianjin, China; 3) New Jersey Institute of Technology, Newark, NJ, USA.

Acute postoperative pain and opioid responsiveness analgesia demonstrate inter-individual variability and polygenic influence. Heritability estimates for acute experimental pain and opioid analgesia range from 12 to 60%. In 241 children of African American (AA) and 277 of European Caucasian (EC) ancestry, we sought to assess select candidate gene associations with morphine dose, high ($\geq 7/10$) pain, and low ($\leq 3/10$) pain and to develop dose prediction models that estimate dose prediction limits based on genetic factors. All subjects in the study were 4-18 years of age and had morphine as the sole intravenous analgesic component of a general anesthetic. All were genotyped using Illumina HumanHap550 or Human610-Quad arrays. Regression analyses revealed age, BMI and physical status as covariates for morphine dose. The Gene Database from the National Center for Biotechnology Information was queried to identify genetic loci affecting acute postoperative or experimental pain and morphine analgesic response as supported by both clinical and basic sciences research. Then association analyses between the candidate genetic loci and morphine dose, postoperative pain were performed via a linear regression model including significant covariates. Few associations replicated: morphine dose (mcg/kg) in African American children and *ABCB1* rs1045642 ($\beta = -9.30$, $p = 0.02$) and *OPRM1* rs1799971 ($\beta = 23.29$, $p = 0.02$); *KCNJ6* rs2211843 and high pain in AA subjects (OR 2.08, $p = 0.01$) and in congruent European Caucasians pain phenotypes; and *COMT* rs740603 for high pain in European Caucasian subjects (OR 0.69, $p = 0.046$). Top SNP candidate gene models with age, BMI and physical status covariates explained a maximum of 24.2% (EC) and 14.6% (AA) of morphine dose variance.

2573F

ATG16L1 and ATG12 are involved in the pathogenesis of gastric mucosal atrophy in Japanese patients. K. Obata¹, H. Ueda¹, M. Hirado¹, M. Taira¹, T. Inamine¹, D. Fukuda², K. Nakao³, H. Isomoto³, K. Tsukamoto³. 1) Dept Pharmacotherapeutics, Nagasaki Univ Grad Sch Biomed Sci, Nagasaki, Japan; 2) Dept Gastroenterology and Hepatology, Nagasaki Univ Grad Sch Biomed Sci, Nagasaki, Japan; 3) Div Med and Clin Sci, Dept Multidisciplinary Int Med, Faculty of Med, Tottori Univ, Yonago, Japan.

Purpose: It has been known that *Helicobacter pylori* (*H. pylori*) infection causes chronic gastritis, which leads to gastric mucosal atrophy, intestinal type of metaplasia, and eventually gastric cancer. Autophagy, a cellular degradation process, is involved in infection and inflammation. We investigated association between single nucleotide polymorphisms (SNPs) of autophagy-related genes and gastric mucosal atrophy of *H. pylori*-positive subjects to identify host genetic determinants of the disease progression. **Methods:** Colonization of *H. pylori* and gastric mucosal atrophy were evaluated based on concentration of anti-*H. pylori* antibody and serum pepsinogen, respectively. Two hundred Japanese subjects infected with *H. pylori* included 94 patients with atrophic gastritis and 106 subjects without atrophy. A total of 25 SNPs in the 4 genes (*ATG5*, *ATG10*, *ATG12*, and *ATG16L1*) were genotyped by polymerase chain reaction (PCR)-restriction fragment length polymorphism, PCR-high resolution melting curve analysis, and PCR-direct DNA sequencing. **Results:** Five SNPs (rs6431655, rs6431659, 6758317, rs7587051, and rs4663136) within *ATG16L1* and rs26537 within *ATG12* were significantly associated with mucosal atrophy of *H. pylori*-positive subjects. Among *ATG16L1* SNPs, rs6431655 showed the strongest association. The frequency of G/G genotype of rs6431655 was significantly increased in patients with mucosal atrophy ($P = 0.005$, OR = 2.838). Conversely, the frequency of C/C genotype of rs26537 within *ATG12* was significantly decreased in the atrophic patients ($P = 0.033$, OR = 0.357). **Conclusions:** The current study demonstrated association between autophagy-related genes and gastric mucosal atrophy induced by *H. pylori* infection for the first time. Genetic variants of autophagy pathway might be one of host genetic factors involved in the development of *H. pylori*-induced atrophy in Japanese.

2574W

The Ophthato™ Knowledgebase : A curated knowledgebase of over 500,000 ocular disease phenotypic records with ability to build complex queries to enable analyses of disease progression, prognosis, genotype-phenotype correlation and drug development. A. Das¹, S. S. Nagasamy², P. Raj³, B. Muthu Narayanan², J. Somasekhar², T. Chandrasekhar², D. Kumar², A. Ghosh³, S. Das³, S. Tejwani³, P. Narendra². 1) MedGenome Inc, Foster city, CA., Select a Country; 2) MedGenome Labs, Bangalore, India; 3) Narayana Nethralaya, Bangalore, India.

Ocular disorders comprise a wide spectrum of diseases affecting anterior or posterior segments of the eye, caused by infections, genetic changes (monogenic or complex), environmental stress and trauma. While advancements in genetic, molecular and proteomics technologies has led to a better understanding of disease pathology, specific treatment modalities and drug development for various complex ocular diseases remain. Challenges still exist in utilizing knowledge from genome wide association studies (GWAS) to identify disease relevant variants due to: a) lack of reproducibility of the same results in a replication study in a different population b) and limited well characterized phenotypic data to draw genotype to phenotype correlations across different populations. Complex diseases with a diverse manifestation of phenotypes and sub-phenotypes require studies on larger cohorts with comprehensive longitudinal phenotype and clinical data to address the unmet research needs in ocular diseases. To enable genomic, pharmacogenomic discovery or clinical research in ocular diseases, MedGenome Inc has launched the Ophthato™ Knowledgebase: a comprehensive collection of clinical, phenotype and biochemical data providing researchers and clinicians with a platform to design studies that address critical unmet needs in eye disorders. The searchable interface enables complex queries to select specific disease cohorts based on demographics, specific tissue affected, disease type and subtype, disease course or severity, drug response and many other clinical and phenotypic parameters. We will present case studies to showcase the utility of the Ophthato Knowledgebase. As a complete solution, we also provide access to genomic sequencing for cohorts of interest to enable a) validation of the disease relevant variants in a different population b) design GWAS studies and perform phenotype-genotype correlations c) novel biomarker discovery.

2575T

Disease individuals with bi-allelic recessive mutations, also show secondary-variant burden in disease-relevant biological modules. M. Kousji^{1,2,3}, O. Soylemez^{2,4,5}, A. Ozanturk³, S. Akle^{4,5}, I. Jungreis^{1,2}, J. Muller^{6,7}, C.A. Cassa^{4,5}, H. Brand^{8,9}, J.A. Rosenfeld¹⁰, M.Y. Wolf^{1,2}, A. Sadeghpour³, K. McFadden³, R.A. Lewis^{10,11}, M.E. Talkowski^{8,9,12}, H. Dollfus⁴, M. Kellis^{1,2}, E.E. Davis³, S.R. Sunyaev^{2,4,5,13}, N. Katsanis³. 1) MIT CSAIL, Massachusetts Institute of Technology (MIT), Cambridge, MA; 2) Broad Institute of MIT and Harvard, Cambridge, MA; 3) Center for Human Disease Modeling, Duke University Medical Center, Durham, NC; 4) Division of Genetics, Brigham and Women's Hospital, Boston, MA; 5) Department of Medicine, Harvard Medical School, Boston, MA; 6) Laboratoire de Génétique Médicale, Institut de Génétique Médicale d'Alsace, INSERM U1112, Fédération de Médecine Translationnelle de Strasbourg (FMTS), Université de Strasbourg, Strasbourg, France; 7) Laboratoire de Diagnostic Génétique, Institut de Génétique Médicale d'Alsace (IGMA), Hôpitaux Universitaires de Strasbourg, 1 place de l'hôpital, Strasbourg, 67091, France; 8) Molecular Neurogenetics Unit and Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, and Department of Neurology, Massachusetts General Hospital, Boston, MA; 9) Program in Population and Medical Genetics and Genomics Platform, The Broad Institute of MIT and Harvard, Cambridge, MA; 10) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 11) Department of Ophthalmology, Baylor College of Medicine, Houston, TX; 12) Program in Bioinformatics and Integrative Genomics, Division of Medical Sciences, Harvard Medical School, Boston, MA; 13) Department of Biomedical Informatics, Harvard Medical School, Boston, MA.

The importance of genetic background on driver mutations is well established but, the mechanisms by which the background interacts with Mendelian loci remains unclear. This challenge is particularly acute in rare disorders with low allele frequencies for both causal variants and secondary-locus contributory alleles (i.e. alleles in loci other than the primary locus), with population-based studies being hampered by the lack of statistical power due to very few events. To address the contribution of background genetic burden, we studied secondary-locus variation in two independent Bardet-Biedl syndrome (BBS) cohorts (n=191) with known recessive biallelic pathogenic mutations. We found that ultra-rare nonsynonymous variants are enriched in secondary loci, both versus population controls (N=850) and versus non-BBS diagnosis individuals with biallelic variants in the same recessive genes (N=50), using either in-cohort (MAF<0.2%; p=0.001; OR=2.11) or population-based (MAF<0.0001%; p=0.02; OR=2.88) minor allele frequencies. The secondary-locus variants preferentially disrupted overlapping functional genomic elements and thus are more likely to be deleterious and detrimental to protein function (p=0.025); were specific for nonsynonymous changes (synonymous variants were equally distributed between cases and controls (p=0.21), arguing against population stratification as a confounder); and concentrated in chaperonin-encoding genes (*MKKS*, *BBS10* and *BBS12*), a finding that was further corroborated by the observation of epistatic interactions involving this complex *in vivo*. These data intimate a complex genetic architecture involving the identification of secondary-locus variants that are likely to contribute to disease penetrance and expressivity, across multiple genes within biologically-relevant pathways. Though our observations are based on the study of a single recessive disorder (BBS), we postulate that this currently-untapped component of genetic architecture complexity is likely to be relevant for several other recessive and/or dominant disorders.

2576F

Are there genetic variants to explain the phenomenon of donor cell survival following blood transfusion? R. Hirani¹, M. Hobbs², A. Statham², M. Dinger^{2,3}, D.O. Irving¹. 1) Research and Development, Australian Red Cross Blood Service, Sydney, New South Wales, Australia; 2) GenomeOne, 370 Victoria Street, Darlinghurst, Sydney NSW 2010; 3) Kinghorn Centre for Clinical Genomics, 370 Victoria Street, Darlinghurst, Sydney NSW 2010.

Background: Despite the introduction of leukodepleted blood components, it has been shown that donor leukocyte engraftment (transfusion-associated microchimerism (TAM)) remains a long-term consequence of red blood cell (RBC) transfusion in some patient groups. The mechanism by which donor leukocytes survive following blood transfusion is unclear and it has been speculated that genetic factors may be one possible reason. This study was conducted to determine whether there is any association between an individual's genetic profile and the establishment of TAM in blood transfusion recipients. **Methods:** Australian trauma patients (n = 86) transfused with RBC units between 2000 and 2012, with an injury severity score of greater than 15, were recruited. Twelve of these patients were found to have TAM. To date, whole exome sequencing was conducted on four patients with TAM and four patients without TAM. Alignment of the exome data to the human genome reference sequence assembly GRCh37 was conducted and analysis of differential genetic variations was carried out. Variations unique to at least 50% of patients in each group were further investigated, with a primary focus on genes associated with immune regulation. **Results:** Preliminary analysis has been performed. Many of the changes found in 2 of the four patients with TAM, were present in highly heterogeneous genes including zinc finger protein genes and the human leukocyte antigen genes (HLA-B and HLA-DRB1). Some of the most interesting changes were found in genes with unknown function such as Chr4 70257329 variant ID248066. In addition changes were found in the following genes, BTN3A1 gene (Chr6 26405815 variant ID 340887) which plays a role in the regulation and proliferation of T-cells and in the CCL8 gene (chr 17 32647356 variant ID 847320) which is a chemokine which contributes to tumor associated leukocyte removal. **Conclusion:** Next generation (whole exome) sequencing technologies allow for the exploration of genetic variations across different population groups, which can provide clues for the potential indications of mechanisms by which TAM may establish in certain patients. This preliminary analysis provides potential genetic changes that could involve genes that have a role in T-cell regulation and proliferation. It is expected that further analysis with expanded patient numbers will allow for this genomic data to be explored in more detail.

Phenotype	CYP2C19	CYP2C9	CYP2D6	CY3A5	DPYD	TPMT	SLCO1B1	VKORC1
Poor Met	1	3	6	57				
Intermediate Met	15	19	6	10	2			
Intermediate or Poor Met						7		
Normal Met	28	45	55		65	60		
Rapid Met	14							
Ultra-rapid Met	9							
Normal Fun							47	
Decreased Fun							17	
Poor Fun							3	
Low Warfarin Sensitivity								29
Intermediate Warfarin Sensitivity								27
High Warfarin Sensitivity								11

2577W

Clinical utility of genotype screening in unselected individuals for personalized medicine. K.P. Lubieniecki^{1,2,3}, L.M. Fricke¹, J.M. Lubieniecka^{1,2,3}, M.M. Moore¹, B. Tucker¹, M. Mboob¹, C.F. Boerkoel¹, M.M. Soe¹, C.J. Carlson¹, S.J. Ahuja¹, D.W. Lee¹, M.E. Cornwell¹, C. Hajek^{1,2}, L. Carmichael^{1,2}, P. Cherukuri^{1,2,3}, R.E. Pyatt^{1,2,3}. 1) Sanford Health, Imagenetics, Sioux Falls, SD; 2) Sanford School of Medicine, University of South Dakota, Sioux Falls, SD; 3) Sanford Research Center, Sioux Falls, SD.

Background: Development of high-throughput and high accuracy (>99%) SNP genotyping platforms significantly reduced the per marker cost of genotyping. Inclusion of pathogenic and likely pathogenic variants identified by the American College of Genetics and Genomics (ACMG), as well, as pharmacogenetic variants recommended by Clinical Pharmacogenetics Implementation Consortium (CPIC) on a high throughput SNP screening array can enable preemptive screening for unselected patients. Based on the estimates of secondary finding from whole genome and exome sequencing studies we predict that approximately 1-2% of our patient population would be positive for a Medically Actionable Predisposition (MAP) finding. Based on published population frequencies, the frequency of variants affecting drug metabolism is expected to be at least 30%. **Methods:** To evaluate the technical performance of a custom version of the Illumina Global Screening Array (GSA), a cohort of 67 samples submitted for diagnostic testing to the Sanford Medical Genetic and Genomics Laboratories was analyzed retrospectively. To establish the reportable data set, we first removed variants targeted by probes that failed clinical and/or analytical validation. The remaining set was manually curated for pathogenic and likely pathogenic variants considered medically actionable genetic predispositions for adult onset disorders and medication efficacy. We confirmed all positive findings by Sanger sequencing or Taqman assays. **Results:** We identified 2 carriers of pathogenic variants in MUTYH and one carrier of pathogenic variant in MYBPC3 genes causing susceptibility for MYH-associated polyposis, pilomatixoma and familial hypertrophic cardiomyopathy-4, dilated cardiomyopathy-1A, respectively. All three positive cases were confirmed by Sanger sequencing. Frequencies of the pharmacogenomic phenotypes associated with 8 metabolic genes are presented in Table 1. We discuss the variants reported, the discordance between GSA results and results from orthologous tests, the need for supplemental testing, and the detection rates within the screened population. Table 1

2578T

Accurate prediction of multiple pigmentation phenotypes. *B. Alipanahi, P. Fontanillas, M. Multhaup, S. Shringarpure, C. Wilson, M. Johnson, S. Pitts, A. Auton, R. Gentleman, 23andMe Research Team. 23andMe, Inc., Mountain View, CA.*

Genetic studies, in particular, genome-wide association studies (GWAS), have established that pigmentation traits (eye color, skin color, and hair color) are highly polygenic in humans. They also have shown that the genetic architecture of these traits is largely overlapping and have suggested epistasis between the main loci. The shared genetic architecture of pigmentation phenotypes has not been exploited for predicting pigmentation thus far. Most prediction efforts have achieved relatively high accuracy in predicting the extremes of the human pigmentation distributions but suffer lower predictive accuracy for the intermediate phenotypes. Here, we applied a two-hidden layer deep neural network on self-reported pigmentation phenotypes from a large cohort of 23andMe research participants of European ancestry for joint-prediction of eye, hair, and skin colors. The collected phenotypes have eight, six, and six levels for eye color, hair color, and skin color, respectively, and provide fine-grained self-reporting of these phenotypes. Our model is trained on over 1,100 SNPs and 103,000 individuals and shows excellent prediction performance: When tested on out-of-sample data it achieves an aggregate AUC of 86%, 81%, and 76% for predicting eye color, hair color, and skin color levels. Although trained on categorical responses, the distribution of predicted eye color and hair color scores approximately matches the natural distribution of quantitatively-measured M index for eye color and C' score for skin color [1]. Our analysis identifies significant epistatic interactions between *SLC45A2* and *SLC24A4*, *IRF4*, *SLC24A4*, *MC1R*, and the *HERC2/OCA2* locus, among others. Notably, the effects of these interactions are different on each pigmentation phenotype. This model is a useful tool to predict the pigmentation phenotypes while preserving their inherent correlations, to understand and investigate the genetic architecture of these traits, and to analyze their involvement in skin cancer risk or other pigmentation-associated disorders. References [1] Genome-wide association studies of quantitatively measured skin, hair, and eye pigmentation in four European populations. Candille et al., PLoS One. 2012;7(10):e48294.

2579F

Itchy and scratchy: Polygenic risk scores predict self-reported mosquito attraction and bite response. *J.R. Ashenurst, S. Laskey, M. Multhaup, O. Sazonova, N. Furlotte, B. Koelsch, The 23andMe Research Team. 23andMe, Inc, Mountain View, CA, CA.*

Mosquitos present not merely an annoyance but also a significant threat to global public health. As such, susceptibility and response to mosquito bites are phenotypes of substantial interest. Previous research identified 15 independent genetic loci associated with self-reported attractiveness to mosquitoes, bite size, or itch intensity (Jones et al., 2017), and suggested shared phenotypic and genetic etiology. In the present study, we use a large ($n = 311,556$) cohort of unrelated 23andMe research participants of European descent to assess the predictive power of polygenic risk scores (PRS) for each of these three phenotypes. We performed novel GWAS of each phenotype, yielding 139, 105, and 128 independent markers ($p < 5e-6$) for mosquito attraction, bite size, and bite itchiness respectively. Tissue and gene set enrichment analyses implicated genes associated with these loci in blood and lymph tissues, and in gene sets involved in hematopoiesis, myelopoiesis, and leukocyte activation. Additive linear models trained in the GWAS discovery sample using these SNPs as features explained 1-2% of the variation in the three phenotypes in a held-out replication cohort ($n = 77,890$). Addition of age and sex features to create a combined risk score (CRS) increased the variance explained to 6-12%. Despite their modest predictive power, CRS meaningfully stratify participants by phenotype. For example, we observe a four-fold increase (59.5% vs 14.7%) in the fraction of participants rating themselves as "more attractive" to mosquitos than their peers among those above the 95th vs. those below the 5th percentiles of the CRS. The three self-reported phenotypes and corresponding PRS were highly correlated (mean Pearson's $r = 0.55$ and 0.69 respectively), and previous analyses implied a causal relationship from bite size to perception of mosquito attractiveness (Jones et al., 2017). As such, prediction using the combination of variants associated with these highly correlated phenotypes may yield stronger models of each. Indeed, PRS trained using the union of the three sets of markers (285 variants) resulted in minor improvement in model performances (mean increase in $r^2 = 0.04\%$), and highly correlated PRS (mean $r = 0.94$). Notwithstanding their modest predictive power, these PRS could be used to identify individuals who might benefit from protective strategies to avoid mosquitos.

2580W

Enrichment of deleterious variation on down-regulated haplotypes due to the regulation-dependent penetrance of coding variants. *R. Brown, S. Sankararaman.* University of California, Los Angeles, Los Angeles, CA.

Recent work with allele-specific expression data has shown that rare deleterious coding variants tend to lie on haplotypes with reduced gene expression (Castel, 2018); consistent with a model in which natural selection tends to reduce the frequency of pairs of haplotypes associated with higher penetrance at these variants. Here we study the population genetics of deleterious mutations whose penetrance is modulated by linked regulatory variation and empirically investigate the distribution of regulatory-linked enrichment for rare deleterious variation. We use simulations to understand the selection dynamics and steady-state population-level distributions of deleterious coding variation as a function of regulation and gene length. Simulations demonstrate that a single de novo deleterious variant is able to drift to a higher frequency and persists for longer when it occurs on down-regulated haplotypes. Since genes are subject to repeated random de novo mutation at multiple sites, the difference in selection efficacy results in down-regulated haplotypes carrying a greater number of deleterious variants than up-regulated haplotypes. The difference in the number of deleterious variants carried by the up- and down-regulated haplotypes increases with increasing gene length, mutation rate and eQTL effect. We show that these patterns persist in real data using 1000 Genomes phased-genotypes, eQTLs calls from GTEx and CADD scores for the predicted variant deleteriousness. We observe an enrichment of deleterious variation on down-regulated haplotypes on average across genes. Down-regulated haplotypes contain 1.21 (s.e. 0.036) alternate alleles per kb across all CADD scores while up-regulated haplotypes contain 1.10 (s.e. 0.024) indicating decreased tolerance of variation (paired t-test p-value = 4.9×10^{-6}). On average 47.2% of down-regulated haplotypes contain at least one variant at or exceeding a CADD score of 10 while only 46.2% of up-regulated haplotypes do, a marginally significant difference (paired t-test p-value = 0.042). These observations demonstrate that regulation affects the selection strength against deleterious variation in coding regions and results in an enrichment of deleterious variation on down-regulated haplotypes. Since most GWAS hits occur outside of coding regions, we further hypothesize that GWAS hits in regulatory regions may be tagging down-regulated haplotypes that are enriched for rare deleterious coding variants. .

2581T

Genome-wide association studies of human facial width-to-height ratios.

M. Lee¹, J. Roosenboom¹, E. Feingold², M.L. Marazita¹, J.R. Shaffer², S.M. Weinberg¹. 1) Center for Craniofacial and Dental Genetics, Department of Oral Biology, University of Pittsburgh, Pittsburgh, PA, USA; 2) Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA, USA.

The ratio of facial width to height is sexually dimorphic in humans and has been correlated with testosterone levels and related behavioral effects in numerous studies. Because most aspects of facial morphology are heritable, these types of quantitative facial traits are also likely to be influenced by genes. This has been borne out by GWAS, which have showed associations between many common genetic variants and normal-range variation in facial morphology. To better understand the genetic basis of facial width-to-height ratio, we calculated 5 facial indices based on various horizontal and vertical distance measures derived from either direct anthropometry or 3D facial surface images. We performed GWAS on these indices in 2,156 healthy unrelated individuals genotyped on the Illumina OmniExpress + Exome v1.2 array and imputed to the 1000 Genomes phase 3 reference panel. All measures were adjusted by sex, age and body size, then subjected to GWAS in PLINK adjusting for ancestry principal components. We also performed sub-analyses in males and females separately, and in post-pubertal individuals (>14y). We observed five genome-wide significant associations ($p < 5 \times 10^{-8}$). Two genes showed strong associations with the index involving total facial width by total facial height: *DNAH5* (rs1971860, 5p15.2; $p = 4.22 \times 10^{-8}$) in the over 14 subset and *DYNC2H1* (rs601446, $p = 4.17 \times 10^{-8}$) in the female subset. *WDR72* (rs141781189; $p = 2.94 \times 10^{-8}$) showed a significant association with indices specifically involving the upper face. Both *HLS-DOA* (rs409443; $p = 2.44 \times 10^{-8}$) and *CNTNAP4* (rs117438273; $p = 3.79 \times 10^{-8}$) were associated with the index involving the lower portion of the face. Additionally, several suggestive loci ($p < 5 \times 10^{-6}$) were observed. These results provide further insight into genetic pathways and mechanisms underlying normal variation in human facial morphology. Moreover, they suggest that traits commonly held to reflect the action of androgens during the human growth phase are also influenced by additional factors.

2582F

Low-frequency genetic variants contribute to human craniofacial morphology. D. Liu¹, M.K. Lee², J. Roosenboom², J.T. Hecht³, G.L. Wehby⁴, L.M. Moreno⁵, C.L. Heike⁶, M.L. Marazita², E. Feingold¹, S.M. Weinberg², J.R. Shaffer¹. 1) Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA, 15261; 2) Department of Oral Biology, University of Pittsburgh, Pittsburgh, PA, 15219; 3) Department of Pediatrics, University of Texas McGovern Medical Center, Houston, TX; 4) Department of Health Management and Policy, University of Iowa, Iowa City, IA; 5) Department of Orthodontics, University of Iowa, Iowa City, IA; 6) Department of Pediatrics, Seattle Children's Craniofacial Center, University of Washington, Seattle, WA.

The genetic basis of normal variation in human craniofacial traits is still poorly understood. Several GWAS and candidate genes studies have focused on identifying associations with common genetic variants, with moderate success to date. The influence of low-frequency variants has not been evaluated, and their contribution to facial morphology remains elusive. In order to better understand the genomic architecture of normal-range facial traits, we used burden-style tests to evaluate the role of low-frequency variants against 24 quantitative facial measurements derived from the 3D surface images. A cohort of 2447 healthy individuals of European ancestry were genotyped on Illumina OmniExpress+Exome v1.2 array. Two complementary statistical tests, Sequence Kernel Association Test (SKAT) and Combined Multivariate and Collapsing (CMC) method were performed using two ranges of minor allele frequency cutoffs (1% and 5%). A total of ~92,000 variants grouped into ~12,950 genes were included, and significance threshold was defined by Bonferroni method taking into account the number of phenotypes and the number of genes tested. We observed associations involving two genes. *GRAMD1B* was associated with the morphology of the lower lip (SKAT p-value = 6.34×10^{-12} , CMC p-value = 1.64×10^{-7}). There were 3 variants with MAF<1% in this gene and single variant linear regression showed a significant effect for a missense variant (rs191981781; p-value = 2.83×10^{-11} ; MAF=0.0008). *OR2C1* was associated with the height of the entire face (CMC p-value = 3.05×10^{-8} , SKAT p-value = 3.86×10^{-6}). There were 7 variants with MAF<5% and single variant analysis showed suggestive significance for one variant (rs62000975; p-value = 1.90×10^{-6}). *GRAMD1B* and *OR2C1* have not previously been implicated in human facial traits, and we consider them as potential candidate genes worthy of follow up in future studies. Our findings demonstrate that low-frequency variants contribute to normal-range human facial morphology.

2583W

The role of glycaemic, lipid, obesity and blood pressure risk factors as mediators of the effect of height on coronary artery disease and type 2 diabetes mellitus: A Mendelian randomisation study. E. Marouli¹, F. Del Greco M.², C.M. Astley^{3,6}, J. Yang^{7,8}, R.J.F. Loos⁴, Z. Kutalik⁵, J.N. Hirschhorn^{3,6}, P. Deloukas¹, *Genetic Investigation of Anthropometric Traits (GIANT) Consortium*. 1) William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK; 2) Institute for Biomedicine, Eurac Research, Affiliated Institute of the University of Lübeck, Bolzano, Italy; 3) Boston Children's Hospital, MA, USA; 4) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 5) Institute of Social and Preventive Medicine, Lausanne University Hospital, Lausanne, Switzerland; 6) Broad Institute of Harvard and MIT, Cambridge, MA, USA; 7) Institute for Molecular Bioscience, University of Queensland, Brisbane, Australia; 8) Queensland Brain Institute, The University of Queensland, Brisbane, Queensland, Australia.

Evidence from observational studies suggests that shorter stature is associated with cardiometabolic diseases. However, it is unclear whether this is a causal relationship or whether it is confounded by other factors. Using a Genetic Score for height (defined as adult stature) and Mendelian randomisation (MR), we investigated the causal effect of height on coronary artery disease (CAD) and type 2 diabetes (T2D). We also assessed whether these effects are mediated by known cardiometabolic risk factors (e.g. BMI, glycaemic traits, lipids, blood pressure) based on genetic evidence. We used UK Biobank (449,094 unrelated British individuals) to investigate by several MR analyses: (1) the causal effect of height on CAD (23,755 cases vs 425,339 controls) and T2D (29,427 cases vs 416,908 controls), (2) the causal effect of height on possible mediators (known cardiometabolic risk factors), and (3) the extent to which these factors mediate any effect that height may have on CAD and T2D. In parallel, we used publically available summary statistics data from large consortia (GIANT, CARDIOGRAMplusC4D, DIAGRAM, MAGIC, GLGC) to perform two-sample MR analyses. Two-sample MR analyses showed that one SD higher genetically determined height (~6.5 cm) was associated with decreased risk of CAD (OR: 0.83, 95% CI: 0.80-0.87). This association remained significant (OR: 0.86, 95% CI: 0.79-0.94) after performing sensitivity analyses. The height-CAD effect was reduced by 1-2%, in terms of magnitude, after adjustment for glycaemic (glucose, insulin, HbA1c, 2hGlu), lipid (Total Cholesterol, Low Density Lipoprotein, High Density Lipoprotein, Triglycerides) or blood pressure (BP; Systolic BP, Diastolic BP, Pulse Pressure) traits in Multivariate MR analyses. We observed a causal effect of height on T2D (OR: 0.93, 95% CI: 0.90-0.96, p=0.01), but it was entirely driven by its indirect effect through BMI (OR: 0.95, 95% CI: 0.92-0.99, p=0.12). There was no evidence for a causal relation between adult height and T2D adjusted for BMI (OR: 0.99, 95% CI: 0.96-1.02, p=0.95). In summary, one SD higher genetically determined height reduces CAD risk by 17% and the association does not seem to be mediated by glycaemic, lipid and blood pressure traits, suggesting the observational association is due to a causal effect. On the other hand, we observed no direct causal effect of height on T2D risk, once its effect on BMI is taken into account.

2584T

DNA methylation in nasal epithelial cells and sensitization to aeroallergens.

A. Morin¹, C. McKennan², K. Naughton¹, D. Vercelli³, D. Nicolae^{1,2,4}, K. Bønnelykke⁵, H. Bisgaard⁵, C. Ober¹, the COPSAC study. 1) Department of Human Genetics, University of Chicago, Chicago, IL; 2) Department of Statistics, University of Chicago, Chicago, IL; 3) Department of Cellular and Molecular Medicine, Arizona Center for the Biology of Complex Diseases, and Asthma and Airway Disease Research Center, University of Arizona, Tucson, AZ; 4) Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, IL; 5) Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital, University of Copenhagen, Copenhagen, Denmark.

DNA methylation is an epigenetic mark that responds to the environment and, therefore, can serve as a proxy for understanding how exposures influence gene expression and disease pathogenesis. Airway epithelia are in direct contact with inhaled exposures and are likely the site of interactions. In particular, sensitization to aeroallergens during early life impacts the development of childhood diseases such as asthma and allergies. To better understand the mechanism through which environmental exposures impact disease, we collected DNA from nasal epithelial cell brushings at age 6 from 517 children in the Copenhagen Prospective Studies on Asthma in Childhood (COPSAC)²⁰¹⁰ birth cohort. DNA methylation profiles were assessed using the Illumina 850k EPIC array. We used a linear model (limma) with sex and confounding variables identified by PCA as covariates to assess differential DNA methylation between children with and without sensitization to aeroallergens. Batch effects were adjusted using ComBat. We also removed additional unwanted variation by protecting the phenotype of interest using an in-house method (unpublished). We identified 20 CpGs associated with sensitization to aeroallergen (FDR 5%): 16 were less methylated and four were more methylated in sensitized children. Among these were 7 CpGs that were associated with atopic asthma or Immunoglobulin (Ig) E levels in previous studies, including cg04983687 (*ZFPM1* intron; $p=2.48e-14$; $q=1.94e-8$), cg20372759 (intergenic between *METTL1* and *CYP27B1*; $p=1.32e-10$; $q=3.44e-5$) and cg21220721 (*ACOT7* exon; $p=7.18e-10$; $q=1.12e-4$). Of the 13 newly identified CpGs, 9 are specific to the EPIC array, which was not used in any previous studies. The most significant association on the EPIC array was at cg16911809 ($p=2.35e-11$; $q=9.18e-6$), which is located in an intron of *CCL26*, a chemokine that has been previously associated with asthma and allergies in multiple studies. Furthermore, another CpG located in an intron of *CCL26* was differentially methylated in bronchial epithelial cells from asthmatic and non-asthmatic subjects. These data suggest a link between nasal epithelial cell DNA methylation and sensitization to aeroallergens. Ongoing analyses are exploring the impact of DNA methylation in on gene expression in the same samples. This work is supported by NIH R01 HL129735.

2585F

A characterization of eQTLs directions in complex traits inspired by the omnigenic model. O. Naret, J. Fellay. EPFL, Lausanne, Switzerland.

The genetic architecture of extreme polygenic traits and its biological underpinning remain open questions. As an extreme example of polygenicity, it was estimated through genome-wide association studies (GWAS) that up to 100,000 independent variants exert a causal effect on height. While some of the top signal variants could be mapped to meaningful biological pathways or features, most of them fall in the regulatory regions of supposedly irrelevant genes. This leaves the bulk of the biology of height heritability unexplained. Relying on the observation that any gene can be associated with a trait as long as it is expressed specifically or broadly in a relevant cell or tissue, Boyle et al. (2017) proposed the omnigenic model to classify them in two categories. First, a minority of core genes that are specific to the biology of the trait. Second, a majority of peripheral genes that interact with core genes through the small world property of the regulatory network. Under this model, we tested if the effect of peripheral genes could result from a competition for resources with core genes, at the level of transcription. Under this assumption, the peripheral genes direction of expression should be correlated with the direction of their association with the phenotype of interest (as expressed in the + or - sign of their effect size), while opposite correlations should be observed for the core genes. To test our model, we leveraged the gene expression data reported by the GTEx consortium in different tissues, and the summary statistics from UK Biobank GWAS for height, BMI, and depression. We did not observe any evidence of associations between the direction of RNA expression and the direction of the genetic association. The small world property of the regulatory networks can imply complex interactions that will be challenging to understand. To accept the omnigenic model, the causality of peripheral genes needs to be further characterized.

2586W

Genetic predisposition for myalgic encephalomyelitis/chronic fatigue syndrome: Pilot study.

L. Nathanson^{1,2}, **M. Perez**^{1,2}, **R. Jaudoo**^{1,3}, **T.J.A. Craddock**^{1,3}, **K. Gaunt**^{1,2}, **K. Gemayel**^{1,2}, **A. Del Alamo**^{1,2}, **M.A. Fletcher**^{1,2}, **N.G. Klimas**^{1,2}. 1) Institute for Neuro Immune Medicine, Nova Southeastern University, Fort Lauderdale, FL; 2) Kiran C Patel College of Osteopathic Medicine, Nova Southeastern University, Fort Lauderdale, FL; 3) College of Psychology, Nova Southeastern University.

Background: Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) is a debilitating disease with unknown causes. It is associated with a variety of symptoms including fatigue, memory problems, muscle and joint pain, gastrointestinal issues, neurological problems, hormonal imbalance and immune dysfunction. Currently, treatment relies solely on symptom management but does not address the underlying mechanisms of disease. It is known that single nucleotide polymorphisms (SNPs) play an important role in gene expression changes that can manifest as phenotypic changes. Prior to this ongoing study, there existed no known databases of SNPs in patients diagnosed with ME/CFS. **Objectives:** Our objectives are to identify possible genetic predisposition to ME/CFS. We are creating a novel database of SNPs that are specific for ME/CFS patients to identify the relative frequency in our cohort of specific SNPs warranting further research. **Methods:** A genetic database was created on-site through the use of a secure user-friendly online platform, REDCap, for participants to upload their raw genetic data, acquired from 23andMe. The uploaded de-identified genetic data acquired from RedCap have been modified and filtered to include only nonsense and non-synonymous SNPs from protein coding regions (exons), microRNAs, and SNPs that are close to splice sites. The frequencies of each SNP have been calculated within our cohort, compared to the public database (Kaviar) and those SNPs of differing prevalence are noted for future analysis. **Results:** Over 800 ME/CFS patients have been recruited for the study so far. Recruitment is ongoing, and new data are being collected. 383 ME/CFS patients uploaded their data to RedCap. Analysis of this data revealed 387 SNPs that have the frequency at least 10% in either healthy or ME/CFS patients cohort and have at least five-fold difference between healthy and ME/CFS patients cohorts. Pathway analysis showed that the most affected are cell signaling (including MAPK-, calcium-, Wnt-, cytokine-cytokine receptor signaling, etc.) and immune pathways. **Conclusion:** This data show possible genetic predisposition for ME/CFS. Studies of larger cohorts are needed.

2587T

Genome-wide association study identifies gene polymorphisms associated with the analgesic effect of fentanyl in the preoperative cold pressor-induced pain test.

D. Nishizawa¹, **K. Takahashi**^{1,2}, **K. Yoshida**^{1,2}, **S. Kasai**¹, **J. Hasegawa**¹, **K. Nakayama**¹, **Y. Ebata**¹, **Y. Koukita**², **K. Fukuda**³, **T. Ichinohe**², **K. Ikeda**¹. 1) Psychiatry and Behavioral Science (Addictive Substance Project), Tokyo Metropolitan Institute of Medical Science, Tokyo, Japan; 2) Department of Dental Anesthesiology, Tokyo Dental College, Tokyo, Japan; 3) Department of Oral Health & Clinical Science, Tokyo Dental College, Tokyo, Japan.

Purpose: Opioid analgesics are widely used for the treatment of moderate to severe pain. The analgesic effects of opioids are well known to vary among individuals. Such individual differences in opioid sensitivity may at least partly be attributable to genetic factors. The present study focused on the genetic factors that are associated with interindividual differences in opioid and pain sensitivity. **Methods:** We conducted a multistage genome-wide association study (GWAS) in subjects who were scheduled to undergo mandibular sagittal split ramus osteotomy and were not medicated until they received fentanyl for the induction of anesthesia. We preoperatively conducted the cold pressor-induced pain test before and after fentanyl administration. Whole-genome genotyping was performed with iScan System (Illumina, San Diego, CA, USA) and mainly two kinds of BeadChip, Human1M v1.0 and Human1M-Duo v3. **Results:** As a result of GWAS with approximately 300,000 polymorphic markers in surgical patients, the rs13093031 and rs12633508 single-nucleotide polymorphisms (SNPs) near the *LOC728432* gene region and rs6961071 SNP in the *tcag7.1213* gene region were significantly associated with the analgesic effect of fentanyl, based on differences in pain perception latency before and after fentanyl administration. The associations of these three SNPs that were identified in our exploratory study have not been previously reported. The two polymorphic loci (rs13093031 and rs12633508) were shown to be in strong linkage disequilibrium. Subjects with the G/G genotype of the rs13093031 and rs6961071 SNPs presented lower fentanyl-induced analgesia. However, no SNPs were identified as genome-wide significant in their associations with pain sensitivity. **Conclusions:** Our findings provide a basis for investigating genetics-based analgesic sensitivity and personalized pain control, in which the G alleles of the rs13093031 and rs6961071 SNPs are associated with lower opioid sensitivity.

2588F

Multiple signals at known loci revealed by analysis of C-reactive protein levels in the Trans-Omics for Precision Medicine (TOPMed) whole genome sequencing data. L.M. Raffield¹, L. Polfus², S. Aslibekyan³, E. Benjamin⁴, J.C. Bis⁵, R.P. Bowler⁶, B. Cade⁷, A. Correa⁸, J. Dupuis⁹, P. Durda¹⁰, N.S. Jenny¹⁰, B.J. Kral¹¹, M.G. Larson⁹, N. Pankratz¹², B.M. Psaty¹³, A.P. Reiner¹⁴, S. Salimi¹⁵, R.P. Tracy¹⁶, B. Wang¹⁷, L.P. Zhao¹⁸, P. Auer¹⁹, *Trans-Omics for Precision Medicine (TOPMed) Inflammation Working Group.* 1) Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Center for Genetic Epidemiology, Keck School of Medicine, Department of Preventative Medicine, University of Southern California, Los Angeles, CA; 3) Department of Epidemiology, University of Alabama, Birmingham, AL; 4) Framingham Heart Study, Framingham, MA; Department of Medicine, Section of Cardiovascular Medicine and Preventive Medicine, Boston University School of Medicine; Department of Epidemiology, Boston University School of Public Health, Boston, MA; 5) Cardiovascular Health Research Unit, University of Washington, Seattle, WA; 6) Department of Medicine, National Jewish Health, Denver, CO; 7) Division of Sleep and Circadian Disorders, Brigham and Women's Hospital, Division of Sleep Medicine, Harvard Medical School, Boston, MA; 8) Department of Pediatrics and Medicine, University of Mississippi Medical Center, Jackson, MS, USA; 9) Framingham Heart Study, Framingham, MA; Department of Biostatistics, Boston University School of Public Health, Boston, MA; 10) Department of Pathology & Laboratory Medicine, Larner College of Medicine at the University of Vermont, Burlington, VT; 11) GeneSTAR Research Program, Division of General Internal Medicine, Department of Medicine, The Johns Hopkins University School of Medicine; Division of Cardiology, Department of Medicine, The Johns Hopkins University School of Medicine, Baltimore, MD; 12) Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN; 13) Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology and Health Services, University of Washington, Seattle; and Kaiser Permanente Washington Health Research Institute, Seattle, WA; 14) Department of Epidemiology, University of Washington, Seattle, WA; 15) School of Medicine, Division of Endocrinology, Diabetes and Nutrition, and Program for Personalized and Genomic Medicine, University of Maryland School of Medicine, Baltimore, MD; 16) Department of Pathology & Laboratory Medicine, Department of Biochemistry, Larner College of Medicine at the University of Vermont, Burlington, VT; 17) Department of Biostatistics, Boston University School of Public Health, Boston, MA; 18) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA; 19) Zilber School of Public Health, University of Wisconsin-Milwaukee, Milwaukee, WI.

C-reactive protein (CRP) is a well-established biomarker of acute and chronic inflammation and a predictor of cardiovascular disease, immunity-related disorder, and mortality risk. The largest genome-wide association study meta-analysis in European populations (>80,000 subjects) identified eighteen loci; however, loci may contain multiple distinct signals with differing regulatory roles, and additional genetic determinants remain to be discovered in diverse ancestral populations. We performed a multi-ancestry whole genome sequence association analysis of single nucleotide and small insertion/deletion polymorphisms with C-reactive protein levels in 23,283 individuals (average age 59.2 years, 32% male; predominantly European (64.7%) and African American (28.1%) ancestry) from nine cohorts from the Trans-Omics for Precision Medicine (TOPMed) program. The analysis was performed on the University of Michigan ENCORE server using the Scalable and Accurate Implementation of Generalized mixed model (SAIGE) R package to control for sample relatedness. Inverse normalized CRP was adjusted for sex, age, study and self-reported ancestry indicator variables, as well as principal components derived from the sequence data. Analyses for variants with a minor allele frequency >0.1% (n=25,949,369) showed little evidence of inflation ($\lambda=1.045$). At a significance threshold of 1×10^{-9} , we identified 8 previously known loci (*CRP*, *APOE*, *HNF1A*, *LEPR*, *GCKR*, *IL6R*, *IL1F10*, *NLRP3*). Notably, six of these loci had different lead variants than those identified in previous studies; for example, the *APOE* locus was refined to rs429358 (a coding variant which defines *APOE*- ϵ 4 allele status) as opposed to noncoding variants which likely tagged this signal. The *CRP* locus demonstrated evidence of twelve conditionally distinct signals using TOPMed sequencing data and Genome-wide Complex Trait Analysis Conditional & joint (GCTA-COJO) analysis, including low frequency (minor allele frequency <5%, rs375458630, rs78069458) and non-European (e.g., functional DARC null red cell phenotype variant rs2814778, rs111245605) signals. We also found evidence for multiple distinct signals at *LEPR* and *HNF1A*. We are currently refining these association signals using functional annotation and fine-mapping approaches and seeking additional replication. For further discovery of CRP-related loci, larger sample sizes from diverse populations will be required.

2589W

The impact of genotype imputation on the statistical power of rare variant association tests. D.M.H. Tong¹, R.D. Hernandez^{1,2,3,4,5}. 1) University of California, Berkeley - University of California, San Francisco Graduate Program in Bioengineering, San Francisco, CA; 2) Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, CA; 3) Institute for Human Genetics, University of California, San Francisco, San Francisco, CA; 4) Quantitative Biosciences Institute, University of California, San Francisco, San Francisco, CA; 5) Institute for Computational Health Sciences, University of California, San Francisco, San Francisco, CA.

We are now entering the era of million-member genome-wide association studies (GWAS), with incredible success so far. Despite these successes, the common genetic variants associated thus far only explain a portion of the expected trait heritability. Rare variants, with minor allele frequencies below 1%, constitute the vast majority of human genetic variation discovered from sequencing studies, and population genetic modeling suggests that there are some scenarios in which rare variants make up a substantial source of that "missing heritability". Rare variant association tests (RVATs) were designed to make associations with these rare variants, but the anticipated explosion in rare variant associations has not happened. One likely reason is that the statistical properties of these RVATs are poorly understood across different potential genetic architectures and study designs. This limitation has hampered our ability to predict the power of RVAT study designs, and as such, we are attempting to use the same methods on different data blindly. Here, we present a simulation study to interrogate RVAT power across different genetic architectures and study designs, by simulating realistic studies using whole genome sequencing (WGS) and genotyping followed by imputation against massive imputation reference panels. We find that power to associate rare variants is low across all situations, that power is lowest when rare variants represent most of the trait heritability, and that large-scale genotyping followed by imputation results in only a slight decrease in power compared to whole genome sequencing. Our results highlight the fact that existing RVATs are severely under-powered, and that future RVATs will need to take into account the forces of natural selection and demography to improve our understanding of trait architectures.

2590T

Genetic sex differences are generated by embryonic developmental mechanisms. *M. Traglia, L.A. Weiss.* Department of Psychiatry and Institute for Human Genetics, University of California San Francisco, 401 Parnassus Ave, LangPorter, San Francisco, CA.

Sex differences in susceptibility, prevalence, symptoms as well as co-morbidities have been observed in a large set of complex disorders and quantitative phenotypes risk factors for common diseases. During puberty, evident physical differences in secondary sex characteristics emerge. We thus used available anthropometric traits to identify SNPs differentially associated by sex with height, weight, waist and hip measurements [GIANT, 2013]. We then showed that the set of anthropometric sex-heterogeneous SNPs (AH SNPs) strongly contribute to the genetic architecture of complex traits and disorders including autism, WTCCC-studied diseases, and GIANT consortia traits [Mitra et al., 2016; Traglia et al 2017]. This evidence suggests that observed disease differences by sex might be generated by the same pleiotropic mechanisms determining secondary sex characteristics, and we therefore wanted to better-define the relevant mechanisms. Based on the observation of a small but significant overlap between AH SNPs and androgen and estrogen-responsive genes (empirical $P < 0.01$), we hypothesized enrichment in functional elements. We thus annotated 8,140 SNPs with the nearest genes using ANNOVAR [Wang et al, 2010] and investigated whether the 1,367 independent AH genes were overrepresented in KEGG and Reactome pathways, GO biological processes and molecular functions ($FDR < 0.05$) using PANTHER [Mi, et al 2017]. To assess significance, we generated 100 permuted SNP lists (matching the number of AH SNPs and the association P in sex-combined analysis) and we calculated an empirical P . After multiple test correction and permutation, we observed significant overrepresentation of these genes in crucial signaling pathways for early body development such as WNT and cadherin signaling, angiogenesis, axon guidance mediated by netrin and gonadotropin-releasing hormone receptor pathways (emp $P < 0.05$). These genes and pathways are mainly implicated in ectoderm (emp $P = 0.03$) and nervous system development (emp $P = 0.04$). During gastrulation, the ectoderm differentiates into the nervous system and into the sensory epithelia and mammary glands that are all significantly enriched in AH genes (emp $P < 0.01$). The observation that fundamental prenatal development processes, responsible for the biology of traits and diseases later in life, are enriched in sexually dimorphic key genes and pathways, demonstrate the essential link between genetic sex differences and pathophysiology.

2591F

The “avocado gene” and nutrigenetics: New evidence for olfaction’s role in food preferences and genetic correlations with health outcomes. *J.F. Wilson, E.J. Grzeszkowiak, N. Pirastu.* University of Edinburgh, Edinburgh, United Kingdom.

It is estimated that diet is the second strongest risk factor for early death, influencing the incidence of the top killers in developed societies, including cardiovascular disease, cancer, diabetes, and dementia as it contributes to at least 678,000 deaths in the US each year. Nutrigenetics offers a unique opportunity for estimating dietary behaviour from individual genetic makeup and assessing its impact on disease risk. The conventional approach focusses on studying self-reported food consumption and converting specific food items to macronutrients. It has been suggested that food preferences may better predict actual food consumption than food frequency questionnaires. After measuring preferences for >60 different food and drink items in four different European cohorts ($n = 4,781$) on a 9-point scale, we ran GWAMA on each of the studied traits. We detected a number of genome wide significant hits, some of which were population specific. For example, we detected a novel significant association between the olfactory receptor gene *OR2L13* and avocado liking. We also show a range of novel associations for specific foods that have been previously linked to diet-related diseases such as the association between liking of acidic food and *PTPRN2* (related to obesity and colorectal cancer). In addition, we replicated some of the associations for food preferences from previous studies (Pirastu 2016). Furthermore, we calculated the genetic correlations between specific foods and diet-related traits which reveal a previously under-studied genetic link between food-liking and health outcomes. Our results strongly suggest that studying food preferences rather than macronutrient intake increases the chance of detecting genetic signals that may influence food consumption. Despite a relatively small sample size, our study has produced more hits than larger studies on macronutrient intake. We have also demonstrated that olfaction and the genetic variation within olfactory receptor genes may play an essential role in determining food preferences and consequently influencing health. Given the genetic correlations with health outcomes and overlap with the associations for diet-related diseases we conclude that in the future nutrigenetics may play an increasingly important role in personalised nutrition and disease prevention. Further studies with larger sample sizes and additional food and drink items are needed to further explore this under-studied area of human genetics.

2592W

Re-examining high altitude adaptation (HAA) associated genes in tibetan populations with a focus on the evolutionary history of EPAS1 gene. X. Zhang¹, K. Witt Dillon¹, X. Jin², A. Asan², R. Nielsen³, E. Huerta-Sanchez¹. 1) School of Natural Sciences, University of California, Merced; Merced, CA, USA 95343; 2) BGI-Shenzhen; Shenzhen, China 518083; 3) Department of Integrative Biology, University of California, Berkeley; Berkeley, CA, USA 94720.

Tibetan population is one of the three human populations that is well adapted to high altitude hypoxic environment. A number of candidate genes, including EPAS1, were previously reported to be strongly associated in the physiological responses leading to high altitude adaptation (HAA). In addition, the selected EPAS1 haplotype showed evidence of being adaptively introgressed from admixture event between modern humans and Denisovan-like archaic human groups (Huerta-Sanchez et al. 2014). However, many important questions remain to be answered. For example, there was limited evidence to convincingly exclude Neanderthals as a potential source of the beneficial EPAS1 haplotype. Besides, a recent study showed that two waves of introgression with Denisovans likely to have occurred in East Asia (Browning et al. 2018) - which event the EPAS1 haplotype was associated with requires further analysis. In this study, DNA samples of 148 individuals from Tibet were obtained, with altitude ranging from 3000 to 4800 meters above sea level. 23 HAA candidate genes were sequenced with 30x depth for the continuous core segment (> 60kb), plus 20 1kb-long segments that were evenly spaced every 20kb upstream and downstream of each gene (Yi et al. 2010; Beall et al. 2010, Simonson et al. 2010, and Bigham et al. 2010). By comparing the *Fst* between Tibetans and Han Chinese (CHB, 1000 Genomes Project), we identified one highly differentiated locus upstream of the EPAS1 core region (*Fst* = 0.8; intergenic variant). By comparing the pairwise differences between the most-common Tibetan haplotype with archaic individuals, we replicated the previous reports that the most common Tibetan EPAS1 haplotype was most similar to Denisovans than to Han Chinese. The difference of EPAS1 haplotype between Tibetans and Denisovans was larger than the Neanderthal-Neanderthal genome-wide divergence, but smaller than the divergence between Denisovan-Neanderthals. Therefore, we showed that the source of EPAS1 beneficial haplotype in Tibetans was most likely to be Denisovans instead of Neanderthals. By applying the SPrime method (Browning et al. 2018), we showed that the source of EPAS1 beneficial haplotype in Tibetans was likely to have originated from the East Asian-unique Denisovan introgression event (second wave).

2593T

A genome-wide analysis in consanguineous families reveals chromosomal loci in specific language impairment (SLI). M.H. Raza¹, E.M. Andres¹, H. Hafeez², A. Yousaf^{1,4}, S. Riazuddin³, M.L. Rice¹, M.A.R. Basra². 1) Child Language Doctoral Program, The University of Kansas, Lawrence, KS; 2) Institute of Chemistry, University of the Punjab, Lahore Pakistan; 3) Allama Iqbal Medical College, Lahore Pakistan; 4) Department of Biotechnology, International Islamic University, Islamabad Pakistan.

Specific language impairment (SLI) is a developmental disorder that delays language skills in children who are otherwise typically developing. Children with SLI experience lifelong impacts in academic learning and social relationships. Genetic studies have identified a number of chromosomal loci and several candidate genes for SLI. Although these are important findings in the genetics of SLI, they explain a small proportion of the genetic variance. Additional genetic studies are needed in extended SLI families. We performed genome-wide parametric linkage analysis and homozygosity mapping in 14 consanguineous families from Pakistan segregating SLI. Linkage analysis revealed a multi-point LOD score of 4.18 at chromosome 2q in family PKSLI05 under a recessive mode of inheritance. Another linkage with a multipoint LOD score of 3.85 was observed in family PKSLI12 at a non-overlapping locus on chromosome 2q. Two other suggestive linkage loci were found in family PKSLI05 on chromosome 14q and 22q with LOD scores of 2.37 and 2.23, respectively. These loci were also replicated in homozygosity mapping. Reduction to homozygosity was also observed in regions on chromosomes 2q, 14q, 22q, 5p, 8q and 17q. These homozygosity regions occurred in affected individuals of PKSLI families but were absent in most of the unaffected or unknown individuals. Our findings confirm previously analyzed suggestive SLI linkage loci on chromosomes 5, 14, 17 and 22 reported in the Robinson Crusoe Island founder population. These findings indicate that linkage and homozygosity mapping in consanguineous families can improve genetic analyses in SLI, and suggest the involvement of additional candidate genes for this disorder.

2594F

Programmable nuclease-based integration into novel extragenic genomic safe harbor identifies from Korean-population based CNV analysis. S. Moon¹, E-S. Lee², Y.J Kim¹, M.Y. Hwang¹, H. Kim³, B-J. Kim¹. 1) Korea National Institute of Health, Cheongju-si, Chungcheongbuk-do, Korea, Republic of; 2) Department of Chemical and Biological Engineering, Seoul National University; 3) Department of Pharmacology, Yonsei University.

One of the critical needs in biomedical field is the ability to stably insert functional transgenes and other genetic elements into human genome without the disrupting genes or perturbing their transcription. Several diseases have been successfully treated with stable insertion of therapeutic genes such as Parkinson's disease. However, random integration of transgenes can cause insertional mutagenesis altering the expression levels of neighbor genes. The most common approach to integrate transgenes in human cells is using retro-viral vector. Uncontrolled insertional mutagenesis can cause cancer such as leukemia and lymphoma by perturbing proto-oncogenes or tumor suppressors. Instead of using viral vectors, the use of engineered nucleases, Zinc finger Nucleases, Transcription Activator-Like Effector nucleases and CRISPR-Cas system, has become the alternative approach. Despite of the advance in these engineered nucleases, there are only few up-to-date identified and validated genomic safe harbors (GSHs). GSHs are intra/extragenic regions that can support predictable transgene expression while minimizing the neighboring gene perturbation. AAVS1 site on chromosome 19 is the most popular GSH due to its ability to support transgene expression in multiple cell types, yet at the same time it was known that AAVS1 locus can be silenced by the mechanism including DNA methylation. Here, we discovered GSH candidates and examined disease correlation. Moreover, we investigated the neighboring gene perturbation near the site of integration. To these ends, we used 3,601 multi-class copy number variation (CNV) regions from the previous CNV study using 4,694 Koreans. We examined CNV regions which satisfy the following criteria: (1) 5-25% minor allele frequency; (2) distance of >300Kb away from any gene and genes that are implicated in cancer and miRNA; (3) lack of disease correlation. As a result, we selected two candidates on chromosome 3 and 8 which satisfied selection criteria. In particular, based on cohort epidemiological data, we examined the disease correlation using 31 traits related to complex diseases including type 2 diabetes and obesity. Then, we found lack of disease correlations with two candidates. Finally, we were able to achieve efficient site-specific integration and the neighboring gene perturbation near integration site. We proved that these two CNV regions could derive robust transgene expression without significantly altering neighboring genes.

2595W

Taiwan Biobank: An open access resource for conducting the biomedical research. H. Chu^{1,2}, C. Shen^{1,2}. 1) Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan; 2) Taiwan Biobank, Academia Sinica, Taipei, Taiwan.

To understand the cause for common disease is challenging. These conditions are caused by a combination of genetics, environmental exposure, and lifestyle. In order to clarify these complex interactions, the Taiwan Biobank (TWB) is establishing a scientific infrastructure accessible to biomedical researchers. TWB is a large and comprehensive prospective study with 200,000 participants, who age 30–70 years, from the general population without history of cancer, and another 100,000 patients with chronic diseases of public health importance. Our aim is to improve the health of future generations. This study has collected extensive phenotypic and genotypic details from individual participants, including the data from questionnaires, physical examination, biomedical test, image inspection, experimental information, biospecimens, and longitudinal follow-up for a wide range of health-related outcomes. This huge and challenging effort is to elucidate risk factors and molecular pathogenesis of disease. This effort will lead to both disease prevention, progression prediction and facilitation of the development of individualized therapies. In order to study the variance of molecular markers in a large number of participants, TWB aims to generate multi-omics experimental information to study the genome, epigenome, transcriptome, proteome and metabolome from a large number biospecimens of participants. In genomics, TWB has already accomplished 28,000 whole genome genotypings, 2,000 whole genome sequencings, and 1,100 NGS-based HLA typings. TWB has also done 1,500 DNA methylation chips in epigenomics and 400 NMR-based metabolites identifications in metabolomics. The genomic summary statistics have been released through a publicly accessible web-based calculation platform, Taiwan View (<https://taiwanview.twbiobank.org.tw/>). TWB is available for open access to any bona fide researchers who wish to use it to conduct research. Additionally, most of participants have provided inform consent to authorize Taiwan Biobank to link their electronic health records. The release of this large-scale population data, omics information, and electronic health records will greatly benefit human biomedical research.

2596T

Association analyses of common SNPs and indels from whole genome sequencing of orofacial cleft (OFC) trios of European ancestry. *N. Mukhopadhyay¹, E. Feingold², E.J. Leslie³, G.L. Wehby⁴, L.M. Moreno-Urbe⁵, J.T. Hecht⁶, F.W.B. Deleyiannis⁷, S.M. Weinberg^{1,2}, T.H. Beaty⁸, J.C. Murray⁹, M.L. Marazita^{1,2}.* 1) Center for Craniofacial and Dental Genetics, Department of Oral Biology, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA; 2) Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA; 3) Department of Human Genetics, Emory University, Emory, GA; 4) Department of Health Management and Policy, University of Iowa, Iowa City; 5) Department of Orthodontics & the Iowa Institute for Oral Health Research, University of Iowa, IA; 6) Department of Pediatrics, McGovern Medical School and School of Dentistry UT Health at Houston, Houston, TX; 7) Department of Surgery, Plastic and Reconstructive Surgery, University of Colorado School of Medicine, Denver, CO; 8) Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; 9) Department of Pediatrics, University of Iowa, Iowa City, IA.

Orofacial clefts (OFCs) represent the most common type of craniofacial birth defects in humans, and they have a complex and heterogeneous etiology with multiple genetic and environmental risk factors. Through the NIH Gabriella Miller Kids First Pediatric Research Program (Kids First), whole genome sequencing (>30x) was done for 415 nonsyndromic OFC trios of European ancestry, drawn from an ongoing study of OFC families. Cumulatively, this research has included molecular, candidate gene, whole-genome linkage and whole-genome association studies, and has identified several genes controlling risk to OFC. Here, we present results from whole-genome sequencing (WGS) of 327 case-parent trios using approximately 4×10^6 common biallelic single nucleotide polymorphisms (SNPs) and indels with high call quality. All cases had cleft lip only (CL), cleft palate only (CP), or cleft lip with cleft palate (CLP). Tests of association and linkage were assessed for each SNP/indel using the allelic transmission disequilibrium test (TDT). A subset of the trios affected with CL or CLP were also analyzed together (CL/P subgroup). Outcomes include a genome-wide significant association at 8q24 ($p < 10^{-9}$) and near-significant association for markers at *PAX7* ($p < 10^{-7}$) in the CL/P subgroup; both of which have been reported in earlier studies. Suggestive associations ($p < 10^{-6}$) at other recognized risk genes were also observed in the complete set of OFC trios, including the *MAPT* and *CRHR1* genes on chr17q21.31 (Park et al. 2006), and *ARHGAP8* on chromosome 22q13.31 (Ludwig et al. 2012). A new genome-wide significant association was observed in chromosome 12q13.13 (lead SNP rs7312514, $p < 10^{-9}$), and a second locus approaching significance on chromosome 6q16.3 (lead SNP rs111364373, $p \times 10^{-9}$) in the complete set of OFC trios. This preliminary investigation of Kids First WGS data identified several genomic regions associated with risk to OFCs, some recognized and some novel. Detailed investigation to identify novel genes in the new loci, as well as causal variants and/or regulatory pathways for all putative risk genes is currently underway. This work was supported by grants R01-DE016148, X01-HL132363, R01-DE014581, R01-DD000295 and R03-DE026469.

2597F

Finding subgroups of Chronic Obstructive Pulmonary Disease (COPD) in a diverse population using clinical data from Electronic Health Records. *A.O. Basile¹, A. Frase², A. Lucas², M.D. Ritchie², N.P. Tatonetti¹.* 1) Biomedical Informatics, Columbia University, New York, NY; 2) Department of Genetics, Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, PA.

Complex diseases are heterogeneous in nature; instead of representing a single disease they may be comprised of multiple disorders with varying symptoms, clinical presentations, and etiology. Trait heterogeneity is a confounding issue largely overlooked in traditional genetic approaches. It complicates patient diagnoses, and has been shown to decrease both statistical power to detect associations and the risk attributed to susceptibility variants. The copious amounts of biomedical data present in Electronic Health Record (EHR) systems can be leveraged using data-driven approaches to overcome trait heterogeneity and detect homogeneous patient subsets. Similarity Network Fusion with spectral clustering was used to group 57,574 COPD patients of mixed race (20% Hispanic, 11% African and 37% European American) from Columbia University/NY Presbyterian Hospital using EHR extracted clinical lab measures, COPD comorbidities, and demographics. Analysis on 19,627 females identified 2 clusters, with cluster 1 characterized by older patients (mean age 87) who more frequently exhibit renal (33% chronic kidney disease (CKD)) and cardiac (40%) traits, and cluster 2 defined by younger subjects (mean 63) with respiratory traits (54% asthmatic). Analysis on 18,379 males identified 3 clusters. Cluster 1 is composed of younger patients (mean 33) with higher liver enzymes and more frequent asthma diagnoses (44%). Cluster 2 is defined by older patients (mean 84) who exhibit renal and cardiac (46% cases) traits. Cluster 3 is defined by patients with a mean age of 69 and a higher bronchitis diagnosis (43%). For both sexes, the aforementioned traits showed statistically significant (Bonferroni $p < 5 \times 10^{-4}$) differences between clusters when assessed with Wilcoxon or ANOVA and chi-squared tests. Recursive feature elimination with a random forest classifier performed separately on each sex identified similar top predictors in both cohorts including age, BUN, and diagnosis of CKD, asthma, and heart disease. Also, while the number of clusters differed, the sex stratified analyses identified groups with similar clinical characteristics: younger asthmatic patients versus older patients with renal and cardiac traits. Likewise, variables such as chronic cough, HDL, and obesity had consistent patterns without significant differences between the clusters for both sexes. This analysis successfully utilizes EHR-extracted data to elucidate clinical patterns and address heterogeneity in COPD.

2598W**Gene expression and splicing QTLs inform the biology of complex traits.**

R. Bonazzola, GTEx Consortium. Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, Illinois, USA.

Huge progress has been made uncovering thousands of genetic variants robustly associated with complex traits. Most of these are non-coding and are thought to regulate mRNA levels and splicing, whose cascading effects alter phenotypes. We report the analysis of the V8 release of GTEx with 838 whole genomes and 17,382 RNA-sequenced samples from 49 tissues. After uniform harmonization of thousands of GWAS results covering diverse human phenotypes and imputation to a uniform variant set in hg38, we performed colocalization and PrediXcan analyses to identify xQTLs (expression and splicing) that underlie complex trait mechanisms. Here we identify candidate genes at more than half of the examined GWAS loci. We demonstrate that disease-associated genetic variation frequently perturbs distal gene regulation and is often associated with the expression of more than one gene. The unprecedented depth of the GTEx catalog enabled the identification of conditionally independent xQTL signals at a large fraction of human genes. We also find that secondary xQTLs are frequently co-localized with GWAS signals even when the primary signal is not. Moreover, we demonstrate that top xQTLs tend to have smaller GWAS effect sizes than secondary xQTLs. In contrast, jointly finemapped eQTLs (DAP-G) ranked by their effects on expression tend to have more significant downstream effects. These results suggest that the current practice of focusing on top xQTLs should be discouraged. Gene-level analysis with PrediXcan replicated many known disease genes while dramatically expanding the catalog of genes with roles in complex trait biology. We replicate our candidate causal complex trait genes with a test based on an enrichment of rare, loss-of-function coding mutations in genes with related phenotypes, demonstrating that many human phenotypes are affected by an allelic series, where common variation that affects expression has a milder range of related phenotypic effects relative to mendelian mutations. As expected from the widespread sharing of xQTLs across tissues, multi-tissue analysis allows for more discoveries than any single tissue. We find tissue-shared and tissue-specific gene-level association with PrediXcan, suggesting intricate patterns of context specificity of the etiology of complex traits. Given the extensive coverage of human tissues and phenome-wide scope of our analysis, we expect that findings of our analysis will shape the guidelines for future GWAS-xQTL analysis.

2599T**Pedigree-focused linkage analysis of specific language impairment reveals multiple suggestive loci.**

E.M. Andres, M.L. Rice, M.H. Raza. Child Language Doctoral Program, University of Kansas, Lawrence, KS.

Specific language impairment (SLI) is described as a deficit in language despite average or above average nonverbal intelligence (IQ) and no other behavioral diagnoses or hearing loss. The estimated prevalence of SLI is 7%. The causes of SLI are not well understood, but family aggregation studies provide evidence that genetic factors are involved in the transmission of SLI. The current study used pedigree-based parametric linkage analysis of six SLI families that are followed longitudinally for language measures. We performed whole genome SNP genotyping using Illumina Infinum QC Array-24 and genotyping of microsatellite markers within the identified linkage loci. Linkage analysis was completed under autosomal dominant and recessive modes of inheritance with variable disease penetrance. Performance on an age appropriate standardized omnibus language measure was used to assign affected status categorically for all individuals. Single-point LOD scores above 2.0 were identified on chromosomes 4q (family 1), 3p, 9q, 15q (family 2), 6q, 10q, 12p (family 4), and 14q (family 6). The highest LOD scores were consistently found under the recessive mode of inheritance, with the exception of one locus (3p in family 2). Multipoint linkage analysis, under recessive mode of inheritance with full penetrance, showed maximum LOD scores of 2.41 on chromosome 4q, 2.39 on 14q, and a maximum LOD score of 2.65 on 6q, 10q and 12p (all found in family 4). The large size of family 2 does not allow multipoint linkage analysis for the loci on chromosomes 3p, 9q and 15q. These results indicate genetic heterogeneity of SLI and provide support for the previous hypotheses that SLI is polyphenic and polygenic. Further investigation of these loci in other populations may provide additional support for these candidate SLI loci.

2600F

Gene-smoking interaction in Fuchs endothelial corneal dystrophy:

A GWAS analysis. Y.J. Li^{1,2}, R.P. Igo³, N.A. Afshari⁴, S.K. Iyengar², *Fuchs Genetics Consortium*. 1) Biostatistics and Bioinformatics, Duke University Medical Center, Durham, NC; 2) Duke Molecular Physiology Institute, Duke University Medical Center, Durham, NC; 3) Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, Ohio; 4) Shiley Eye Institute, University of California, La Jolla, California.

Fuchs endothelial corneal dystrophy (FECD) is a common ocular disease causing corneal edema, decreased vision, and eventually leading to transplantation. Genetic predisposition is considered the best predictor for FECD with four validated loci (*TCF4*, *LAMC1*, *KANK4* and *LINC00870/ATP1B1*), as reported in our genome wide association study (GWAS) (Afshari et al. 2017. *Nat Commun* 8, 14898). Environmental exposures, UV light and smoking, have been linked to FECD as risk factors, but their role in disease etiology is not well understood. Herein, we focus on the role of smoking (having ever smoked) and the gene-smoking interaction on FECD. We also evaluate the impact of smoking on the four known FECD genes. Genotyping was performed on 3968 samples (1404 FECD cases, 2564 controls) using Illumina HumanOmni2.5-4v1_H array by the Center for Disease Research; Of these, we analyzed 1294 cases and 2317 controls with smoking data for ~1.5 Million variants (MAF > 0.02). Multivariable logistic regression was used to test the effect of smoking on FECD while controlling for age and sex. We then conducted genome-wide SNP-smoking interaction tests with adjustment of SNP, smoking, age, sex, and six principal components. For the four known FECD gene regions, we repeated the same regression model as the one in our prior GWAS except adding the smoking factor. By itself smoking showed significant protective effects on FECD after adjusting for age and sex (OR=0.54, 95% CI= (0.47, 0.63), $p < 0.0001$). GWAS of SNP $\times$ smoking revealed 7 SNPs meeting $p < 10^{-5}$. Among them, three common SNPs (MAF=0.45-0.5) in *CD274*, an immune checkpoint protein, revealed strong SNP $\times$ smoking interaction effects on FECD. SNP rs7042084 showed the most significant SNP $\times$ smoking, with the nucleotide A $\times$ smoking increasing the risk of FECD (OR=1.72, 95%CI = (1.39, 2.14), $p=8.3 \times 10^{-7}$) and the main effects of both SNP and smoking are also significant ($p=0.0009$ and 3.5×10^{-17} , respectively). However, the other two SNPs in *CD274* showed significant SNP $\times$ smoking (max $p=8.0 \times 10^{-6}$) but no significant main effects. Variants at four known FECD genes remain significantly associated with the disease after adjusting for smoking factor. As the SNP $\times$ smoking in these four genes shows only nominal association with FECD, we confirm that their effects on FECD are primarily driven via the main effect. In conclusion, this will be the first report on genes associated with FECD modifiable through one's life style, such as smoking.

2601W

Circadian misalignment accentuates genetic risk to obesity. A.R. Wood¹, S.E. Jones¹, Z. Kutalik², H. Yaghootkar¹, R.N. Beaumont¹, M.A. Tuke¹, K.S. Ruth¹, R.M. Freathy¹, A. Murray¹, J.T. Tyrrell¹, T.M. Frayling¹, M.N. Weedon¹. 1) Genetics of Complex Traits, University of Exeter Medical School, Exeter, Devon, United Kingdom; 2) Institute of Social and Preventive Medicine, University Hospital of Lausanne, Switzerland.

Introduction Gene-lifestyle interactions in humans have been difficult to identify and replicate. Some putative interactions have been observed between self-report behaviours (e.g. sleep) and genetic variants associated with complex traits, but most are not validated with objective measures. Using individuals from the UK Biobank, we tested the hypothesis that misalignment between genetic estimates of circadian timing, based on self-reported chronotype ("morningness"), and accelerometer-derived estimates of circadian timing (activity and sleep timing) would increase susceptibility to obesity. **Methods** We derived objective measures of circadian timing using 82,322 unrelated European individuals from the UK Biobank. These measures included timing of the least and most active periods of the day, and a relative amplitude of these measures to estimate rest-activity rhythms. We tested for interaction effects on BMI between the genetics associated with chronotype as a genetic risk score (GRS) and derived measures of circadian misalignment. Misalignment was defined as intra-individual disagreement between the GRS and accelerometer-based timing measures. For example, individuals were classified as having a misaligned circadian timing if they were in the top 20% for the GRS and bottom 20% for the accelerometer-derived estimates of circadian timing and vice versa. **Results** Circadian misalignment was associated with higher obesity risk. For example, 16% of individuals with a high or low genetic predisposition to being a "morning" person were misaligned to estimates of sleep timing. Misaligned individuals were 1.78kg heavier (for individuals 1.75m tall) than aligned individuals. We observed an association between the chronotype GRS and higher BMI in the misaligned group ($\beta=0.58\text{kg}$, 95% CI 0.26,0.88, $P=3 \times 10^{-4}$). We did not observe the same association within the aligned group ($\beta=-3.15\text{kg}$; 95% CI -6.68,0.37; $P=0.08$). Interaction analyses of the chronotype GRS and circadian misalignment measures provided evidence of accentuated effects on BMI ($P_{\text{int}}=5 \times 10^{-6}$). **Conclusions** Using genetics and accelerometer data, we provide evidence that circadian misalignment accentuates obesity risk. However, individuals misaligned were more overweight at baseline and further analyses are needed to test the robustness of these interactions including negative control experiments. Our results highlight the potential of identifying gene-lifestyle interactions when using objective measures of lifestyle.

2602T

Identifying disease subtypes from genotype data. *H. Aschard, A. Vaysse, A. Bzili.* Centre de Bioinformatique, Biostatistique et Biologie Intégrative (C3BI), Institut Pasteur, Paris, France.

Several common diseases are not single entity but rather composed of multiple heterogeneous disease subtypes. For example, inflammatory bowel disease (IBD) encompasses Crohn's disease and ulcerative colitis, while breast cancer can be either estrogen receptor (ER)-positive or ER-negative. Subtypes are expected for many other diseases, but have not yet been characterized. Identifying such latent disease subtypes is important both for understanding the etiology of diseases, and for improving clinical treatment. Because known disease subtypes display genetic heterogeneity, it has been suggested that genetic data can be used to infer latent disease subtypes. More specifically, previous works proposed applying principal component analysis of genotypes in disease cases, and using the top principal component as a classifier to identify subtypes. While achievable in theory, this approach faces several issues making it of limited utility in existing data. The most important one is a very low power, so that extremely large sample size is required to achieve sufficient discriminatory ability. Here we show first that sample size necessary to identify subtypes can be substantially decreased by cleverly pre-selecting variants with the largest differential effects between subtypes. Consider for example a disease with two sub-types sharing half of their causal variants (the remaining causals being subtype-specific). Our simulations shows that a predictive model using only the subtype-specific variants can achieve high discriminatory ability (e.g. AUC>0.80) in situations where the naive approach using all available genetic variants would have almost no discriminatory power (AUC < 0.55). We further explored various strategies and metrics to pre-select these subtype-specific variants. Our analyses show that they can be partly identified by comparing the observed contingency table of genotypes in cases and controls against the expected one under the assumption of homogeneity of genetic effect (i.e. in the absence of disease subtypes). Thus, our approach has the potential to dramatically reduce the required sample size to achieve detection of subtypes, making application in existing cohorts possible.

2603F

Identifying genetic architecture of facial morphology in Uyghurs. *H. Chen¹, Y. Huang¹, J. Zhang², L. Ma¹, Q. Liu¹, Y. Liu¹, H. Wu¹, K. Xiang¹, W. Yuan¹, C. Shi¹, H. Shi³, B. Su⁴, J. Xie².* 1) Beijing Institute of Genomics, CAS, Beijing, China; 2) Shihezi Medical School, Shihezi, Xinjiang, China; 3) Kunming School of Science and Technology, Kunming, China; 4) Kunming Institute of Zoology, CAS, Kunming, China.

Human facial morphogenesis is generated synthetically by craniofacial, skin and muscle developmental processes. Twin studies found that the heritability of facial features is around 80%. In this study, we conducted a genome-wide association analysis of facial morphological phenotypes measured from 3D facial images of 1,400 Uyghur individuals. Based on seventeen landmarks of facial 3D images, we used four schemes to efficiently characterize the facial morphological variations: traditional inter-landmark distances, geometric measurements of landmarks, principal components of clustered inter-landmark distances, and partial generalized procrustes analysis of landmarks. We identified ten genome-wide significant SNPs of seven genes: *DRD1*, *CNBD1*, *SBF2*, *SOX9*, *ALG10B*, *EDAR* and *CASC17*. In particular, deficiency of *SOX9* can lead to the skeletal malformation syndrome, campomelic dysplasia and cleft palate. In addition, we also found that several pleiotropic SNPs, e.g., rs3827760 in the exonic region of *EDAR*, are associated with facial and multiple other phenotypes. Our study thus provides insights into the genetic mechanisms of facial morphogenesis.

2604W

New genetic loci associated with permanent tooth eruption. *F. Geller¹, J. Fadista¹, E.A. Nøhr², J. Bybjerg-Grauholm³, D.M. Hougaard³, S. Kreiborg⁴, B. Feenstra¹, M. Melbye^{1,5,6}*. 1) Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark; 2) Research Unit of Gynaecology & Obstetrics, Institute of Clinical Research, University of Southern Denmark, Odense; 3) Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark; 4) Institute of Odontology, Section of Pediatric Dentistry and Clinical Genetics, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark; 5) Department of Clinical Medicine, University of Copenhagen, Denmark; 6) Department of Medicine, Stanford University School of Medicine, California, USA.

The timing of permanent tooth eruption is highly heritable and can have orthodontic consequences even later in life. We extended our genome-wide association study of permanent tooth eruption to 14,000 children with dental records between age 6 and 14 years. Age-adjusted standard deviation scores averaged over multiple time points were calculated for both sexes and then analyzed jointly as quantitative trait on dosages of 7.3 million variants imputed from Haplotype Reference Consortium release 1.1 reference panel. We identified 11 new loci at genome-wide significance ($P < 5 \times 10^{-8}$). Two loci (*BMP4*, *RAD51L1*) provided further evidence that there is substantial overlap between genetic variants underlying primary and permanent tooth eruption. Another locus was previously reported for adult height (*PEX2*), also showing an effect on infant length. Further signals were in the vicinity of genes playing a role in growth and development (*VEGFA*, *IRX5*, *FOXF1*). By performing a genome-wide association study on a large group of individuals with extensive childhood dental records, we were able to substantially increase the number of known loci for tooth eruption and these genomic regions harbor several candidate genes for functional follow-up.

2605T

GWAS meta-analysis highlights the hypothalamic–pituitary–gonadal axis (HPG axis) in the genetic regulation of menstrual cycle length. *T. Laisk^{1,2,3}, V. Kukushkina¹, D. Palmer^{4,5}, S. Laber^{6,7}, C.-Y. Chen^{4,8,9}, T. Ferreira⁶, N. Rahmioglu¹, K. Zondervan^{7,10}, C. Becker¹⁰, J.W. Smoller^{6,8}, M. Lippincott¹¹, A. Salumets^{1,3,12,13}, I. Granne¹⁴, S. Seminara¹¹, B. Neale^{4,8}, R. Magi², C.M. Lindgren^{6,7,15}*. 1) Estonian Genome Center, University of Tartu, Tartu, Estonia; 2) Dept of Obstetrics and Gynecology, Institute of Clinical Medicine, University of Tartu, Tartu, Estonia; 3) Competence Centre on Health Technologies, Tartu, Estonia; 4) Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA; 5) Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA; 6) Big Data Institute, Li Ka Shing Center for Health for Health Information and Discovery, Oxford University, Oxford, UK; 7) Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK; 8) Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, USA; 9) Broad Institute of MIT and Harvard, Boston, Massachusetts, USA; 10) Oxford Endometriosis CaRe Centre, Nuffield Department of Women's and Reproductive Health, University of Oxford, UK; 11) Harvard Reproductive Sciences Center and Reproductive Endocrine Unit, Massachusetts General Hospital, Boston, Massachusetts, USA; 12) Department of Obstetrics and Gynecology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland; 13) Department of Biomedicine, Institute of Biomedicine and Translational Medicine, University of Tartu, Tartu, Estonia; 14) Nuffield Department of Women's and Reproductive Health, University of Oxford, UK; 15) Program in Medical and Population Genetics, Broad Institute, Boston, MA, USA.

The normal menstrual cycle requires a delicate interplay between the hypothalamus, pituitary, and ovary. Therefore, its length is an important indicator of female reproductive health. Menstrual cycle length has been shown to be partially controlled by genetic factors, especially in the follicle stimulating hormone beta-subunit (*FSHB*) locus. GWAS meta-analysis of menstrual cycle length in 44,871 women of European ancestry confirmed the previously observed association with the *FSHB* locus and identified four additional novel signals in, or near, the *GNRH1*, *PGR*, *NR5A2* and *INS-IGF2* genes. These findings align with what is known about the endocrine control of menstrual cycle length, and confirm the role of the hypothalamic-pituitary-gonadal axis in the genetic regulation of menstrual cycle length, but also highlight potential novel local regulatory mechanisms, such as those mediated by *IGF2*.

2606F

Genome-wide meta-analysis identifies novel loci associated with free triiodothyronine and thyroid-stimulating hormone. A. Matana¹, M. Popović¹, V. Torlak², T. Boutin³, D. Brdar³, I. Gunjača¹, N. Vidan¹, D. Kaličanin¹, I. Kolčić¹, V. Boraska Perica¹, A. Punda², O. Polašek⁴, M. Barbačić¹, C. Hayward³, T. Zemunik¹. 1) Department of Medical Biology, University of Split, School of Medicine, Šoltanska 2, Split, Croatia; 2) Department of Nuclear Medicine, University Hospital Split, Spinčičeva 1, Split, Croatia; 3) MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, United Kingdom; 4) Department of Public Health, University of Split, School of Medicine, Šoltanska 2, Split, Croatia.

Background: Thyroid hormones are essential for the normal function of almost all human tissues, and have critical roles in metabolism, differentiation and growth. Although twin and family studies have estimated that up to 65% of serum free triiodothyronine (fT3), free thyroxine (fT4) and thyroid-stimulating hormone (TSH) variation is genetically determined, most of the heritability is yet unexplained. Moreover, no genome-wide significant signals have been identified for an association with fT3 levels so far. **Methods:** We performed a genome-wide meta-analysis in up to 1731 euthyroid individuals originating from three Croatian cohorts: City of Split and the Islands of Korčula and Vis. Euthyroid participants were selected based on detailed thyroid status that includes biochemical measurements of thyroid function (including measurements of fT3, fT4, TSH, TgAb and TPOAb) and anamnestic data. A total of 7 411 206 variants, imputed using the 1000 Genomes reference panel, were analyzed for an association. Genome-wide association study (GWAS) within each data set was carried out under an additive model, controlling for age, gender and relatedness. Meta-analysis was performed using the inverse-variance fixed-effects method. Additionally, we also conducted multiple-trait GWAS to test the association between each variant and the two intermediately correlated traits fT3 and fT4 simultaneously. **Results:** The *EPHB2* gene variant rs67142165 reached genome-wide significance for association with fT3 plasma levels ($P=9.27 \times 10^{-9}$), and its significance was confirmed in the multiple-trait GWAS ($P=9.72 \times 10^{-9}$). We also found a genome-wide significant association for variant rs13037502 upstream of the *PTPN1* gene and TSH serum levels ($P=1.67 \times 10^{-9}$). **Conclusion:** To the best of our knowledge, this is the first GWAS of these traits performed in minutely selected euthyroid subjects. We identified the first genome-wide significant association for fT3 level to date, as well as a novel association with TSH level. Furthermore, we carried out the first reported multiple-trait GWAS of fT3 and fT4, which resulted in the identification of *EPHB2* as a potential pleiotropic gene. These findings are biologically relevant and add new knowledge to the genetics of observed thyroid-related traits.

2607W

Genetic analyses of Fuchs endothelial corneal dystrophy in a Japanese population. M. Nakano¹, Y. Tokuda¹, H. Adachi¹, Y. Ikeda², M. Ueno², C. Sotozono², S. Kinoshita³, K. Tashiro¹. 1) Department of Genomic Medical Sciences, Kyoto Prefectural University of Medicine, Kyoto, Japan; 2) Department of Ophthalmology, Kyoto Prefectural University of Medicine, Kyoto, Japan; 3) Department of Frontier Medical Science and Technology for Ophthalmology, Kyoto Prefectural University of Medicine, Kyoto, Japan.

Fuchs endothelial corneal dystrophy (FECD, MIM: 136800) is a degenerative disease that causes progressive endothelial cell loss, resulting in vision loss. In United States, FECD is regarded as a common disease - ~4% of the population over 40 years old suffer from the disease, whereas the prevalence of FECD is rare in Japanese. According to the results of a genome-wide association study (GWAS) and the following genetic study in Caucasian, a single nucleotide variant (rs613872) and the trinucleotide repeat (TNR) variant within the introns of transcription factor 4 (*TCF4*) gene were identified as the significant risk factors of late-onset FECD. However, in a Japanese population, we found that rs613872 was monomorphic and the frequency of FECD patients with expanded TNR was less than the Caucasian patients. Moreover, some novel variants from the genes other than *TCF4* have been identified by recent GWAS in Caucasian. Consequently, these results suggested that it should be important to conduct further genetic analyses using Japanese population to reveal the different epidemiology and/or etiology of FECD arisen from different genetic backgrounds. To this end, we performed an association study using the genotype data of 278,032 autosomal variants obtained by Infinium HumanCoreExome BeadChip (Illumina, Inc) derived from 55 FECD patients and 445 controls. None of the loci, which were previously reported in Caucasian, were confirmed to be significant in our population. In contrast, we identified an interesting locus at 6q15 including a genome-wide significant variant ($P < 5.0 \times 10^{-8}$) from the gene involved in ribosome biogenesis and some suggestive variants ($P < 1.0 \times 10^{-5}$) from an adjacent apoptosis-related gene located on the same LD block. We therefore targeted the 6q15 locus and some other suggestive loci (~500 kb, in total) that seemed to be uniquely associated with Japanese FECD patients by means of HaloPlex HS system (Agilent Biotechnologies, Inc), and generated sufficient amount of sequence data to perform further analyses. We are also trying hard to recruit additional FECD patients to perform replication study in an independent population. Overall, the approach shown here should provide not only the genetic information to reveal the molecular mechanism of FECD pathogenesis in Japanese, but also the foundation to explain the mechanism determining the different epidemiology and/or etiology of different populations.

2608T

Effects of HLA class II genotypes on a response to HB vaccine in Japanese population. N. Nishida^{1,2}, M. Sugiyama¹, H. Sawai², J. Ohashi³, S. Khor², T. Tsuchiura¹, K. Tokunaga², M. Mizokami¹. 1) National Center for Global Health and Medicine, Ichikawa, Japan; 2) Department of Human Genetics, The University of Tokyo, Bunkyo-ku, Japan; 3) Department of Biological Sciences, The University of Tokyo, Bunkyo-ku, Japan.

Approximately 5-10% of individuals, who are vaccinated with a hepatitis B (HB) vaccine designed based on the HBV genotype C, fail to acquire protective levels of antibodies. In our previous study, GWAS and HLA association tests revealed that HLA class II genes and *BTNL2* gene are host genetic factors related with a response to the HB vaccine in the Japanese population. Here, we assessed the effects of HLA class II genotypes on a response to the HB vaccine. HLA association tests were carried out using a total of 1,103 Japanese individuals including 94 low responders (HBsAb ≤ 10 mIU/mL), 323 intermediate responders (10 mIU/mL < HBsAb < 100), and 686 high responders (HBsAb ≥ 100 mIU/mL). Classical HLA class II alleles were statistically imputed using the genome-wide SNP typing data. HLA association tests showed that odds ratios in a comparison between low-responders and high-responders were higher than ones in a comparison between low-responders and intermediate- plus high-responders. These results implied that there may exist ambiguous individual(s) who need to be re-categorized into low or high responders among intermediate responders. We next compared genotype frequencies of HLA class II alleles and haplotypes between low responders and high responders. Odds ratios for low response to the HB vaccine were higher for those with both HLA-DRB1 susceptibility alleles (*DRB1*04*05* and *DRB1*14*06*) than for those with only one susceptibility allele; therefore, effects of susceptibility alleles were additive for risk of low response to the HB vaccine. Notably, compound-heterozygote of susceptibility alleles (*DRB1*04*05*, *DRB1*14*06*, and *DQB1*04*01*) and protective alleles (*DRB1*08*03*, *DRB1*15*01*, *DQB1*05*01*, and *DQB1*06*02*) was significantly associated with protection against low response to the HB vaccine, which indicated that one protective HLA-DRB1 or DQB1 molecule can provide dominant protection. Identification of the HLA-DRB1, DQB1 genotypes associated with susceptibility to and protection from low response to the HB vaccine is essential for future analysis of the mechanisms responsible for immune recognition of hepatitis B surface antigens by HLA-DRB1, DQB1 molecules.

2609F

A whole-genome study identifies genetic risk factors for systemic hypertension in high altitude Indian population. K. Sharma^{1,2}, S. Kohli¹, M.A. Syed², G. Mohammad², T. Thinlas³, M.A.Q. Pasha^{1,4}. 1) CSIR-Institute of Genomics and Integrative Biology, Delhi, India; 2) Department of Biotechnology, Jamia Millia Islamia, New Delhi; 3) Department of Medicine, Sonam Norboo Memorial Hospital, Leh, Ladakh, Jammu and Kashmir, India; 4) Indian Council of Medical Research, New Delhi, India.

Systemic hypertension (SH) is a multifactorial disease that tends to increase at high altitude (HA). A genetic predisposition has been suggested for SH; however, the pathogenesis remains largely unknown. In the present study, we used a single nucleotide polymorphism (SNP) chip representing over 2.5 million SNPs to analyze DNA from hypertensive highlanders (HHLs) and normotensive highlanders (NHLs). To the best of our knowledge, this is the most comprehensive genome-wide association study of a HA cohort conducted to date in relation to SH. Here, 1600 SNPs spanning across 162 genes were identified as candidates for association with SH (adjusted P 2×10^{-3} - 5.19×10^{-6}). Whereas the majority of these SNPs are represented in non-coding regions of the genome, thirty-six SNPs were identified in the coding region of their respective genes. Among these, nine SNPs resulted in missense substitutions (MS). Further, the top MS of dynein axonemal heavy polypeptide 11 (*DNAH11*) gene (rs2214326, $P \leq 0.0005$; missense_Ala3467Thr) was found to be significant (odds ratio (OR) = 1.5, $P \leq 0.005$) when screened in a larger sample size. Integrated analysis of 162 genes indicates 47 biological pathways including novel pathways in relation to SH. Also, decreased level of nitric oxide and apelin were found to be significantly associated with SH ($P \leq 0.05$). Further, copy number variation (CNV) analyses of genome-wide data revealed that the number of autosomal CNVs was significantly higher in HHLs as compared to NHLs ($P \leq 0.005$). Interestingly, deletion variations in HHLs were found to be significantly more when compared with NHLs. Also, the average size of deletion variation in kb was found to be 1.64-fold more in HHLs as compared to NHLs. Our findings are encouraging albeit, further validation of these variants may help identify contributing factors to the pathophysiology of SH, as well as categorize potential targets for medical intervention strategies.

2610W

Novel susceptibility loci for tanning ability in a Japanese population identified by genome-wide association study from ToMMo cohort study.

K. Shido^{1,2}, K. Kojima^{2,4,5}, K. Yamasaki¹, A. Hozawa^{2,4}, G. Tamiya^{2,4}, S. Ogishima^{2,4}, N. Minegishi^{2,4}, Y. Kawai⁶, K. Tanno³, Y. Suzuki^{2,4}, S. Aiba¹, M. Nagasaki^{2,4,5}. 1) Department of Dermatology, Tohoku University Graduate School of Medicine, Sendai, Miyagi, Japan; 2) Tohoku Medical Megabank Organization, Tohoku University; 3) Iwate Tohoku Medical Megabank Organization, Iwate Medical University; 4) Graduate School of Medicine, Tohoku University; 5) Graduate School of Information Sciences, Tohoku University; 6) Department of Human Genetics, Graduate School of Medicine, Tokyo University.

Skin types of the Japanese population are classified into Fitzpatrick skin Type III and IV, which are the most abundant skin types in worldwide. Since the information about genes affecting Fitzpatrick skin Type III and IV is limited, we performed the genome-wide association studies (GWAS) for skin types on 9,187 Japanese individuals, using the dataset from the prospective cohort study by the Tohoku Medical Megabank Organization (ToMMo). Skin type information was obtained from the cohort study questionnaires and classified into the Japanese skin type J-I to J-III, which correspond to Fitzpatrick skin Type III and IV. We identified two SNPs reaching genome-wide significance level of 5×10^{-8} : chr15:28228553 C>T (rs74653330) at 15q12-13 (Oculocutaneous albinism II, OCA2) and chr9:16795790 C>A (rs16935073) at 9p22.3-22.2 (Basonuclin 2, BNC2) in the discovery set of 4,475 populations in Miyagi prefecture as well as in the replication set of 4,712 populations in Iwate prefecture. We additionally found the following six loci in the combined analysis of Miyagi and Iwate population: SLC6A17 rs6689641, RAB32 rs78620166, KITLG rs11104947, MC1R rs33932559, DBNDD1 rs138160245, and MSFD12 rs2240751. The GWAS from the ToMMo cohort study revealed novel susceptible loci of tanning ability in the Japanese population by classifying Fitzpatrick skin Type III and IV into the Japanese skin type J-I to J-III.

2611T

Genetic basis of falling risk susceptibility. K. Trajanoska^{1,2}, F. Day^{1,3}, C.

Medina-Gomez¹, A. Uitterlinden¹, J. Perry^{1,3}, F. Rivadeneira^{1,2}. 1) Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands; 2) Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands; 3) MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Cambridge, UK.

Background: Falls and fall-related injuries are growing healthcare problem in elderly people. Approximately one out of three older adults will experience at least one fall. Multiple factors can increase the risk of falling. We aimed to determine how much of an individual's fall susceptibility can be attributed to genetics. **Methods:** We conducted the first genome-wide association study (GWAS) to evaluate the genetic variation in fall risk in 89,076 cases and 362,103 controls from the UKBiobank. Individuals were genotyped with the Affymetrix UKBiobank chip. Model included age and sex analysed with BOLT-LMM v.2.3.1. Next, we tested the predictive value of polygenic risk scores (PRS) across diverse significance thresholds in relation to falls using individual level data from the Rotterdam Study. Finally, we used LD score regression to: 1). estimate falling risk heritability 2). test for genetic correlation with other traits and 3). identify tissues where expression of genes related to fall-associated variants are enriched. **Results:** We identified two novel fall loci mapping near *PER4* on 7p21.3 (rs2709062, OR=1.03, P=3.4x10⁻⁰⁶) and near *TSHZ3* on 19q12 (rs2111530, OR=1.03, p=1.2x10⁻⁰⁶). Falling risk heritability was modest (2.5%). PRS (10⁻⁰² to 10⁻⁰³ thresholds) were significant explaining small proportion of the trait variance. Falls had strong positive genetic correlation with fracture risk (rg=0.35, se=0.05), insomnia (rg=0.42, se=0.05), neuroticism (rg=0.26, se=0.08) and ADHD (rg=0.44, se=0.14); and strong negative genetic correlation with muscle strength (rg=-0.24, se=0.04), intelligence (rg=-0.12, se=0.04) and social well-being (rg=-0.29, se=0.05). Brain and in particular cerebellum tissue showed the highest gene expression enrichment (FDR<5%) for fall-associated variants. **Conclusions:** Falling risk constitutes a heritable, heterogeneous and polygenic trait. Strong evidence link it to fracture risk and grip strength. Genes expressed in the cerebellum are enriched for falls-associated variants, indicating genetic interplay between postural balance and falls.

2612F

A genome-wide association study identifies uric acid-related genetic variants in the Taiwanese population. J. Wei¹, W. Chen^{1,2}, Y. Liu³, S. Tsai^{4,5}, P. Kuo^{4,6}. 1) Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan; 2) Department of Psychiatry, Taipei City Psychiatric Center, Taipei City Hospital, Taipei; 3) Center for Neuropsychiatric Research, National Health Research Institutes, Miaoli County, Taiwan; 4) Department of Psychiatry, Taipei Veterans General Hospital, Taipei, Taiwan; 5) Division of Psychiatry, National Yang-Ming University, Taipei, Taiwan; 6) Research Center for Genes, Environment and Human Health, National Taiwan University, Taipei, Taiwan, Taiwan.

Uric acid (UA), which has been found to be associated with many physical conditions, is one of the important biomarkers for the health of humans' body as small changes in UA levels can make a substantial impact. Previously, several genetic loci were reported to be associated with serum UA levels in different ethnic groups. We conducted a genome-wide association study (GWAS) to identify UA-related genetic variants in the Taiwanese general population with 10,295 subjects from Taiwan Biobank. All subjects were genotyped using Affymetrix TWB array with 646,735 markers, and serum UA levels were examined. Standard procedures of quality controls for GWAS data were performed, and 10,095 individuals remained in data analysis. Our results demonstrated that 88 loci were associated with UA levels while adjusted for sex and age. Marker rs3775948 (beta= -0.2556, p-value= 6.582E-47) and rs2231142 (beta= 0.2623, p-value= 2.391E-43) showed the most significant association with UA levels, which were located on the *SLA2A9* (solute carrier family 2, member 9) and *ABCG2* (ATP Binding Cassette Subfamily G Member 2) gene regions, respectively. These two variants were independently associated with UA concentration in conditional analyses. These two genes were also implicated in influencing UA levels in previous association studies. Replication studies will be performed in an additional sample, and the effects of these genetic variants on other UA related phenotypes are under investigation.

2613W

Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. Y. Wu¹, J. Zeng¹, F. Zhang¹, Z. Zhu¹, T. Qi¹, Z. Zheng^{2,1}, L. Lloyd-Jones¹, R. Marioni^{4,5}, N. Martin⁶, G. Montgomery¹, I. Deary⁶, N. Wray^{1,2}, P. Visscher^{1,2}, A. McRae¹, J. Yang^{1,2}. 1) Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland 4072, Australia; 2) Queensland Brain Institute, The University of Queensland, Brisbane, Queensland 4072, Australia; 3) The Eye Hospital, School of Ophthalmology & Optometry, Wenzhou Medical University, Wenzhou, Zhejiang 325027, China; 4) Medical Genetics Section, Centre for Genomics and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, EH4 2XU, UK; 5) Centre for Cognitive Ageing and Cognitive Epidemiology, Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh, EH8 9JZ, UK; 6) Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland 4029, Australia.

The identification of genes and regulatory elements underlying the associations discovered by GWAS is essential to understanding the aetiology of complex traits (including diseases). Here, we demonstrate an analytical paradigm of prioritizing genes and regulatory elements at GWAS loci for follow-up functional studies. We perform an integrative analysis that uses summary-level SNP data from multi-omics studies to detect DNA methylation (DNAm) sites associated with gene expression and phenotype through shared genetic effects (i.e., pleiotropy). We identify pleiotropic associations between 7,858 DNAm sites and 2,733 genes. These DNAm sites are enriched in enhancers and promoters, and >40% of them are mapped to distal genes. Further pleiotropic association analyses, which link both the methylome and transcriptome to 12 complex traits, identify 149 DNAm sites and 66 genes, indicating a plausible mechanism whereby the effect of a genetic variant on phenotype is mediated by genetic regulation of transcription through DNAm.

2614T

Replicated association of *CYP4F2* variant with change in plasma vitamin E level after supplementation. J. Xu¹, N.C. Gaddis², R.S. Parker¹, A.G. Clark³, P.J. Goodman⁴, S.M. Lippman^{5,6}, D.B. Hancock⁷, P.A. Cassano¹. 1) Division of Nutritional Sciences, Cornell University, Ithaca, NY; 2) Research Computing Division, RTI International, Research Triangle Park, NC; 3) Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY; 4) SWOG Statistical Center, Fred Hutchinson Cancer Research Center, Seattle, WA; 5) Moores Cancer Center, University of California, San Diego, CA; 6) Department of Medicine, Division of Hematology-Oncology, University of California, San Diego, CA; 7) Behavioral and Urban Health Program, Behavioral Health and Criminal Justice Division, RTI International, Research Triangle Park, NC.

Randomized controlled trials (RCTs) have investigated whether supplementation of vitamin E (vitE), a nutrient with antioxidant properties, can reduce age-related inflammatory disease risks. Evidence from RCTs, however, is inconsistent, which could be due to variation in bioavailability of vitE supplementation among different subgroups of individuals. The only prior genome-wide association study of plasma α -tocopherol response (denoted as α -TOH, the main vitE form in plasma) to vitE supplementation ($n=2,112$ male smokers from Finland) reported two statistically significant single nucleotide polymorphisms (SNPs) (Major et al. J Nutr 2012). We tested these SNPs for independent replication using the vitE-only arm of the Respiratory Ancillary Study (NCT00063453) to the Selenium and Vitamin E Cancer Prevention Trial (SELECT, S0000 and NCT00006392) in men aged 50+ ($n=424$ European Americans [EAs], $n=139$ African Americans [AAs]) with >3 years of supplementation with 400 IU/day α -TOH. Plasma α -TOH was measured at baseline and at year 3, and we tested SNP associations with change in plasma α -TOH level from baseline to year 3. One SNP met the replication threshold ($P<0.025$, accounting for multiple testing): the *CYP4F2* missense SNP rs2108622. *CYP4F2* is a biologically plausible gene for vitE metabolism, given its function as a tocopherol- ω -hydroxylase, which initiates the oxidative catabolism of tocopherols. We found that the rs2108622-T minor allele (frequency=29% in EAs and 8% in AAs) was associated with a greater increase in plasma α -TOH after supplementation (i.e., greater response) with similar magnitudes of effect across ancestries ($\beta=0.15$, SE=0.08, and $P=0.04$ in EAs; $\beta=0.26$, SE=0.18, and $P=0.15$ in AAs; and meta-analysis $P=0.01$). The effect direction was consistent with the previous GWAS, where rs2108622-T was associated with a higher plasma α -TOH level after supplementation. Follow-up analysis showed that rs2108622 is a potential expression quantitative trait locus for *CYP4F2* in whole blood tissue: its T allele was associated with lower *CYP4F2* expression ($P=1.7\times 10^{-12}$, $n=369$) in the Genotype-Tissue Expression Project. In summary, this independent replication finding adds to the evidence that rs2108622 is associated with plasma response to vitE supplementation and supports the hypothesis that genetic variation influences bioavailability of vitE supplementation across populations. Funding: R01HL071022, R21HL125574, HHSN268201100037C, U10 CA37429, 5UM1CA182883.

2615F

Identifying novel genetic defects for radiation hypersensitive phenotypes through Next-Gen sequencing. M. Gupta¹, X. Liu², S. Teraoka², P. Concanon², A. Quinlan¹. 1) Department of Human Genetics, University of Utah, Salt Lake City, UT, USA; 2) Genetics Institute and Department of Pathology, Immunology and Laboratory Medicine, University of Florida, Gainesville, FL, USA.

Although ionizing radiation (IR) is an effective therapeutic agent for cancer treatment, it is also a potent carcinogen. Sensitivity to the cell-killing effects of IR vary across human population with a subset of individuals displaying extreme hypersensitivity, usually attributable to inherited defects in DNA damage response pathways. The present study aims to identify causative genetic defects in IR hypersensitive individuals that share clinical features with known Mendelian DNA damage response disorders Ataxia-telangiectasia and Nijmegen breakage syndrome but lack mutations in the causative genes *ATM* and *NBN*. This approach can shed light on novel proteins/pathways involved in DNA damage responses to radiation, and previous studies from our group have been successful in identifying and validating the role of new genes such as *MTPAP* and *ATIC* in radiation sensitivity (RS). Here, we performed whole genome and exome sequencing in a cohort of 84 radiation sensitive individuals to identify rare or novel loss of function (LoF) SNVs and INDELs, as well as copy number variants (CNVs). SNVs and INDELs were annotated with functional consequence predictions using GEMINI and prioritized under the assumption of a recessive model. We identified three novel frameshift, potential LoF variants in three genes (*LIMCH1*, *CTRB1* and *SLC2A5*) not known to be involved in RS or DNA damage pathways. In addition, we have also adopted an epistatic model to identify rare/novel heterozygous mutations in gene-pairs which might disrupt essential protein-protein interactions (PPIs) among RS genes. This approach led to identification of potentially deleterious novel and rare mutations in *ATM-ATR*, *TP53-TP53BP1* and *BRCA2-BARD1* gene pairs. Apart from SNVs, we have also detected large homozygous and heterozygous deletions and duplications across multiple genes in 5 individuals. We will report progress on the validation of these putative radiation sensitivity genes, through a two step process. Initially the genes are knocked down using shRNA in a cell line of normal radiosensitivity followed by testing for increased susceptibility to death by ionizing radiation. Genes passing this screen are then further tested by knockdown followed by colony survival assays. In conclusion, our robust methodology to detect and validate RS gene candidates can provide new insights into the mechanisms underlying radiation hypersensitivity.

2616W

Gene expression analyses in thrombotic storm indicate dysregulation of coagulation and extracellular matrix pathways.

K. Nuytemans¹, A.J. Griswold¹, L. Gomez¹, N. Hofmann¹, P. Whitehead-Gay¹, C.S. Kitchens², D. Erkan³, L.R. Brandao⁴, A.H. James⁵, R. Kulkarni⁶, T.L. Ortel⁶, M.A. Pericak-Vance¹, J.M. Vance¹. 1) University of Miami, John P. Hussman Institute for Human Genomics, Miami, FL; 2) University of Florida, Gainesville, FL; 3) Hospital for Special Surgery, Weill Medical College of Cornell University, New York, NY; 4) Hospital for Sick Children, University of Toronto, Toronto, ON, Canada; 5) Duke University Medical Center, Durham, NC; 6) Michigan State University, East Lansing, MI.

Thrombotic storm (TS; www.thromboticstorm.com) is a rare and severe clinical phenotype that occurs in a small subset of patients with venous thromboembolic disease. It is characterized by more than two acute arterial and/or venous thromboemboli, frequent events in unusual locations, progressive/recent unexplained recurrence and frequent refractory and/or atypical response to therapy. As gene expression during thrombotic events compared to baseline presentation will help elucidate the underlying pathophysiology of active disease, our goal was to identify transcriptomic changes in blood in four TS patients during and after thrombotic storm event. All patients presented with both venous and arterial thrombi at time of event and two out of four patients had a history for antiphospholipid antibodies. Average age at thrombotic event was 25.7y (21-32). We performed gene expression analysis of all coding genes during and after the event, using a paired analyses model to correct for baseline expression differences across patients. A total of 71 genes (56 up-regulated, 15 down-regulated) were differentially expressed at FDR<0.05 (fold change range -2.5 – 6.0). Of the known thrombotic factors, we observed an upregulation of anticoagulation factor Protein S (*PROS1*), indicating a clear attempt of the endothelial cells to control the clotting. Other genes of interest include *ADAMTS1* and 2 and *ITGA2B* and *B3*; all elements involved in extracellular matrix/glycocalyx formation. Pathway analyses confirm and implicate coagulation cascades, extracellular matrix organization and actin cytoskeleton (upregulation), as well as O-glycosylation of proteins and negative regulation of cell proliferation (downregulation). These results support a dysfunction of coagulation pathways and indicate dysregulation of genes involved in extracellular matrix/glycocalyx stability and O-glycosylation at the time of thrombotic storm events. Interestingly, these data support our recent report on rare, high risk variants in genes involved in chondroitin sulfate, an O-glycosylation chain on proteins of the glycocalyx protecting the endothelium in TS patients. Studies as these are critical for our understanding of underlying mechanisms in TS and thrombosis in general; and might inform on therapeutic biomarkers for thrombotic disease.

2617T

Inflammation derived molecular pathogenesis of keratoconus: An insight into disease onset.

M. Azam¹, R. Qamar^{1,2}, A. Ahmed¹, S. Tahir¹, M.K. Niazi^{3,4}. 1) Translational Genomics Laboratory, Department of Biosciences, COMSATS University Islamabad; 2) Pakistan Academy of Sciences, Islamabad, Pakistan; 3) Armed Forces Institute of Ophthalmology, Rawalpindi, Pakistan; 4) Army Medical College, Rawalpindi, Pakistan.

Keratoconus (KC) is a corneal disorder in which cornea becomes thin and bulges outward in a cone-shape, resulting in myopia, irregular astigmatism substantially leading to visual impairment. It is a common corneal anomaly that is estimated to be 229 per 100,000 individuals in the population of the Sub-continent. KC is a multifactorial disease; both genetic and environmental factors contribute in its development. To date KC pathogenesis is undefined; however, the involvement of inflammatory pathways in disease development is noticeable. In our previous study, the role of endogenous factors; promoter region polymorphism rs1800629 (-308G>A) in the inflammatory pathway component *TNF-α*, tumor necrosis factor receptor 2 (*TNFR2*), v-rel avian reticuloendotheliosis viral oncogene homolog A (*RELA*) and interleukin 6 (*IL-6*) exhibited significant upregulation in the keratoconic tissues suggesting a strong correlation of *TNF-α*-induced higher expression of inflammatory pathway molecules in KC manifestation, thus implicating KC as an inflammatory condition. Therefore, in the present study, the role of other endogenous factors involved in inflammatory pathway whose deregulated expression could lead to KC development; interleukin 1β (*IL-1β*), hypoxia-Inducible Factor 1 (*HIF-1α*), matrix metalloproteinase 16 (*MMP16*) and eleven collagen genes (*COL1A1*, *COL1A2*, *COL4A4*, *COL6A1*, *COL6A2*, *COL8A1*, *COL11A1*, *COL11A2*, *COL12A1*, *COL16A1* and *COL28A1*), were investigated through expression profile studies in keratoconus corneal tissues (cases) and deceased donor corneal tissues (controls). Real-time quantitative PCR was performed and Mann-Whitney U test was applied for statistical analysis. Significant difference in expression of *IL-1β* (p=0.028), *HIF-1α* (p=0.028), *COL4A4* (p=0.029), *COL16A1* (p=0.029) and *COL28A1* (p=0.028) was observed between keratoconus and control corneal tissues. The current study is among one of the few studies suggesting a possible role of *IL-1β*, *HIF-1α*, *COL4A4*, *COL16A1* and *COL28A1* along with other inflammatory components involved in KC development. To the best of our knowledge, this is the first study regarding molecular pathogenesis of KC patients in Pakistan. This study primarily highlights the interrelated and intra-related association of multiple components of inflammatory pathway in the development of disease where their deregulation may imply towards keratoconus onset.

2618F

Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. T. Qi¹, Y. Wu¹, J. Zeng¹, F. Zhang^{1,2}, A. Xue¹, L. Jiang¹, Z. Zhu¹, K. Kemper¹, L. Yengo¹, Z. Zheng^{1,3}, R. Marion^{1,4}, G. Montgomery¹, I. Deary⁵, N. Wray^{1,2}, P. Visscher^{1,2}, A. McRae¹, J. Yang^{1,2,3}. 1) Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland 4072, Australia; 2) Queensland Brain Institute, The University of Queensland, Brisbane, Queensland 4072, Australia; 3) The Eye Hospital, School of Ophthalmology & Optometry, Wenzhou Medical University, Wenzhou, Zhejiang 325027, China; 4) Medical Genetics Section, Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, EH4 2XU, UK; 5) Centre for Cognitive Ageing and Cognitive Epidemiology, Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh, EH8 9JZ, UK.

Understanding the difference in genetic regulation of gene expression between brain and blood is important for discovering genes for brain-related traits and disorders. Here, we estimate the correlation of genetic effects at the top associated *cis*-expression or –DNA methylation (DNAm) quantitative trait loci (*cis*-eQTLs or *cis*-mQTLs) between brain and blood (r_s). Using publicly available data, we find that genetic effects at the top *cis*-eQTLs or mQTLs are highly correlated between independent brain and blood samples ($r_s = 0.70$ for *cis*-eQTLs and $r_s = 0.78$ for *cis*-mQTLs). Using meta-analyzed brain *cis*-eQTL/mQTL data ($n = 526$ to $1,194$), we identify 61 genes and 167 DNAm sites associated with 4 brain-related phenotypes, most of which are a subset of the discoveries (97 genes and 295 DNAm sites) using data from blood with larger sample sizes ($n = 1,980$ to $14,115$). Our results demonstrate the gain of power in gene discovery for brain-related phenotypes using blood *cis*-eQTL/mQTL data with large sample sizes.

2619W

MR-pheWAS with stratification and interaction: Searching for the causal effects of smoking heaviness. L.A.C. Millard^{1,2,3}, M.R. Munafò^{1,4}, K. Tilling^{1,2}, G. Davey Smith^{1,2}. 1) MRC Integrative Epidemiology Unit (IEU), Bristol Medical School, University of Bristol, Bristol, United Kingdom; 2) Population Health Sciences, Bristol Medical School, University of Bristol; 3) 3 Intelligent Systems Laboratory, Department of Computer Science, University of Bristol, Bristol, UK; 4) UK Centre for Tobacco and Alcohol Studies, School of Experimental Psychology, University of Bristol, Bristol, UK.

Background Mendelian randomization (MR) is an established instrumental variable approach for estimating the causal effect of an environmental exposure on a downstream outcome. The gene x environment (GxE) study design can be used within an MR framework to determine the extent to which a MR estimate of a causal effect may be biased if the genetic instrument affects the outcome through pathways other than via the exposure of interest (known as horizontal pleiotropy). MR phenotype-wide association studies (MR-pheWAS) search for the causal effects of an exposure, and a recently published tool (PHESANT) means that it is now possible to do this comprehensively, across thousands of traits, in UK Biobank. In this study, we introduce the GxE MR-pheWAS approach, and search for the causal effects of smoking heaviness – stratifying on smoking status (ever versus never) – as an exemplar. If a genetic variant is associated with smoking heaviness (but not smoking initiation), and this variant affects an outcome (at least in part) via tobacco intake, we would expect the effect of the variant on the outcome to differ in ever versus never smokers. In the absence of horizontal pleiotropy (i.e. if this effect is entirely mediated by tobacco intake), we would expect to see an effect in ever smokers but no effect in never smokers. **Methods and findings** We used PHESANT to search for the causal effects of smoking heaviness, instrumented by the genetic variant rs16969968, among never and ever smokers respectively, in UK Biobank. We ranked results by: 1) the strength of effect of rs16969968 among ever smokers, and 2) the strength of interaction between ever and never smokers, accounting for the number of tests performed. We identified both established and novel causal effects of smoking heaviness, where associations of rs16969968 with a given outcome differed among ever versus never smokers, indicating an effect of the genetic variant on the outcome via tobacco intake. We replicated previously established causal effects of smoking heaviness, including a detrimental effect on lung function and pulse rate. Novel results included an effect of smoking heaviness on several brain phenotypes. **Conclusions** We have demonstrated a novel GxE MR-pheWAS approach to identify causal effects, while simultaneously assessing the extent to which estimated effects may be biased by horizontal pleiotropy. Our proof-of-principle analyses identified both known and novel causal effects of smoking heaviness.

2620T

Lack of associations of human genetic variations with gut microbiome diversity in healthy individuals. P. Scepánovic^{1,2}, S. Mondot³, C. Hammer⁴, C. Alanio⁵, E. Patin^{6,7}, J. Bergstedt⁸, L. Abej^{9,10}, M.L. Albert¹, D. Duffy¹¹, O. Lantz^{12,13}, L. Quintana-Murci^{6,7}, J. Fellay^{1,2}, *The Milieu Intérieur Consortium*. 1) School of Life Sciences, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland; 2) Swiss Institute of Bioinformatics, Lausanne, Switzerland; 3) MICALIS Institute (INRA/AgroParisTech), Jouy-en-Josas, France; 4) Department of Cancer Immunology, Genentech Inc., San Francisco, CA 94080, USA; 5) Institute for Immunology, Perelman School of Medicine, University of Pennsylvania, PA, USA; 6) Unit of Human Evolutionary Genetics, Department of Genomes and Genetics, Institut Pasteur, Paris, France; 7) Centre National de la Recherche Scientifique, UMR2000, Paris, France; 8) Department of Automatic Control, LTH, Lund University, Lund, Sweden; 9) Laboratory of Human Genetics of Infectious Diseases, Necker branch, Inserm U1163, Paris Descartes University, Imagine Institute, Paris, France; 10) St Giles laboratory of Human Genetics of Infectious Diseases, Rockefeller branch, The Rockefeller University, New York, NY, USA; 11) Immunobiology of Dendritic Cells laboratory (INSERM U1223/Institut Pasteur), Paris, France; 12) Institut Curie, PSL Research University, Inserm U932, 75005, Paris, France; 13) Center of Clinical Investigations, CICBT1428 IGR/Curie, 75005, Paris, France.

Introduction. The gut microbiota is an important determinant of human health. It contains a large number of bacterial species with tremendous compositional variation across individuals. This diversity is influenced by multiple environmental factors but also by host genetic variation, as shown by heritability analysis. In the framework of the *Milieu Intérieur* Consortium, a population-based study aimed at dissecting the different sources of immune response variance within a healthy population, the commensal intestinal microbiota of 1000 healthy individuals was assessed by 16S rRNA sequencing. We conducted genome-wide association studies (GWAS) to detect associations between human genetic variants and gut microbiota diversity. **Methods.** A total of 1000 healthy individuals of west European ancestry, with a 1:1 sex ratio, and stratified across five-decades of life (age 20 – 69), were recruited in France and genotyped using the Illumina OmniExpress and HumanExome arrays. 16S rRNA profiles were obtained from stool samples of 862 individuals. We used linear regression to search for associations between >5 million SNPs and multiple indicators of variation in the gut microbiota, including α - (Simpson's index) and β - (weighted UniFrac distances) diversity, as well as quantitative measurements of operational taxonomic units (OTUs). Furthermore, we used sequence kernel association test to investigate the potential associations of rare variants. The first two principal components of the genotyping data, age and sex were used as covariates in all tests. **Results and discussion.** No genome-wide significant associations were identified after Bonferroni correction for the number of phenotypes tested in the single-marker or in the burden (rare variant) analyses. We did not replicate previously reported associations between β -diversity and host genetics. Our study suggests that host genetic variation is likely to play a minor role in shaping the global diversity and the taxonomic composition of the gut microbiome in healthy populations. Because the putative effect of any associated gene variant will be small, our results also emphasize the need for adequate statistical power and large sample sizes in future assessments.

2621F

Aberrant expressions of *miRNA-206* target, *FN1*, in multifactorial Hirschsprung disease. N.Y.P. Budi¹, A.S. Kalim¹, W. Santiko¹, F.D. Musthofa², K. Iskandar², A. Makhmudi¹, G. Gunadi¹. 1) Pediatric Surgery Division, Departement of Surgery, Faculty of Medicine, Public Health and Nursing, UGM, Yogyakarta, Indonesia, Yogyakarta, Special Distric of Yogyakarta, Indonesia; 2) 6Department of Child Health, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada/UGM Academic Hospital, Yogyakarta 55291, Indonesia.

Background: MicroRNAs (miRNAs) have been associated with the Hirschsprung disease (HSCR) pathogenesis, however, the findings are still inconclusive. We aimed to investigate the effect of *miRNA-206* and its targets, *fibronectin 1* (*FN1*), *serum deprivation response* (*SDPR*), and *paired box 3* (*PAX3*) expressions on multifactorial HSCR in Indonesia, a genetically distinct group within Asia. **Methods:** We determined the *miRNA-206*, *FN1*, *SDPR* and *PAX3* expressions in both the ganglionic and aganglionic colon of HSCR patients and control colon by quantitative real-time polymerase chain reaction (RT-PCR). **Results:** Twenty-one sporadic HSCR patients and thirteen controls were ascertained in this study. The *miRNA-206* expression was up-regulated (2-fold) in the ganglionic colon and down-regulated (1.5-fold) in the aganglionic colon compared to the control group (12.4 ± 3.0 vs. 14.1 ± 3.9 vs. 13.1 ± 2.7), but these differences did not reach significant levels ($p=0.48$ and $p=0.46$, respectively). Interestingly, the *FN1* expression was significantly increased in both the ganglionic (38-fold) and aganglionic colon (18-fold) groups compared to the control group (5.7 ± 3.0 vs. 6.8 ± 2.3 vs. 11.0 ± 5.0 ; $p=0.001$ and $p=0.038$, respectively). Furthermore, the expressions of *SDPR* were similar in the ganglionic, aganglionic and control colon groups (2.4 ± 0.6 vs. 2.2 ± 0.4 vs. 2.1 ± 0.6 ; $p=0.16$ and $p=0.39$, respectively), while no change was observed in the *PAX3* expression between the ganglionic, aganglionic, and control colon groups (3.8 ± 0.8 vs. 4.1 ± 0.8 vs. 3.7 ± 1.1 ; $p=0.83$ and $p=0.44$, respectively). **Conclusion:** Our study is the first report of aberrant *FN1* expressions in the colon of patients with HSCR. These abnormal *FN1* expressions might contribute to the pathogenesis of HSCR.

2622W

Potential role of *STAT4* genetic variation in treatment decision-making for hepatitis B e antigen-positive chronic hepatitis B. *H. Chen, J. Sun, B. Zhou, X.E. Liang, R. Fan, J.L. Hou, D.K. Jiang.* Infectious Diseases and Hepatology Unit, Nanfang Hospital, Southern Medical University, Guangzhou, China.

AimCurrently, peginterferon alfa (PegIFN α) and nucleos(t)ide analogues (NUCs) are the two main first-line therapies for chronic hepatitis B (CHB). Unfortunately, both of them produce response in only minority of CHB patients. Previous study showed that a genetic variant in *STAT4* (rs7574865) is a reliable predictor of response to PegIFN α therapy for hepatitis B e antigen (HBeAg)-positive CHB patients (Jiang DK et al., *Hepatology* 2016). Our aim was to further evaluate the performance of rs7574865 in PegIFN α treatment as well as NUCs treatment for HBeAg-positive CHB patients. **Methods**A total of 1,860 HBeAg-positive CHB patients, including 254 patients (PegIFN α Cohort 1) and 708 patients (PegIFN α Cohort 2) treated with PegIFN α , 334 patients (NUCs cohort 1) and 564 patients (NUCs cohort 2) treated with NUCs, were enrolled. Treatment efficacy was assessed using sustained virologic response (SVR), which was defined as HBeAg seroconversion along with HBV DNA level <2000 IU/mL at week 72 for PegIFN α cohorts or at week 104 for NUCs cohorts. Early responders were defined as hepatitis B s antigen (HBsAg) < 1500 IU/mL and HBV DNA < 10⁵ copies/mL at week 24. Genotyping of rs7574865 was performed by Sequenom MassArray platform. Cochran-Armitage trend test and logistic regression analysis were used to evaluate the association of rs7574865 with SVR. **Results**The association of rs7574865 with SVR was significant or marginally significant in PegIFN α cohorts (P-trend = 0.01, 0.07 and 0.004 in PegIFN α cohort 1, 2 and their combination, respectively), but not significant in NUCs cohorts (P-trend = 0.88, 0.59 and 0.81 in NUCs cohort 1, 2 and their combination, respectively). When stratified the patients in combined PegIFN α cohort according to HBV genotype, the SVR rates in HBV genotype B patients with rs7574865 GG, GT and TT genotype were 46%, 30% and 21%, respectively (P-trend = 0.002); while in HBV genotype C patients were 17%, 19% and 19%, respectively (P-trend = 0.66). For HBV genotype B patients, we further evaluated the performance of rs7574865 in early responders and non-early responders. We found that, in PegIFN α cohort 1, 2 and their combination, the SVR rates in early responders with rs7574865 TT genotype were as high as 100%, 67% and 72%, while in non-early responders with GG genotype were only 19%, 15% and 15%, respectively. **Conclusion** By clarifying the performance of rs7574865 in CHB patients, our research may guide physicians and patients in choosing better therapy.

2623T

Next-generation sequencing (NGS) based pharmacogenomics panel testing revealed *CYP2C9* as a critical gene for warfarin dosing in Taiwan. *Y.C. Chen¹, Y.S. Liu¹, P.L. Chen^{1,2,3,4}.* 1) Medical Genomics and Proteomics, National Taiwan University of Medicine, Taipei, Taiwan; 2) Department of Medical Genetics, National Taiwan University Hospital, Taipei, Taiwan; 3) Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan; 4) Graduate Institute of Molecular Medicine, National Taiwan University College of Medicine, Taipei, Taiwan.

Pharmacogenomics is the study of the relationship between inherited genetic variations and drug response in individuals. It plays an important role in precise drug dosing. Here we established a pharmacogenomics testing platform using capture-based target region enrichment followed by next-generation sequencing (NGS). Our NGS panel covered 345 major pharmacogenomics genes including pharmacokinetics genes and pharmacodynamics genes, such as various CYP genes, *ABCB1*, *CFTR*, *DRYR*, *EGFR*, *HLAs*, *KRAS*, *RYR1*, *TPMT*, *UGT1A1* and *VKORC1*, dedicated to identifying specific pharmacogenomic variations and biomarkers of adverse effects. We then applied the panel to 34 individuals, including 6 samples from Taiwan Biobank as controls and 28 patients with adverse drug reactions recruited from National Taiwan University Hospital. Our data confirmed that specific genetic variants are the main determinants for dosing of medications, especially warfarin. For example, our data suggested that, in Taiwan, patients carrying *CYP2C9**2 (c.430C>T, p.R144C, rs1799853) or *CYP2C9**3 (c.1075A>C, p.I359L, rs105910) would decrease activity of warfarin metabolism, therefore required lower doses to obtain similar efficacy of anticoagulation. Moreover, the distribution of cytochrome P450s genes/alleles is quite variable between population. Thus, it is important to genotype CYPs and identify the frequencies of CYP alleles with implications for personalized drug therapy and health programs. Therefore we set up a complete reference and analysis pipeline for CYP genes from whole genome sequencing data of ~1000 Taiwan Biobank participants.

2624F

Genetic and transcriptomic variation in human hepatic pharmacogenes.

K.G. Claw¹, T.A. Thornton², D.A. Nickerson³, E.G. Schuetz⁴, K.E. Thummel¹.

1) Department of Pharmacaceutics, University of Washington, Seattle, WA; 2) Department of Biostatistics, University of Washington, Seattle, WA; 3) Department of Genome Sciences, University of Washington, Seattle, WA; 4) St. Jude's Children's Research Hospital, Toronto, Canada.

In the area of pharmacogenomics, the liver is an ideal system to study due to its role in defining the efficacy and toxicity response to drugs that target the liver (e.g., warfarin and statins). Nearly 75% of all commonly prescribed medications are metabolized to some degree in the liver, making it a controlling site for the clearance of thousands of therapeutic agents. Well-known pharmacogenes include the cytochrome P450s, UGT and SULT genes which are responsible for most of the drug biotransformation reactions that occur in the liver. However, knowledge about the regulation of these genes remains incomplete. We used next-generation sequencing to target the total mRNA, microRNA and long non-coding RNA and sequenced the exons of 84 pharmacogenes from 279 human livers, collected as part of an organ procurement program by St. Jude Children's Research Hospital and University of Washington School of Pharmacy. In addition, human liver microsomal and cytosolic protein and enzyme activities were collected for specific genes of interest. We identified 10,770 single nucleotide variants (SNVs) in the 84 genes and flanking regions. With this robust dataset, we identified associations between common genetic variants and transcription levels of pharmacogenes. Pharmacogenes of interest were *CYP3A5* and *VKORC1*. *CYP3A5/4* is a member of the P450 family and accounts for about 35% of drug metabolism. Within *CYP3A5*, a total of 57 variants were identified, of which 20 were singletons and 16 SNVs had MAF $\geq$ 1%. *VKORC1* was previously identified in genome-wide association studies as a significant contributor to variability in warfarin dose and the vitamin K cycle. Within *VKORC1*, a total of 26 variants were identified, of which 9 were singletons. Overall, we identified a total of 483 genetic variants that were significantly associated with the transcription levels of 35 pharmacogenes. Six common SNVs in *VKORC1* and 5 SNVs in *CYP3A5* were associated with mRNA transcription levels. Some variants were previously identified as reduced function variants, suggesting that our results are consistent with known associations. Further investigations may yield novel regulatory candidate genes that further account for interindividual variability. Regulatory elements influencing specific anabolic and catabolic pathways and genes, such as those controlling warfarin response (*VKORC1*, *CYP4F2*, *GGCX* and *CYP2C9*), can be further characterized with functional assays in liver tissue.

2625W

Are drug targets with genetic support twice as likely to be approved? – Revised estimates of the impact of genetic support for drug mechanisms on the probability of eventual drug approval. J.F. Degner, J.W. Davis, E.A. King. Computational Genomics, AbbVie, North Chicago, IL.

High drug prices can in large part be attributed to the cost of bringing new molecular entities (NMEs) from early evidence of disease relevance through the drug development process to eventual regulatory approval. On average, it is estimated to cost \$1.8 billion to take an NME to market. Despite strong vetting for disease activity, only 10% of candidate NMEs in early stage clinical trials are eventually approved and this probability of approval has a direct relationship to total cost per approved drug (Paul *et al.* 2010; *Nat. Rev. Drug Discov.*). Analyzing historic data of the progress of drug compounds through the drug development pipeline, Nelson *et al.* 2015 (*Nat. Genet.*) concluded pipeline drug targets with human genetic evidence of disease association are twice as likely to lead to approved drugs. The specific claim of doubled approval probability, if true, could lead to fewer failed clinical programs thereby lowering drug development costs and costs to consumers. Despite our excitement about the implied transformative effect of current pharmaceutical investment in genomics research, there is room for skepticism. Determining drug genetic evidence requires many decisions, but it is not clear that safeguards against overfitting were employed. Since this paper, the number of SNP-trait associations in the GWAS Catalog has more than quadrupled, new drug target-indications pairings have been trialed, and development of others has advanced. We can now revisit these estimates on updated data and determine if genetic evidence predicts recent successes. Our work confirms genetically supported drug targets were more likely to progress since 2013, with the strongest effect on Phase II to III transitions. The effect of OMIM genetic evidence may be even larger than originally reported (2.2 fold effect on Phase II to III progression). However, GWAS genetic evidence effect estimates tend to be lower than originally reported (1.4-fold effect on approval from Phase I using updated clinical data) and near zero using only post-2013 genetic association data. Both estimates depend strongly on the similarity between the associated trait and drug indication and increase when controlling for target class. Target success prediction may be improved through a more nuanced understanding of the effects across different modalities of genetic architecture and by better methods for joint modeling of genetic evidence with other components of drug success probability.

2626T

A missense change in *CFB* is a potential predictor of sustained virologic response to peginterferon alfa therapy of HBeAg-positive chronic hepatitis B patients. D. Jiang, H. Chen, B. Zhou, X. Liang, R. Fan, J. Sun, J. Hou. Infectious Diseases and Hepatology Unit, Nanfang Hospital, Southern Medical University, Guangzhou, China.

Objective Nucleos(t)ide analogues (NUCs) and peginterferon alfa (PegIFN- α) are currently the main two first-line therapies for chronic hepatitis B (CHB). Unfortunately, both of them produce response in only minority of CHB patients. Predictors of treatment response are expected to optimize treatment of CHB. Till present, 14 SNPs were reported to be associated with susceptibility of CHB. This study was to assess whether these SNPs are potential predictors of response to treatment of CHB. **Methods** A total of 1,860 HBeAg-positive CHB patients, including 254 patients (PegIFN α Cohort 1) and 708 patients (PegIFN- α Cohort 2) treated with PegIFN α , 334 patients (NUCs cohort 1) and 564 patients (NUCs cohort 2) treated with NUCs, were enrolled. Treatment efficacy was assessed by sustained virologic response (SVR), which was defined as HBeAg seroconversion along with HBV DNA level < 2000 IU/mL at week 72 for PegIFN α cohorts or at week 104 for NUCs cohorts. All 14 SNPs were genotyped by Sequenom MassArray platform. Cochran-Armitage trend test and logistic regression analysis were used to evaluate the association of SNPs with SVR. **Results** None of these SNPs was found to be significantly associated with SVR in both NUCs cohorts. But in both PegIFN α cohort 1 and 2, rs12416, which is a missense change of complement factor B (*CFB*), was significantly associated with SVR. That is, the SVR rate of patients with rs12416 CC genotype were 28.38%, 20.76% and 22.82% in PegIFN α cohort 1, 2 and their combination; while for patients with rs12416 TT/CT genotype, the SVR rate were only 5.26%, 6.82% and 6.35% in PegIFN α cohort 1, 2 and their combination (P-trend = 0.03, 0.02 and 0.002, respectively). This association remained significant even after adjusting baseline predictors and rs7574865 in *STAT4*, which was reported to be a reliable predictor of PegIFN α therapy in our previous study (Jiang DK et al., *Hepatology* 2016). We also found cumulative effect of rs12416 and rs7574865, i.e., the SVR rate decreased steadily as the number of unfavorable alleles of the two SNPs increasing. For example, in PegIFN α cohorts combination, the SVR rate of patients carried with 0, 1, 2, 3 and 4 unfavorable alleles were 32.93%, 24.23%, 18.71%, 3.13% and 0.00%, respectively (P-trend = 0.0001). **Conclusion** For HBeAg-positive CHB patients with PegIFN α therapy, rs12416 is an independent predictor of SVR. Moreover, by integrating rs12416 with rs7574865, we can further discriminate SVR patients from Non-SVR patients.

2627F

High frequency of PM phenotype in pharmacogenomic actionable genes: Results support the need for pre-emptive genotyping across global populations to improve therapy. D.K. Sanghera¹, C. Bejar¹, G.S. Wander², S. Ralhan². 1) University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma, USA; 2) Hero DMC Heart Institute, Ludhiana, India.

Diversity in drug response is attributed to both genetic and non-genetic factors. However, there is a paucity of pharmacogenetics information across ethnically and genetically diverse populations of India and the world. Here, we have analyzed 20 single nucleotide polymorphisms (SNPs) from 12 pharmacogenomics genes in Punjabi Sikhs of Indian origin (N= 1,616), as part of the Sikh Diabetes Study (SDS). We compared the allele frequency of poor metabolism (PM) phenotype among Sikhs across other HapMap populations. The PM phenotype of CYP1A2*1F for slow metabolism of caffeine and carcinogens was significantly higher in Indians (SDS 43%, GIH (Gujarati) 51%) compared to Europeans (CEU 29%) (p=0.01). Similarly, Indians had a significantly higher frequency of CYP2C9*3 (11% SDS, 13% GIH) vs. 5% in CEU (p= 0.008) and 'T' allele of CYP4F2 (36%) SDS, (43%) GIH vs. (21%) CEU (p=7.8x10⁻⁷); both associated with high risk of bleeding with warfarin. Both Sikhs (0.36) and Gujaratis (0.34) had a higher frequency of the NAT2*6 allele (linked with slow acetylation of isoniazid) compared to Europeans (0.27). Additionally, the low activity 'C' allele of MTHFR (rs1801131) was highest in Sikhs compared to all other ethnic groups [SDS (44%), GIH (39%) and CEU (34%) (p=4.9x10⁻³)]. SNPs in the MTHFR affect the metabolism of statins, 5-fluorouracil, and methotrexate-based cancer drugs. These findings underscore the need for establishing a global benchmark for pre-emptive genotyping of drug metabolizing genes before beginning therapeutic intervention. **Acknowledgments** Authors thank all the participants of AIDS/SDS who made this study possible. This work was supported by NIH grant (R01DK082766) funded through the NIDDK, and Harold Hamm Diabetes Center, and Presbyterian Health Foundation of Oklahoma grants. Technical assistance provided by Bishwa Sapkota is acknowledged.

2628W

Pharmacogenomic-associated medication prescribing trends in the Medicare and Medicaid population: Data-driven pharmacogenomic subpanel design. *E.R. Scott^{1,2}, A.O. Owusu Obeng^{2,3}, M.R. Botton^{1,2}, R.L. Wallsten¹, A.J. Chen¹, E.E. Schadt^{1,2}, S.A. Scott^{1,2,3}*. 1) Sema4, a Mount Sinai venture, Stamford, CT 06902; 2) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029; 3) Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029.

Pharmacogenomic (PGx) associations with high levels of evidence include hundreds of genomic variants and unique medications across dozens of disease states, which results in PGx-associated medications being prescribed by a broad array of medical specialties. To better characterize PGx-associated prescribing practices across the U.S. and potentially employ these data to inform on clinical PGx panel design, publicly available Centers for Medicare & Medicaid Services (CMS) Part B and Part D provider prescribing data from 2014-2016 were interrogated. Analyses were restricted to physicians with approved reimbursement claims for at least 50 CMS beneficiaries ($n > 450,000$), and the percentage of medicare beneficiaries prescribed a PGx-associated medication was calculated for each physician-medication pair. PGx prescribing rates were calculated by summing all PGx-associated medication prescriptions and dividing by a physician's total unique beneficiaries. Summary statistics of the identified PGx prescribing rates were assessed based on physician specialty, zip code, state, pharmacogene, and Clinical Pharmacogenomic Implementation Consortium (CPIC) evidence classification. CMS data recorded more than 28,000,000 CPIC Level A and 30,000,000 CPIC Level B PGx-associated medication prescriptions in each of the three years analyzed. Importantly, ~60% of physicians prescribed at least one of the 77 CPIC Level A (average of 2.9 ± 2.4) or Level B (average of 3.3 ± 2.86) medications each year. PGx prescribing rates varied significantly across physician specialty ($p < 2e-16$, MANOVA), with the highest rates observed in General Practice, Internal Medicine, and Family Practice. In addition to clustering groups of medical specialties by PGx prescribing rates, these data allowed for data-driven construction of novel PGx subpanels for 20 medical specialties. Moreover, choropleth maps of physician specialty prescribing rates for pharmacogene medication groups revealed dramatic variation across states, with Puerto Rico, Hawaii, and Oregon registering the greatest PGx-prescribing intensity. Taken together, our analysis of national provider prescribing data detected pervasive prescription of PGx-associated medications with clustered variation across geographic territories, medical specialties, and pharmacogenes. Importantly, these results also enabled the design of novel PGx gene panels based on medical practice specialty, which will likely have utility for PGx clinical implementation. .

2629T

Phenotypic and pharmacogenomics identification of intraoperative phenylephrine responses using electronic health records. *Y. Zhang¹, M.S. Poler², V. Abedi³, S.A. Perdergrass⁴, M.T.M. Lee¹*. 1) Genomic Medicine Institute, Geisinger, Danville, PA; 2) Department of Anesthesiology, Geisinger, Danville, PA; 3) Biomedical Translational Informatics Institute, Geisinger, Danville, PA; 4) Biomedical Translational Informatics Institute, Geisinger, Rockville, MD.

Background: Patients treated with phenylephrine for hypotension may have substantial differences in blood pressure response. Genetic variation may contribute to this variable response to phenylephrine. In this study we used clinical data with measures of phenylephrine response from patients undergoing surgery and receiving anesthesia to identify genetic associations with phenylephrine response. **Methods:** Data used to define the phenylephrine response were retrieved from electronic health record. A total of 4027 subjects were available for study after data cleaning. Unsupervised machine learning was used to identify sub-populations of patients treated with phenylephrine based on infusion rate and systolic blood pressure during infusion. We performed GWAS on the subset of patients with genetic data using average infusion rate, as well as with the patient sub-groups identified through machine learning. Pathway analyses were also performed. **Results:** We identified three response sub-groups of patients to phenylephrine treatment: resistant, normal, and sensitive responders. Though resistant responders had higher incidence of potentially confounding conditions such as diabetes and heart failure, the overall comorbidity rates were low in all three groups. Subsequent GWAS identified 12 loci with P -value $< 1 \times 10^{-5}$ associated with phenylephrine response. Vasoconstriction pathways were enriched (P -value < 0.05) in the pathway analysis. **Conclusions:** Novel loci associated with average infusion rate and patient subsets were identified by GWAS. Our study highlights the value of integrative analysis to provide insights into the variability of response to phenylephrine treatment that could lead eventually to precision medicine strategies and new care paths for patients being treated with phenylephrine.

2630F

Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. A.V. Khera^{1,2}, M. Chaffin¹, K.G. Aragam^{1,2}, M. Haas¹, C. Roselli¹, S.H. Choi¹, P. Natarajan^{1,2}, E.S. Lander¹, S.A. Lubitz^{1,2}, P.T. Ellinor^{1,2}, S. Kathiresan^{1,2}. 1) Broad Institute of MIT and Harvard, Cambridge, MA; 2) Massachusetts General Hospital, Boston, MA.

A key public health need is to identify individuals at high risk for a given disease to enable enhanced screening or preventive therapies. Because most common diseases have a genetic component, one important approach is to stratify individuals based on inherited DNA variation. Proposed clinical applications have largely focused on finding carriers of rare monogenic mutations at several-fold increased risk. Although most disease risk is polygenic in nature, it has not yet been possible to use polygenic predictors to identify individuals at risk comparable to monogenic mutations. Here, we leverage recently developed algorithms and large population biobanks to develop and validate genome-wide polygenic scores for five common diseases. The approach identifies 8.0%, 6.1%, 3.5%, 3.2% and 1.5% of the population at greater than three-fold increased risk for coronary artery disease (CAD), atrial fibrillation, type 2 diabetes, inflammatory bowel disease, and breast cancer, respectively. For CAD, this prevalence is 20-fold higher than the carrier frequency of rare monogenic mutations conferring comparable risk. Moreover, practicing clinicians could not identify the 8% of individuals at >3-fold risk based on GPS_{CAD} using conventional risk factors in the absence of genotype information. We propose that it is time to contemplate the inclusion of polygenic risk prediction in clinical care and discuss relevant issues.

2631W

A genome-wide association meta-analysis of the nicotine metabolite ratio and five other smoking related traits in smokers of European descent. J. Buchwald¹, M. Chenoweth², T. Palviainen¹, G. Zhu³, S. Gordon³, C. Benner⁴, T. Korhonen¹, S. Ripatti^{1,4,5}, P. Madden⁶, T. Lehtimäki^{7,8}, O. Raitakar^{9,10}, V. Salomaa¹¹, R. Rose¹², M. Pirinen^{1,4,13}, N. Martin³, J. Kaprio^{1,4}, R. Tyndale², A. Loukola¹. 1) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; 2) Campbell Family Mental Health Research Institute, CAMH, and Departments of Pharmacology & Toxicology and Psychiatry, University of Toronto, Toronto, Canada; 3) QIMR Berghofer Medical Research Institute, Brisbane, Australia; 4) Department of Public Health, University of Helsinki, Helsinki, Finland; 5) Broad Institute of MIT and Harvard, Cambridge, United States; 6) Department of Psychiatry, Washington University School of Medicine, Saint Louis, United States; 7) Department of Clinical Chemistry, Fimlab Laboratories, and Finnish Cardiovascular Research Center, Tampere, Finland; 8) Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland; 9) Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland; 10) Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland; 11) Department of Health, National Institute for Health and Welfare, Helsinki, Finland; 12) Department of Psychological and Brain Sciences, Indiana University, Bloomington, United States; 13) Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland.

Smoking is the main preventable cause of mortality. Smoking behaviour is influenced by the nicotine metabolism rate, which varies among individuals and is highly heritable. Understanding its genetics will facilitate optimizing personalised smoking cessation interventions. We present the largest (N=5185) genome-wide association study (GWAS) to date of a biomarker of nicotine metabolism rate, the nicotine metabolite ratio (NMR, ratio of 3-hydroxycotinine to cotinine). In our sample of Finnish, North American and Australian smokers, we also ran GWASs for two self-reported tobacco exposure measures (cigarettes smoked per day (CPD) and pack-years (CPD x years smoked)), two objective exposure biomarkers (plasma cotinine (Cot), and a more sensitive biomarker constructed as the sum of cotinine and 3-hydroxycotinine (Cot+3HC)), and one biomarker of smoking intensity (Cot/CPD). Cot and 3HC were determined by liquid chromatography/mass spectrometry. To our knowledge, these represent the first GWASs of Cot/CPD and Cot+3HC. In total, we found 1902 genome-wide significant SNPs located in seven distinct loci. The table below shows for each locus and phenotype: - the top SNP and p-value - the relevant gene - whether the top SNP was a known eQTL or meQTL - the total phenotypic variance explained (PVE) by the shown top SNPs - the total number of SNPs (n(Credible set)) with a direct effect on the phenotype (via fine-mapping analyses using step-wise conditional regression with GCTA and a shotgun stochastic search with FINEMAP (www.christianbenner.com))

chr	NMR	CPD	PhenotypesPack-years	Cot	Cot+3HC	Cot/CPD
1		rs860873 4e-8 CNN3 eQTL				
4	rs34638591 1e-12 TMPRSS11E eQTL			rs294775 4e-20 UGT2B10	rs294775 5e-21 UGT2B10	rs294778 7e-16 UGT2B10
5		rs2337033 3e-8 TENM2	rs2337033 7e-9 TENM2			
9					rs12684930 2e-8 SMARCA2	
15		rs72740955 7e-12 CHR-NA5 eQTL meQTL	rs72740955 6e-10 CHR-NA5 eQTL meQTL	rs2036527 4e-27 CHR-NA5 eQTL meQTL	rs2036527 8e-28 CHR-NA5 eQTL meQTL	
19	rs56113850 5e-259 CY-P2A6 eQTL meQTL			rs2316205 3e-11 CY-P2A6 eQTL meQTL		rs56113850 8e-14 CYP2A6 eQTL meQTL
PVE (%)	24.8	1.5	1	4.8	4.6	2.6
n(Credible set)	11	3	2	3	3	2

The chr 4 association for NMR is novel, and *TMPRSS11E* has previously only been reported in GWASs of corpuscular haemoglobin and volume. This is also the first GWAS to find associations for addiction phenotypes in *CNN3*, *TENM2* and *SMARCA2*, among others. Our study reveals novel genetic loci associated with smoking-related traits and sheds light on the genetic similarities and differences between nicotine metabolism and tobacco exposure phenotypes.

2632T

A multi-trait GWAS (MTAG) for puberty timing in men. *F.R. Day¹, B. Hollis¹, P. Timmers², P. Joshi², D. Thompson³, A.S. Busch^{1,4}, K.K. Ong¹, J.R.B. Perry¹, 23andMe.* 1) MRC Epidemiology Unit, Cambridge, UK; 2) Centre for Global Health Research, Usher Institute for Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, Scotland; 3) Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, UK; 4) Department of Growth and Reproduction, Juliane Marie Centre, Copenhagen, Denmark.

Background The timing of puberty is highly variable and has important consequences for long-term health. Most genetic studies have been performed in women, while studies in men are few and smaller in scale. **Methods** Here, we combine genome-wide association study (GWAS) statistics for related recalled male puberty timing traits from the 23andMe (continuous age at voice breaking) and UK Biobank (relatively early/late age at voice breaking and relatively early/late age of first facial hair) studies, generating an effective sample size of 205,354 men, nearly four-fold larger than previously. **Results** We identify 78 GWAS signals for male puberty timing, including 29 not previously associated with puberty timing in either sex, and 5 with significant heterogeneity in their effects on puberty timing between sexes ($P_{\text{heterogeneity}} < 6.4 \times 10^{-4}$) in/near: *ENDOD1*, *LEPR*, *TMEM18*, *TMEM38B* and *TMEM102*. Novel signals include three previously identified for natural hair colour (*HERC2*, *IRF4*, and *C16orf55*). Systematic testing across 124 reported hair colour signals demonstrate a stronger causal influence of darker natural hair colour on earlier puberty timing in males ($P = 6.3 \times 10^{-43}$) than females (age at menarche: $P = 2.2 \times 10^{-3}$), which supports similar phenotypic associations, and possibly reflects common effects of pituitary hormones in pigmentation and puberty timing. Mendelian randomization analyses infer a robust effect of higher body mass index (BMI) on earlier puberty timing in males (weighted median test: $B = -0.259$ years/SD, $P = 1.4 \times 10^{-16}$) but weaker than that in females ($B = -0.623$ years/SD, $P = 6.7 \times 10^{-47}$), consistent with the more variable phenotypic relationship in males than females. Additionally, we identify several negative genetic correlations between male puberty timing and health-related outcomes, including diabetes, hypertension, smoking and alcohol intake. Mendelian randomization analyses infer a causal relationship between later male puberty timing and greater lifespan, corresponding to 7 months per year later puberty (self-hazard protection ratio = 0.061, $P = 7.0 \times 10^{-5}$). **Conclusions** Puberty timing in males and females show only moderately overlapping genetic architectures ($r_g = 0.68$, $P = 2.6 \times 10^{-213}$). Our findings demonstrate the utility of MTAG to combine data across studies with related measures to provide novel insights into the regulation and consequences of puberty timing.

2633F

A genetic atlas for reproductive health and disease in UK Biobank. *B. Hollis, A.S. Busch, F.R. Day, K.K. Ong, J.R.B. Perry.* MRC Epidemiology Unit, University of Cambridge, Cambridge, United Kingdom.

Background Reproductive health and fertility vary considerably within populations and impact on social, economic and clinical outcomes. Despite this, progress identifying the genetic determinants and mechanisms driving this variability has lagged behind that of other complex traits. To help address this, we have performed the first large-scale, systematic assessment of genetic determinants of reproductive traits in half a million individuals from the UK Biobank study. **Methods** We performed genome-wide association studies for 175 reproductive quantitative traits and disease outcomes in the UK Biobank study (max $N = 467,522$). Individual traits were curated using a combination of hospital in-patient episode statistics and self-reported data. Analyses were conducted using linear mixed models and our threshold for individual variant significance reflected the total number of tests performed ($P < 2.85 \times 10^{-10}$). **Results** A total of 870 SNP-trait associations reached significance across 735 distinct genomic regions. We identify at least one associated variant for 56 of the 175 traits (32%), including the first identified genetic determinants for many traits. While many of the associations were specific to particular diseases, some SNPs showed widespread associations with clinically relevant endpoints; rs11031005 near *FSHB* was associated with nine of the assessed outcomes including endometriosis, excessive and frequent menstruation and increased chance of having had an oophorectomy; rs78378222 in *TP53* was also associated with oophorectomy as well as having had a hysterectomy. We were also able to identify other loci such as the *ESR1*/*SYNE* locus (encoding estrogen receptor-1) associated with bilateral oophorectomy and endometriosis, uterine fibroids and breast cancer, and *WNT4* (endometriosis, bilateral oophorectomy, uterine fibroid and novel association with pelvic organ prolapse) where there were an number of index SNPs across the phenotypes. Several loci also showed important effects on sex-specific phenotypes, for example *SCGB1C1* which is associated with prostate hyperplasia and uterine fibroids/polyps. **Conclusions** Our study represents the first large-scale, multi-trait assessment of the genetic basis of reproductive health, identifying hundreds of novel associated genetic variants and genes. We aim to develop this into a community resource 'ReproPhen' to enable investigators from diverse biomedical disciplines to utilise these genetic data.

2634W

P2RX7 functional variants in localized aggressive periodontitis. T. Harris^{1,2}, M. Wallace^{3,4}, H. Huang¹, H. Li³, A. Mohiuddeen¹, Y. Gong⁵, S. Walle⁶, T. Kompotiati¹, P. Harrison¹, I. Aukhil¹, L. Shaddox^{1,2}. 1) Dept. of Periodontology, University of Florida College of Dentistry, Gainesville, FL, USA; 2) Dept. of Oral Biology, University of Florida College of Dentistry, Gainesville, FL, USA; 3) Dept. of Molecular Genetics and Microbiology, University of Florida College of Medicine, Gainesville, FL, USA; 4) University of Florida Genetics Institute, Gainesville, FL, USA; 5) . Dept. of Pharmacotherapy and Translational Research, University of Florida College of Pharmacy, Gainesville, FL, USA.

The purpose of this study is to investigate the involvement of the P2X7 receptor in the inflammatory response in localized aggressive periodontitis (LAP). Peripheral blood was obtained from 220 African-Americans (103 with LAP and 117 Healthy patients). Four P2RX7 SNPs encoding non-synonymous changes (rs1718119, rs2230911, rs208294, and rs3751143) were genotyped from DNA, and whole blood was stimulated with lipopolysaccharide (LPS) from *E.coli* (Ec) and *porphyromonas gingivalis* (Pg) followed by Luminex assay of secreted chemokines/cytokines. P2RX7 SNP frequencies were compared between both populations and genotypes tested for correlation with LPS-stimulated cyto/chemokines levels. Three P2RX7 SNPs were in perfect linkage disequilibrium, with a total of 4 haplotypes and 9 diplotypes identified among all patients. Diplotypes were grouped based on known haplotype functional effects: wild-type (WT), net gain of function, net loss of function, and mixed effect. These four diplotype group frequencies were compared within and between the two patient populations (LAP and controls), and tested for association with levels of LPS-stimulated cytokines and chemokines. In another study, P2RX7 was compared to the cytokine/chemokine data based on number of ancestral alleles at each SNP. To measure the effect of extrinsic activation of the receptor, peripheral blood samples from 41 African-Americans (21 with LAP and 20 Healthy unrelated controls) were stimulated under three different treatments; *E.coli* LPS alone, LPS+ATP, and LPS+ATP+P2X7 selective inhibitor. Levels of IL-2, IL-1 β , and IL-12p40 were then measured. Stepwise regression found significant genetic association of the rs1718119 A allele with LAP versus healthy controls ($p=0.0082$). Correlations were found between three P2XR7 SNPs and LPS stimulated levels of several cyto/chemokines ($P<0.05$). Significant cyto/chemokine differences were found between the homozygous ancestral allele group and the hetero-/homozygous non-ancestral allele group for all SNPs. A number of associations with cyto/chemokines were found both between and within the diplotypes for all patients ($P<0.05$). The P2X7 activation assays showed statistically significant differences for IL-1B levels in both LAP and healthy groups based on P2RX7 genotypes, although there was no difference between the two populations. This is the first study to detect evidence of a biological relationship between P2RX7 genetic variants and the inflammatory response in LAP.

2635T

Genome-wide association study reveals sex-specific genetic architecture of facial attractiveness. B. Hu¹, N. Shen¹, J. Li^{2,3}, H. Kang¹, J. Hong³, J. Fletcher^{4,5}, J. Greenberg^{3,6}, M. Mallick^{3,6}, Q. Lu^{1,7}. 1) Department of Statistics, University of Wisconsin-Madison, Madison, WI; 2) Department of Psychology, University of Wisconsin-Madison, Madison, WI; 3) Waisman Center, University of Wisconsin-Madison, Madison, WI; 4) La Follette School of Public Affairs, University of Wisconsin-Madison, Madison, WI; 5) Department of Sociology, University of Wisconsin-Madison, Madison, WI; 6) School of Social Work, University of Wisconsin-Madison, Madison, WI; 7) Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI.

Facial attractiveness is a fascinating and complex human trait of great interest in both academia and industry. It has profound influence on human behavior and literature on its sociological and psychological implications is rich. However, although twin-based studies have suggested substantial heritability for facial attractiveness, its genetic basis remains poorly understood, partly due to the scarcity of well-phenotyped attractiveness data with matched genotype information. In this study, we report the first genome-wide association study (GWAS) of facial attractiveness based on 3,928 samples in the Wisconsin Longitudinal Study (WLS). Facial attractiveness in WLS was measured based on each individual's high school yearbook photo by 12 raters (six females and six males). We stratified our association analyses by sex (of samples and raters) to better understand the distinct genetic architecture of male and female attractiveness as well as its perception among males and females. We identified two genome-wide significant loci on 10q11.22 (rs2999422, $p=9.2e-10$) and 2p22.2 (rs10165224, $p=3.4e-8$), and highlighted more than a dozen candidate genes via eQTL analyses and cross-tissue transcriptome-wide association mapping. Sex-specific genetic architecture of facial attractiveness was a recurrent pattern observed in all our analyses. Sex-stratified GWAS analyses identified distinct lists of associated loci. Additionally, several top candidate genes (e.g. *ANTRXL*) were specifically expressed in human tissues involved in reproduction and hormone synthesis, a pattern that was further confirmed via heritability enrichment analysis - testis and ovary were the most enriched tissues for male and female attractiveness, respectively (enrichment=3.0~5.7 across analyses). Further, we identified potentially pleiotropic loci shared with body-mass index (BMI), facial morphology, skin pigmentation, and hair color. Interestingly, only female attractiveness (especially when rated by males) showed strong and negative genetic correlation with BMI ($p=4.7e-5$) while male attractiveness was more correlated with blood lipid levels ($p=5.1e-4$) which are known to be involved in testosterone synthesis. Finally, we demonstrated that the selection pressure on negative associations of male attractiveness was substantially stronger than in females. These results revealed sex-specific genetic architecture of facial attractiveness and provided fundamental new insights into its genetic basis.

2636F

Association of the NTCP S267F variant disabling of HBV receptor function with a reduced host risk to HBV infection and disease progression. P. An¹, Z. Zeng², C.W. Winkler¹. 1) Basic Research Laboratory, Center for Cancer Research, National Cancer Institute, Basic Science Program, Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research, Frederick, MD; 2) Department of Infectious Diseases, Peking University First Hospital, Beijing, China.

Sodium taurocholate cotransporting polypeptide (NTCP) is a cellular receptor for entry of Hepatitis B virus (HBV). The natural S267F variant in NTCP specifically present in the Asian population attenuates HBV receptor function of NTCP. We assessed the impact of S267F on host susceptibility to HBV resistance, HBV infection clearance, and HBV-related cirrhosis and hepatocellular carcinoma (HCC) in 1117 Han Chinese patients with distinct HBV infection outcomes using multivariate logistic regression analysis. The 267F variant allele prevalence was elevated in the group of HBV-resistance healthy controls compared with HBV-infected patients (odds ratio (OR) 0.32, 95% confidence interval (CI) 0.15-0.68, dominant model). The 267F variant genotypes were also associated with reduced risk for cirrhosis (OR 0.15, 95% CI 0.03-0.70) or HCC (OR 0.21, 0.06-0.72), when compared with those with chronic HBV infection. There was no association of the S267F variant with spontaneous HBV clearance. We demonstrated for the first time that NTCP S267F variant does not act synergistically with *HLA-DPA1* rs3077 variant, an HBV infection restriction factor, to modify HBV infection and chronic liver disease progression. In summary, carriage of the loss-of-function S267F variant in the HBV hepatocyte receptor NTCP confers innate resistance to HBV infection as well as restriction of progression to cirrhosis and liver cancer among those with chronic HBV infection. Our population genetic data support an *in vivo* role of NTCP as the primary entry receptor of HBV. (Funded by the National Cancer Institute Contract HHSN261200800001E).

2637W

Genotype imputation of structural polymorphism with whole genome sequencing based haplotypes reference panel. Y. Kawai¹, T. Mimori², K. Ueno³, Y. Hitomi⁴, O. Gervais⁵, S.S. Khor¹, M. Kawashima^{1,3}, N. Nishida^{1,4}, M. Nakamura⁵, M. Nagasaki², K. Tokunaga¹. 1) Department of Human Genetics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; 2) Department of Integrative Genomics, Tohoku Medical Megabank Organization, Tohoku University, Sendai, Japan; 3) Japan Science and Technology Agency (JST), Tokyo, Japan; 4) The Research Center for Hepatitis and Immunology, National Center for Global Health and Medicine, Ichikawa, Japan; 5) Clinical Research Center, National Hospital Organization (NHO) Nagasaki Medical Center, Omura, Japan.

Genotype imputation based on the large haplotype reference has been successful in genome-wide association studies (GWASs) to date. Prediction of missing genotypes by genotype imputation not only boosts the power of GWAS but also enables to identify causal variants in post-GWAS analyses. Although the genotype imputation is routinely used in GWAS, the majority of studies have focused on the imputation of single nucleotide polymorphisms (SNPs). However, the structural variants (SVs) such as copy number variations (CNVs) and mobile element insertions (MEIs) are source of the genetic diversity among human populations and their variations can contribute phenotypic variations of complex traits. We discovered the SVs from the whole genome sequencing data of a general population in Japan. We found 17,314 CNVs, 7,072 MEIs and 869,590 short tandem repeats from 418 individuals. The SVs were phased together with tag SNPs, in which their probes were on SNP arrays used in GWAS, to conduct the genotype imputation. We evaluated the accuracy of the SV imputation by cross validation. The imputation accuracy tends to be higher in high frequency SVs as it has been reported for SNP imputation. We estimated a heritability of primary biliary cholangitis (PBC) explained by the imputed SVs. We found a small but significant fraction (~3%) of heritability of PBC can be accounted for by the SVs which were likely overlooked in previous studies which were estimated by SNPs.

2638T

Imputing the effects of genetic variation on mRNA splicing in rare human samples. A. Shah¹, Y.I. Li^{2,3}. 1) Committee on Genetics, Genomics, and Systems Biology, University of Chicago, Chicago, IL; 2) Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, IL; 3) Department of Human Genetics, University of Chicago, Chicago, IL.

Over 90% of all disease-associated variants identified in genome-wide association studies (GWAS) are located outside protein coding regions, suggesting that these variants have gene regulatory function. While much of the work understanding disease-associated variants has focused on mapping the effects of genetic variation on gene expression levels, we have shown previously that mRNA splicing is the primary mechanism that links genetic variation to disease (Li et al., 2016), such that a substantial number of GWAS hits may affect disease risk by impacting RNA splicing. To better understand how variation in RNA splicing affects disease, we investigated the impact of genetic variation on RNA splicing in a large collection of 53 tissues from the Genotype-Tissue Expression (GTEx) project. Using a novel splicing quantification method, LeafCutter (Li et al., 2018), we identified over forty thousand unannotated splicing events across all tissues and mapped thousands of splicing quantitative trait loci (sQTLs) in each tissue. We found that 10-20% of splicing events show tissue specific patterns of inclusion, and that up to 40% of testis-specific splicing events were not observed in any other tissue. We found that sQTLs were enriched in nearly all GWAS traits examined, including autoimmune and neurodegenerative diseases. Notably, most sQTLs did not affect gene expression levels, confirming the widespread importance of measuring RNA splicing alongside expression. Finally, we quantified the concordance of sQTL effects on splicing events that are present across GTEx tissues using flashr, a statistical method that accurately estimates sharing of genetic effects on multiple phenotypes and/or samples (Wang and Stephens, 2018). Strikingly, our analysis revealed that 95% of sQTLs are shared across samples, compared to 70% for expression quantitative trait loci (eQTLs). The high amount of shared genetic effects on RNA splicing across tissues offers the possibility of accurately imputing the effects of genetic variation on RNA splicing in rare, and as of yet, inaccessible human samples. For example, using the imputed effects of genetic variants collected from large neuropsychiatric disease GWASs, we can study the impact of genetic variation on RNA splicing in developing fetal brains.

2639F

Mosaic loss of chromosome Y (LOY) in leukocytes helps explain why men live shorter lives. L.A. Forsberg^{1,2,3}, J. Halvardson¹, M. Danielsson¹, J. Mattisson¹, J. Dumanski^{1,2}, B. Torabi Moghadam¹, H. Davies¹. 1) Dept. of Immunology, Genetics and Pathology, Uppsala University, Sweden; 2) Science for Life Laboratory, Uppsala University, Sweden; 3) Beijer Laboratory of Genome Research, Uppsala University, Sweden.

Men live shorter lives, but the reason(s) why has been poorly understood. Recent analyses show that mosaic loss of chromosome Y (LOY) in leukocytes of men is associated with increased risk for all-cause mortality, mortality and diagnosis of non-hematological cancer, increased risk for diagnosis with Alzheimer's disease (AD), autoimmune conditions and major cardiovascular events. Published data show that LOY is the most common human mutation and unpublished data show that more than half of 93-year-old men harbor LOY-cells in their blood. Known risk factors for LOY include age, smoking as well as inherited SNP-variants. Hence, as a common and male specific genetic risk factor, LOY helps explain why men in the entire world on average dies about 5 years younger compared to women. How could LOY in leukocytes increase the risk for disease in other organs? One hypothesis to explain the associations between LOY in blood cells with increased risk for disease in other tissues, is that some normal functions of immune cells, such as orchestrating defense against and eradication of various disease and malignant processes, could be compromised after being affected with LOY. To address this question, we study the functional consequences of LOY in leukocytes using various tools. For example, we investigate dysregulated gene expression as an effect of LOY in single cell transcriptomic assays (10X Chromium and ddSEQ platforms) and well as in FACS-sorted immune cell fractions with AmpliSeq. Moreover, LOY mediated changes in plasma proteins are studied using PEA. Our preliminary data support the hypothesis that LOY could influence important immune cell properties and functions. For example, unpublished single cell RNA data shows that the gene expression of immune cells with LOY is profoundly dysregulated in a cell type specific and functionally relevant manner. Furthermore, new analyses of FACS-sorted cells show that cancer and AD patients may have LOY in different types of immune cells. LOY in leukocytes is associated with increased risk for death and major causes of male morbidity and mortality. LOY could therefore become a new and predictive biomarker for identification of men with increased risk for disease. To improve the clinical utility of LOY we are developing new tools for improved LOY-detection. Identification of men with LOY in leukocytes could lead to earlier diagnoses of ongoing disease in affected men, and thus, open up for better medical treatment options.

2640W

Investigating the impact of smoking on mental and physical health outcomes using genome-wide data.

C. Chen^{1,2,3}, *R. Walters*^{1,3}, *L. Abbott*⁶, *D. Pulmer*^{1,3}, *D. Howrigan*^{1,3}, *A. Ganna*^{1,3}, *C. Churchhouse*^{1,3}, *C.E. Carey*^{1,3}, *J. Smoller*^{2,3}, *B. Neale*^{1,2,3}. 1) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 2) Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, MA; 3) Broad Institute of MIT and Harvard.

The increasing amount of publicly available genome-wide association study (GWAS) summary results and biobanks that cover a wide range of human traits linked to genome-wide data have enabled us to comprehensively investigate the relationships between mental and physical health outcomes at an unprecedented scale. Here we aim to examine the impact of smoking, measured as average smoking pack-year over adult lifetime (ASP), on a wide range of health outcomes and traits using genome-wide data. We leverage data from existing large-scale GWAS summary statistics as well as individual level phenotype and genotype data from the UK Biobank. We performed genome-wide association analysis on ASP using 101,726 subjects from the UK Biobank with 10,889,592 imputed SNPs after QC. The ASP was derived as average number of smoking pack-years by an individual each year after age of 16. The association analysis was done with linear regression adjusted for sex and the first 10 PCs. The genome-wide association analysis identified 13 independent genome-wide significant loci, with λ_{GC} of 1.22 and LD score regression intercept of 1.017 (SE = 0.0081; ratio = 0.067 with SE = 0.0305) indicating high polygenicity with minimal inflation due to confounding. The SNP-based heritability of ASP was 0.124 (SE = 0.0083; $p = 1.06E-50$). Using the association summary statistics, we first explored the genetic correlation (r_g) between ASP and 230 phenotypes from publicly available GWAS summary statistics and 496 phenotypes from UK Biobank using LDhub. We identified 296 significant r_g after Bonferroni correction. The most significant results include self-reported poor health ($r_g = 0.631$; $p = 3.35E-118$) and years of schooling ($r_g = -0.574$; $p = 2.48E-92$). We further explored the causal effects of ASP on 40 health outcomes and traits using bidirectional 2-sample Mendelian randomization (MR). Multiple methods were used to detect and control for potential pleiotropic biases in the MR analysis, including MR-PRESSO, MR-Egger regression, and weighted median-based MR. The most significant result is that higher years of schooling decrease ASP, which also shows strong pleiotropic effects ($\beta = -0.375$; $p = 7.44E-10$; MR-PRESSO pleiotropy test $p < 0.001$; N of SNP = 70). Our analyses highlighted that smoking has complex relationships with a wide range of mental and physical health outcomes and traits.

2641T

Visualization of UK Biobank genetic association results in PheWeb.

P. VandeHaar, *S.A. Gagliano*, *G.R. Abecasis*. Department of Biostatistics and Center for Statistical Genetics, University of Michigan School of Public Health, Ann Arbor, MI.

The UK Biobank cohort consists of >400,000 individuals, >2000 traits, and >39 million HRC-imputed or genotyped variants. Manually sifting through all of the resulting genetic association results, variable by variable, is neither an efficient nor comprehensive way to foster discoveries. Efficiently translating a massive amount of results into new understandings of health and disease requires making the information accessible to researchers in an easy-to-use manner. We use PheWeb, our web-based tool for visualizing and browsing through hundreds or thousands of genetic association results, to visualize the UK Biobank results. This instance of PheWeb provides a comprehensive view of association results for the largest genotyped population cohort. These results are based on the SAIGE generalized mixed model association test, which accounts for case-control imbalance and relatedness. PheWeb's visualization capabilities are particularly useful for studies with many phenotypes, and users can build a PheWeb of their own data by following instructions on GitHub (<https://github.com/statgen/pheweb>). We are continually implementing new features to aid in data analysis, such as the ability to conduct gene-burden tests on the fly, selection of customized panels for linkage-disequilibrium, and highlighting known GWAS loci. The UK Biobank PheWeb is freely available for all researchers at <http://pheweb.org/UKBiobank>.

2642F

Does the *CPT1A* p.P479L variant increase risk for early neonatal hypoglycemia? S. Collins^{1*}, E. Hildes Ripstein^{2*}, B. Thompson³, K. Barker⁴, C. Greenberg², L. Arbour¹, * indicates shared first authorship. 1) Dept of Medical Genetics, University of British Columbia, Victoria BC Canada; 2) Dept of Pediatrics and Child Health, University of Manitoba, Winnipeg MB Canada; 3) Cadham Provincial Laboratory, Winnipeg MB Canada; 4) Dept of Health, Government of Nunavut, Canada.

Background: The *CPT1A* p.P479L variant, a possible risk factor for hypoglycemia, is overrepresented in Indigenous populations in the Arctic and coastal British Columbia. We assessed blood glucose (BG) levels to determine if full term Inuit neonates without other risk factors for hypoglycemia had an increased risk if either heterozygous or homozygous for the variant. **Methods:** Clinical charts were reviewed of 503 Inuit infants born to mothers residing in Kivalliq Nunavut between 2010 and 2014 (four years). BG levels in the first 12 hours of life were abstracted with additional variables (maternal diabetes mellitus, infant weight, gestation, feeding type and clinical course) and then linked to genotypes. We used standard neonatal BG thresholds to define hypoglycemia as per the Canadian Paediatric Society (<2.0 at 2hrs or <2.6mmol/L within 12hrs). Charts with missing genotypes (n=36) or risk factors for hypoglycemia (n=142) were excluded. **Results:** Genotype distribution was 4.9% wildtype (PP, n=16), 23.4% heterozygous (PL, n=76) and 71.7% homozygous (LL, n=233; total cohort=325). Mean BG was 2.63±0.07mmol/L and 26.5% had at least one hypoglycemic BG level within 12hrs of life. Infants with LL and PL genotypes had a higher proportion of documented hypoglycemia than infants with the PP genotype (29% and 22% vs 13%, respectively); however, the difference was not statistically significant (FET=0.266 and 2-wayANOVA F=0.28, p=0.753). Proportions of persistent hypoglycemia (2 or more consecutive hypoglycemic levels; 13% PP, 9% PL, 8% LL) and intravenous dextrose treatment (13% PP, 5% PL, 7% LL) did not significantly differ between genotypes (FET=0.579 and 0.420, respectively). Of note, approx. 14% of infants tested at 0-2hrs (n=238) had a BG level <2.0mmol/L, which was equally distributed within genotypes (FET=0.612). **Conclusion:** There was no statistical significant difference in documented early neonatal hypoglycaemia between genotypes of the *CPT1A* allele. Although not significant, it was notable that 29% of homozygotes and 22% of heterozygotes without other neonatal risk factors for hypoglycemia had documented hypoglycemia. Hypoglycemia in wildtypes was similar to that expected for a healthy newborn population (13%). However, there was no difference in persistent hypoglycemia or IV dextrose treatment between the genotypes. Further study into whether the *CPT1A* p.P479L variant influences the risk of neonatal hypoglycemia, and under what circumstance, is indicated.

2643W

The common polymorphism of the *KISS1* gene promoter associated with endometriosis in Russian population. T. Ivashchenko, A. Shagina, N. Ganbarli, M. Yarmolinskaya. Oot's institute of obstetrics and gynecology, St.Petersburg, st.Petersburg, Russian Federation.

Endometriosis is one of the most common gynecological diseases. Etiology and pathogenesis of endometriosis remain obscure so far. It is usually treated as a complex multifactorial disease with obvious genetic predisposition, immunology failure and possible noxious environmental factor involvement. Kisspeptins, the product of the *KISS1* gene, plays an essential role in the regulation of reproductive functions, acting primarily at the hypothalamic level of the gonadotropic axis. It was shown, that the *KISS1* expression in the ectopic glandular endometrium was significantly increased as compared with the eutopic glandular endometrium from patients with and without endometriosis. We examined the common polymorphism of the *KISS1* gene promoter (rs5780218 (c.-145insT); rs3924587 (c.-89G>A)) in women with and without endometriosis by PCR-RFLP analysis. We observed significantly higher frequencies of c.-145insT alleles in the patients group compared to the control group (20% versus 9%). There was no significant difference in frequency alleles for rs3924587 of the *KISS1* gene. The results suggest that polymorphism c.-145insT of the *KISS1* gene associated with endometriosis in Russian population.

2644T

Genetic association of ER stress sensors IRE1 and XBP1 with cranial vault shape and nasal shape in humans. S. Weinberg¹, Y. Zhou¹, J. Roosenboom¹, M. Lee¹, J. Hecht², C. Heike³, G. Wehby⁴, K. Christensen⁵, L. Moreno⁶, E. Feingold⁶, M. Marazita¹, P. Claes⁷, J. Shaffer⁸. 1) Department of Oral Biology, University of Pittsburgh, Pittsburgh, PA, 15219; 2) Department of Pediatrics, University of Texas McGovern Medical Center, Houston, TX; 3) Department of Pediatrics, Seattle Children's Craniofacial Center, University of Washington, Seattle, WA; 4) Department of Health Management and Policy, University of Iowa, Iowa City, IA; 5) Department of Epidemiology, Institute of Public Health, University of Southern Denmark, Odense, Denmark; 6) Department of Orthodontics, University of Iowa, Iowa City, IA; 7) Katholieke Universiteit Leuven, Belgium; 8) Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA, 15261.

Introduction Endoplasmic reticulum (ER) unfolded protein response (UPR) pathway protein IRE1 (coded by the *ERN1* gene) and XBP1 play essential roles in bone formation and chondrogenesis, which are critical processes in craniofacial growth and development. For example, transgenic mice with heightened ER stress display pseudochondroplasia and cranial vault malformation. Whether genetic variants in *XBP1* and *ERN1* influence craniofacial variation in humans remains unknown. Therefore, we employed an *in silico* candidate gene approach testing the genetic association of common variants in *ERN1* and *XBP1*, with facial and cranial vault shape in a sample of healthy individuals. **Methods** 3D facial images were obtained from 2,329 participants in the 3D Facial Norms (3DFN) repository. Caliper measurements of the cranial vault were obtained from both the 3DFN study and the Pittsburgh Orofacial Cleft (POFC; N=2274) study. Whole-genome genotype data were collected for both studies using Illumina platforms, with imputation to the 1000 Genomes Study reference. 2259 single nucleotide polymorphisms (SNPs) with MAF > 1% at *ERN1* and 2562 SNPs on *XBP1* (+/- 200kb) were tested for genetic association with 63 multi-dimensional facial phenotypes in the 3DFN cohort and 3 cranial vault phenotypes in both 3DFN and POFC cohorts while adjusting for ancestry and pertinent covariates. **Results** We observed significant associations in the 5' flanking region of *XBP1* for Cephalic Index (rs35679096, p=0.00017). The associated SNP is in LD with several SNPs showing promoter and enhancer chromatin marks in many different cell types, including in bone-forming osteoblast lineage cells and mesenchyme stem cell-derived chondrocytes, which are cartilage-forming cells. We also detected associations between *ERN1* variant rs78885494 and several nasal shape phenotypes (p=5E-5 to 3E-6). This SNP is in strong LD with two other nose-associated SNPs, rs74940919 and rs8078549 (p=2E-5 to 9E-6), which are located in known promoter and enhancer regions in both chondrocyte and osteoblast lineage mesenchyme stem cells. **Conclusion** Although the mechanism remains unclear, our findings suggest that genetic variants at the *XBP1* and *ERN1* loci are associated with normal human cranial vault and nasal shape, respectively. Funding: R01-DE027023 U01-DE020078 R01-DE016148 .

2645F

The genetic basis of social effects: Genome-wide association study of social genetic effects on 170 phenotypes in laboratory mice. A. Baud¹, F.P. Casale^{1,2}, J. Nicod^{3,4}, O. Stegle^{1,5}. 1) European Molecular Biology Laboratory, European Bioinformatics Institute, Cambridge, UK; 2) Microsoft Research New England, Cambridge, Massachusetts, USA; 3) Wellcome Trust Centre for Human Genetics, Oxford, UK; 4) The Francis Crick Institute, London, UK; 5) European Molecular Biology Laboratory, Genome Biology Unit, Heidelberg, DE.

When two individuals such as peers or parent and offspring interact, heritable traits of one individual can influence the phenotype of the other, giving rise to associations between genotypes of one individual and phenotype of the other. Such associations are called social genetic effects (SGE), indirect genetic effects or "the nature of nurture". We previously demonstrated that SGE, in aggregate, affect a broad range of phenotypes in laboratory mice, and three studies of educational attainment in humans have shown that SGE are also at play in our species. In this study we took the next step and mapped SGE. Much like "traditional" direct genetic effects (DGE, effects of an individual's genotypes on their own phenotype), SGE are amenable to mapping and genome-wide association study of SGE (sgeGWAS) has the potential to greatly advance our understanding of how social partners influence each other's health. We carried out the first sgeGWAS in an outbred population, mapping SGE for 170 behavioral, physiological and morphological phenotypes in outbred laboratory mice. We developed a statistical framework for sgeGWAS that accounts for common confounding factors and identified 21 genome-wide significant SGE loci (FDR < 10%). In addition, for all 170 traits, we calculated the genetic correlation between DGE and SGE and mapped DGE (dgeGWAS). Our results provide the following insights into the mechanisms of social effects and their genetic architecture: first, SGE and DGE arise from partially different loci and/or loci with different effect sizes, which implies complex mechanisms for social effects going beyond "phenotypic contagion" or "spread" (e.g. my anxiety affects my partner's anxiety). Secondly, several genome-wide significant DGE associations but no SGE associations had large effect sizes, suggesting sgeGWAS is unlikely to uncover "low hanging fruits". Finally, our results suggest that a similar number of variants likely contribute to SGE and DGE. The framework we developed for sgeGWAS and the insights we gained from analysing 170 diverse phenotypes will inform the design, implementation and interpretation of sgeGWAS in humans. A better understanding of the role of the social environment could result in improved healthcare. Our study is available as a preprint: Baud, A., *et al.*, *Genome-wide association study of social genetic effects on 170 phenotypes in laboratory mice*. *bioRxiv* (2018). Available from: <https://doi.org/10.1101/302349>.

2646W

Deciphering sex-specific genetic architectures using Bayesian methods.

S.A. Funkhouser¹, G. de los Campos². 1) Genetics Program, Michigan State University, East Lansing, MI; 2) Department of Epidemiology and Biostatistics, Michigan State University, East Lansing, MI.

Many human diseases, ranging from obesity-related conditions to psychiatric disorders, have a different prevalence in men than in women. These differences may be due to differential environmental exposures or differential genetic architectures (genetic sex differences). The existence of genetic sex differences is supported by heritability estimates and by genome-wide estimates of effect correlations. However, mapping loci possessing sex-specific effects remains elusive, in part because any causal locus generally explains a very small ($R^2 < 1 \times 10^{-3}$) proportion of phenotypic variance. The methods commonly used in association analyses (single-marker regression, SMR) are severely underpowered to detect small effects, namely after multiple test correction. Bayesian multiple-marker regression techniques have been proposed for GWAS. These methods do not suffer the burden of multiple test correction and can provide better resolution in locating non-zero effects. In this study, we developed a Local Bayesian Multiple-Marker Regression (LBMMR) method, jointly fitting male and female SNP effects for overlapping chromosome segments spanning multiple LD blocks. This framework provides posterior probabilities of a SNP effect in either men or women as well as posterior probabilities of sex difference (PPDiff) for individual SNPs and chromosome segments. We use data from the UK-Biobank ($N = 258,928$ unrelated white Caucasians) to evaluate these methods using simulations and using real human phenotypes. Our simulation (in which sex differences in SNP effects spanned $R^2 = 7.9 \times 10^{-5}$ to $R^2 = 1 \times 10^{-3}$) shows LBMMR has substantially higher power (LBMMR: 0.74; SMR: 0.51) and a lower false discovery rate (LBMMR: 0.02; SMR: 0.48) than SMR for finding sex-specific SNP effects. In applying the LBMMR method to real traits—height, heel bone mineral density (BMD), BMI, and waist-to-hip ratio (WHR)—correlations of effects between men and women are estimated to be 0.941, 0.901, 0.825, and 0.624, respectively. For height, we confirm previous findings that the cross-sex correlation of effects is near one, yet newly discover four SNPs with sex-specific effects (PPDiff ≥ 0.9). Likewise, we have discovered 30 SNPs with sex-specific effects for WHR, largely near previously implicated genes including *VEGFA* and *ADAMTS9*, but also in previously unreported loci. In summary, LBMMR can provide new insight into sex differences by enhancing the precision and power to detect loci with sex-specific effects.

2647T

Meta-analysis of global-to-local genome-mapping on modular facial shape reveals 75 associated loci.

K. Indencleef^{1,2}, J.D. White³, J. Li^{1,2}, A. Ortega-Castrillón^{1,2}, J. Mohammed⁴, T. Swigut⁴, M.K. Lee⁵, T. González-Zarzar⁶, A.A. Zaid¹³, J. Roosenboom⁵, J.R. Shaffer⁷, E. Feingold^{7,8}, S. Richmond⁹, R.J. Eller¹⁰, S. Walsh¹⁰, M.L. Marazita^{8,9}, J. Wysocka^{4,11}, G. Hens¹², S.M. Weinberg^{5,7,13}, M.D. Shriver³, P. Claes^{1,2}. 1) Department of Electrical Engineering, ESAT/PSI, KU Leuven, Leuven, Vlaams-Brabant, Belgium; 2) Medical Imaging Research Center, MIRC, UZ Leuven, Leuven, Belgium; 3) Department of Anthropology, The Pennsylvania State University, University Park, PA, USA; 4) Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, CA, USA; 5) Center for Craniofacial and Dental Genetics, Department of Oral Biology, University of Pittsburgh, Pittsburgh, PA, USA; 6) Department of Biology, The Pennsylvania State University, University Park, PA, USA; 7) Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA, USA; 8) Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA, USA; 9) Applied Clinical Research & Public Health, School of Dentistry, Cardiff University, College of Biomedical & Life Sciences, Cardiff University Heath Park, Cardiff CF14 4XY, UK; 10) Department of Biology, Indiana University-Purdue University Indianapolis, Indianapolis, IN, USA; 11) Department of Developmental Biology, Stanford University School of Medicine, Stanford, CA, USA; 12) Department of Otorhinolaryngology, University Hospitals Leuven, Belgium; 13) Department of Anthropology, University of Pittsburgh, Pittsburgh, PA, USA.

The similarity of twins and relatives indicates that facial shape is under genetic control. Nonetheless, the number of replicated genetic variants showing associations with facial variation is small. In prior work, we used a novel data-driven approach to identify 15 loci involved in normal-range facial variation. Here, we report a substantial update and extension of that work. We meta-analyzed three large cohorts of European-ancestry with 3D facial shape data and genome-wide SNP data: PSU ($N=1,743$), PITT ($N=2,323$), and AL-SPAC ($N=3,566$). To derive comparable sets of phenotypic data across studies, we established a dense correspondence among the 3D facial surfaces (~10,000 quasi-landmarks) and applied a hierarchical modular decomposition approach, previously reported by our group, on the entire dataset to construct phenotypes from the 3D facial surfaces. This data-driven process results in 63 distinct “global-to-local” facial modules identified based on pairwise spatially-dense landmark correlations. SNPs were tested for association with modular shape PCs via multivariate canonical correlation analysis (calculated separately in each dataset), controlling for sex, age, height, weight, and population stratification. To accommodate the unprescribed selection of facial traits, we utilized a round-robin approach, in which each of the three datasets is, in turn, used to discover and define the linear combination of phenotypic variables that is maximally correlated with the SNP. Subsequently, the two other datasets are projected onto this learned phenotype, creating univariate phenotypic variables which are then also tested for genotype-phenotype associations, unrestricted by the SNPs reaching significance in the discovery dataset. The p-values across a single discovery with follow-up replications were then combined using Fisher’s method, and only results with all three meta p-values (one for each dataset as discovery) reaching genome-wide significance ($p < 5 \times 10^{-8}$) were retained. With this conservative approach, we identified 75 genome-wide significant loci associated with at least one facial module. By combining increased sample sizes and a novel approach for meta-analyzing multivariate 3D facial modules, we were able to quadruple the number of replicating loci involved in normal-range facial variation. This study provides new insights into the genetic basis of human facial shape and is both anthropologically and clinically relevant.

2648F

The landscape of pervasive horizontal pleiotropy in human genetic variation is driven by extreme polygenicity of human traits and diseases. *D.M. Jordan, M. Verbanck, R. Do.* Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY.

Understanding the nature and extent of horizontal pleiotropy, where one genetic variant has independent effects on multiple observable traits, is vitally important for our understanding of the genetic architecture of human phenotypes and for the field of human genetics. Undetected pleiotropy can bias genome-wide association studies (GWASs) and Mendelian randomization (MR) studies, and can cause unexpected effects when genetic evidence is used to select drug targets or in clinical applications of gene editing technology. Many recent studies have pointed to the existence of horizontal pleiotropy among human phenotypes, but the exact extent remains unknown, largely due to difficulty in disentangling the inherently correlated nature of observable traits. We have developed a statistical framework to isolate and quantify horizontal pleiotropy in human genetic variation, using a two-component pleiotropy score computed from summary statistic data derived from published GWASs. This score uses a statistical whitening procedure to remove correlations between observable traits and normalize effect sizes across all traits. When applied to simulated association summary statistics, our score successfully detects horizontal pleiotropy under a range of different models. When applied to real human phenotype data, using association statistics for 1,564 traits measured in 337,119 individuals from the UK Biobank, our score detects a significant excess of horizontal pleiotropy (genomic inflation factor $\lambda_{GC} = 4.94 - 9.95$). We find that horizontal pleiotropy is pervasive throughout the human genome and across a wide range of phenotypes and biological functions, but is especially prominent in regions of high linkage disequilibrium and among phenotypes known to be highly polygenic and heterogeneous. We additionally identify 6,645 loci with extreme levels of horizontal pleiotropy, 4,234 of which have never been previously reported in any published GWAS. These loci represent an under-recognized class of genetic variation with weak effects on many distinct phenotypes, but no specific marked effect on any one phenotype. 42% of these loci replicate using independent datasets of GWAS summary statistics, where a permutation analysis shows only 15% expected to replicate by chance. These results highlight the central role horizontal pleiotropy plays in the genetic architecture of human phenotypes, and the importance of modeling horizontal pleiotropy in genomic medicine.

2649W

Detailed annotation of genetic signals associated with human nose morphology. *J. Roosenboom¹, M.K. Lee¹, E. Feingold², M.L. Marazita¹, M.D. Shriver³, J. Wysocka⁴, P. Claes⁵, J.R. Shaffer², S.M. Weinberg¹.* 1) University of Pittsburgh, Center for Craniofacial and Dental Genetics, Pittsburgh, PA, USA; 2) University of Pittsburgh, Department of Human Genetics, Pittsburgh, PA, USA; 3) Department of Anthropology, Penn State University, University Park, PA, USA; 4) Department of Developmental Biology, Stanford University School of Medicine, Stanford, CA, USA; 5) KU Leuven, Medical Imaging Research Center, ESAT-PSI, Leuven, Belgium.

The nose is one of the most heritable regions in the face and its development is under strong genetic control. This is in line with the results of a recently published GWAS on facial morphology, where the majority of significant signals were associated with nasal shape. In this study, we took a deeper look into these previously described GWAS signals using multiple bio-informatics tools, in an attempt to gain insights into the biological role nominated genes might be playing during facial development. All significant and suggestive SNPs from GWAS ($p < 5 \times 10^{-6}$) associated with nasal morphology were selected and 208 genes surrounding these variants (± 500 kb) were identified. First, a Genomic Regions Enrichment of Associations (GREAT) analysis was performed, showing strong significant enrichment for biological processes involved in cartilage and bone formation, and neural crest cell differentiation. Furthermore, these signals were significantly associated with abnormal nose morphology in mice and humans. Second, using recently published large-scale transcriptome data, we investigated the tissue specificity of the implicated genes during craniofacial morphogenesis. Indeed, 16 of the genes surrounding the GWAS hits were strongly enriched in one of the facial prominences, with 15 of these genes showing differential expression in either the frontonasal (50%) or the maxillofacial (44%) prominence. Furthermore, the majority of these genes (69%) were strongly enriched in the ectodermal tissue during development. Thirdly, the SysFACE tool was used to identify six genes from the list of candidate genes that were strongly enriched in the frontonasal tissue. Finally, a STRING analysis was done to investigate the protein-protein interactions among these six candidates, in an attempt to unravel their mechanism of action. Based on this analysis, four genes (*SOX9*, *RUNX2*, *SMAD7* and *TWIST1*) showed strong experimental evidence of protein interactions and co-expression. These results provide strong biological validation of signals observed from GWAS of facial morphology and help us to focus our efforts in follow-up functional testing.

2650T

Locally dependent screening: Strategies for developing an accurate genomic predictor using big data. *M. Sun, G. de los Campos.* Michigan State University, East Lansing.

Modern genomic data sets include hundreds of thousands (even millions) of SNPs. However, even for highly complex traits the number of SNPs that are associated to disease risk and can contribute to prediction accuracy is considerably smaller than the number of genotyped SNPs. Therefore, SNP screening/selection is a critical step when building DNA-risk assessments. The standard approach for building risk assessments selects SNPs using summary statistics from single marker regressions (SMR). This approach has low power to detect small-effect variants and, by ignoring linkage disequilibrium, leads to redundant SNPs-sets and high rates of false discovery. To confront this problem, we propose a screening procedure based on Local-Regularized-Regression (LRR) methodology. The approach exploits the 'local' nature of linkage disequilibrium and applies sparsity-inducing penalized regressions (e.g., Lasso) within overlapping chromosome segments of the genome, each covering several LD-blocks (e.g., all the SNPs in a 10Mb region). The scanning procedure can be parallelized (one task per chromosome segment) allowing scanning the whole genome in reasonable computing time, even for very large sample size. The LRR is fitted over values of the sparsity-inducing parameter; this regularization path renders subsets of "active SNPs" and provides a natural approach for variable selection. Using simulations (based on real genotypes from ~270K ($K=1000$) unrelated white Caucasian individuals of the UK-Biobank) we show that the proposed methodology has substantially higher power and much lower false discovery rate than SMR (the method predominantly used in GWAS). We applied the proposed methodology to data from the UK-Biobank to identify relevant SNP sets for several complex human traits, including height, body mass index, body composition, bone mineral density and blood-biomarkers. The SNPs identified were then used to build a predictive model. The prediction accuracy (R^2) achieved by SNPs selected using LRR was compared with that of SNP-sets selected using p-values from SMR. Our empirical results show that the proposed screening method always yields better predictors than selection based on summary statistics from SMR for sets including up to 10-20K SNPs. This result aligns well with the simulation results. The improvement in prediction R^2 achieved with the proposed screening procedure reached up to 8 percentage-point in the R^2 -scaled scale (a 76% relative gain).

2651F

Inbreeding effects on damaging homozygote loads in gene-sets associated with 543 complex phenotypes. *A. Blant¹, Z.A. Szpiech², T.J. Pemberton¹.* 1) University of Manitoba, Winnipeg, Manitoba, Canada; 2) University of California San Francisco.

Genomic regions of autozygosity (ROA) reflect homozygosity for haplotypes inherited identical-by-descent from an ancestor common to both parents. They vary widely in length, density, and genomic distribution at both the individual- and population-level, where longer ROA arise from recent parental relatedness as a result of inbreeding or limited mate choice, while shorter ROA arise from baseline linkage disequilibrium patterns. Over the past two decades, the study of ROA patterns has provided key insights into the demographic and sociocultural processes that have shaped genomic variation patterns in the human genome. We have previously shown that longer ROA are enriched for damaging allele homozygotes, and strongly damaging allele homozygotes in particular, while the degree of enrichment in individual genomes varies appreciably across worldwide populations. Here, we explore the rates-of-gain in damaging allele homozygotes with increasing ROA coverage in gene-sets associated with 543 complex diseases and traits (relative to genes not associated with Mendelian and complex diseases and traits) in whole genome sequences available for 2,436 individuals included in Phase 3 of The 1000 Genomes Project. We find that the rate-of-gain in damaging allele homozygotes with increasing ROA coverage differs significantly from the rate-of-gain in non-damaging homozygotes in gene-sets associated with complex diseases and traits; and the magnitude of this difference increases as a function of ROA length. Notably, over half of all gene-sets acquire damaging allele homozygotes in longer ROA at a rate that significantly outpaces that of non-damaging homozygotes. Comparisons of the rate-of-gain in damaging allele homozygotes between gene-sets associated with complex diseases/traits versus unassociated genes highlights the gene-sets that acquire damaging allele homozygotes at either significantly higher or lower rates than the rest; which emphasizes that such patterns can vary both as a function of ROA length as well as across worldwide populations. These findings highlight the influence of cultural and demographic processes on the shaping of genetic risk for complex diseases as well as how variation in complex traits can arise through pattern modification of functionally important genetic variation in a population-specific manner.

2652W

Improving SNP-based gene expression prediction. *D. Zhou, E. Gamazon, N. Cox.* Vanderbilt Genetic Institute, Nashville, TN.

PrediXcan, a powerful gene-based complex trait mapping approach, leverages genetically determined gene expression level and associates it with phenotypes. Gene expression prediction is a fundamental step of PrediXcan. Currently, the selection criterion of input SNPs for expression prediction is fixed, i.e. 1 Mb both sides of the gene body or TSS (only considered local SNPs). While we supposed that gene specific local (or cis) window size and other parameters will bring some improvement for expression prediction. In the current study, we tried to improve the gene expression prediction using different parameters to build the model. The local window size (100kb, 500kb, 1Mb, and 5Mb), with or without distal (or trans) variant, and the alpha value of the model (0.5 for elastic net and 1 for LASSO) are considered for model training. In order to test the performance, we cut the GTEx whole blood sample data into three parts (223 samples for each part). Firstly, we built 20 models with different combinations of parameters using the first part of the data. Then, the second part of the data was used to find the best model according to the highest r^2 . Meanwhile, we constructed the regular prediction model using the first two parts of the data together. At last, we compared the performance of the prediction model with best combinations to the regular way based on the third data set. Results show that, in global, the performance is no difference between the gene specific and uniform method. It indicated that using the additional data did bring some information like the suitable window size, but the improvement would be pulled back by the sample size. We noticed that most of the genes showed improvement using specific parameters are benefit from smaller local window size. So we used a modified SNP selection criterion (100kb as local size and the rest of the genome as distal) without cutting down the sample size to find out if it is better than the regular way. The result showed that the r^2 is significantly higher using smaller window size when we have less than 200 samples for model training. When the sample size is greater than 300, the performance is almost the same. This study indicated that using some part of the samples to find the suitable input SNP list will bring benefits for model training, but it will also pay the price of loss of power. Smaller local window size will reduce the noise when the sample size is limited in the training set.

2653T

Discovering genetically correlated variants in genome-wide association studies with multiple traits. *K. Collins¹, E. Eskin^{1,2}.* 1) Computer Science, University of California, Los Angeles, Los Angeles, CA; 2) Human Genetics, University of California, Los Angeles, Los Angeles, CA.

Genetic correlations estimated from Genome-Wide Association Studies (GWAS) have been reported for dozens of pairs of traits over the past few years demonstrating a shared genetic basis for traits. These results show that many traits share both the same causal variants and that the variant effect sizes are correlated. However, the vast majority of methods for discovering genetic associations examine only one trait at a time, and virtually all reported genetic association studies utilize this single trait approach. It is natural to assume that by combining more than one trait in the association analysis, there would be more power to identify genetically correlated variants. In this project, we assume that the effect sizes of traits follow the Fisher polygenic model and are correlated according to genetic correlation estimates obtained from GWAS. We introduce a method that uses a likelihood ratio for summary statistics, an approach shown to be the most statistically powerful hypothesis test for any level of significance. Under the null hypothesis, summary statistics are derived from a set of variants with no genetic effect; therefore, any deviation in the summary statistics from zero is fully attributed to environmental effects. In the alternative model, however, there is at least one non-zero genetic effect and the distribution is according to the Fisher polygenic model. Using the distribution of genetic effects as a conjugate prior results in the alternative model being normally distributed such that the variance is a sum of the environmental effect and genetic effect. This approach results in substantial power increase for simulations where we assume the effect sizes follow the model. We also show that our method has a more moderate power increase in many scenarios where the simulations deviate from the model. We apply our model on real data from the UK Biobank and show that utilizing multiple traits increases power for detecting variants that affect more than one trait. We also examine how real data differs from the Fisher polygenic model which can give insights into the shared genetic basis of traits and into future methods for discovering variants affecting multiple traits.

2654F

The search for quantitative iris pigmentation markers using automation in genome-wide association studies: From phenotype to genotype. *R. Eller¹, G. Zhu², N. Herrick¹, A. Wollstein⁴, M. Kayser³, N. Martin², S. Walsh¹.*

1) Biology, Indiana University Purdue University – Indianapolis, Indianapolis, IN; 2) Epidemiology, QIMR Berghofer Medical Research Institute, Herston, Queensland; 3) Department of Genetic Identification, Erasmus MC University Medical Center, Rotterdam, The Netherlands; 4) Woodmark Consulting AG, Grasbrunn, Bayern, Deutschland.

Iris pigmentation has often been classified in a categorical fashion despite being a complex trait that takes on a range of appearances. Grouping iris color into light, intermediate, and dark colors has proven useful in early Genome Wide Association Studies (GWAS) to elucidate some of the major genes that affect iris pigment production. However, here we apply a phenotyping strategy that has been previously published by the group (Wollstein et al. 2017). This approach tries to explain more of the intricate details of iris content such as the percentage of pre-defined color classes to identify variants that have a smaller effect size, and those associated with the yellow pigment type pheomelanin. In addition to moving toward a more quantitative color phenotyping method, we looked at iris morphological traits including heterochromia with differing colors and the presence of a concentric distribution with a similar pigment. After performing preliminary GWAS using the more quantitative range of phenotypes, we found hits in several known pigmentation genes and encouraging signals for several novel loci associated with our quantitative color classes. Additional GWAS including replication will need to be performed, though preliminary findings suggest moving to a more quantitative approach will lead to the elucidation of lower penetrance variants that may have been missed in previous searches. While performing GWAS is a standard in the field, its power of detection is often limited by sample size and delayed by data preparation. Finding additional subjects to increase power will remain a logistical challenge, however cleaning genetic and phenotypic data can be automated. Therefore, we produced Odyssey, an autonomous genetic preparation pipeline that can quickly organize genetic data from arrays to clean, phase, impute, perform quality control, convert into formats accepted by genetic analysis tools such as PLINK, and assist in the setup of GWAS analyses. Thus, researchers can minimize the time it takes to prepare data and/or eliminate the outsourcing of data preparation. Using Odyssey to prepare our genetic data in tandem with our in-house eye quantifier program, which automatically quantifies each image pixel into select quantitative color classes, we substantially decreased the time it took to go from data to analyses without issue. This can be seen in the preliminary findings where published signals for iris color association were observed for many loci.

2655W

GWAS of 92 neurology-focused plasma proteins reveals over one hundred novel loci. *B.P. Prins¹, P.G. Bronson², J.E. Peters¹, D. Sangurdekar³, J. Danesh¹, K. Estrada², S. John³, A.S. Butterworth¹.*

1) MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, 2 Worts' Causeway, Cambridge CB1 8RN, UK; 2) Statistical Genetics and Genetic Epidemiology, Biogen, Cambridge, MA, 02142, USA; 3) Translational Biology, Biogen, Cambridge, MA, 02142, USA.

Many neurological and psychiatric diseases are substantially heritable and are often highly polygenic. For many of these disorders GWAS analyses have highlighted regions that are likely to harbour genes that are causally involved in disease pathogenesis. Identification of the causal genes and their protein products (the intermediate effectors), might permit development of therapeutics targeting these pathways. Furthermore, genetic variants that specifically influence protein levels can be utilised as instrumental variables to infer causality using Mendelian randomisation analyses. However, to date only a handful of powerful proteome association studies have been performed. Here we performed a proteome association study on a set of 92 neurologically relevant proteins using the Olink Proseek Multiplex Neurology I proteomics panel in plasma from 4,662 healthy individuals from the INTERVAL cohort. We found 67 proteins associated with nearby proteins (cis associations) ($P < 5 \times 10^{-4}$) for which most sentinel SNPs were in genes (85%) and also had trans associations ($N=54$). Furthermore, 12 pQTLs were cis-only, and 15 were trans-only. 69/178 of significantly associated cis-variants were previously identified using the SOMAscan platform (Sun et al, 2018); direction of effect was correlated between the two studies ($r^2=0.89$, $P=3.1 \times 10^{-16}$). Several loci overlap with those associated with relevant disorders, such as bipolar disorder, schizophrenia, and Alzheimer's disease, helping to pinpoint the causal pathways at these loci. Additionally, several variants highly correlated with our sentinel SNPs appear to affect methylation levels in relevant tissues (frontal- and temporal cortex). Finally, Mendelian randomisation analyses suggest several potential novel candidate therapeutic targets. Taken together, these results provide insights into the identification and interpretation of targetable causal intermediates for a wide range of relevant diseases.

2656T

***Alu* variants can affect gene expression or splicing and contribute to disease risk.** L.M. Payer¹, J.P. Steranka¹, D. Ardeljan¹, T.A. Cooper², K.H. Burns¹.

1) Johns Hopkins University, Baltimore, MD; 2) Baylor College of Medicine, Houston, TX.

Alu elements are the most common mobile DNA in the human genome and the most polymorphic. *de novo* insertions of *Alu* sequence are known to underlie some Mendelian disorders. A handful of polymorphic *Alu* have been suggested to elicit functional effects and contribute to common disease risk. However, the potential functional repertoire of *Alu* polymorphisms has not been fully evaluated. Recent studies have yielded catalogs of common *Alu* variants; we wanted to systematically identify functional *Alu* variants and test their potential contribution to disease risk. We first identified 809 previously published *Alu* variants in linkage disequilibrium (LD) blocks ($r^2 > 0.8$) with GWAS (Genome-Wide Association Studies) signals, a significant enrichment of *Alu* variants at GWAS loci ($p = 0.013$, permutation test). We propose that a subset of these variants modulate disease risk. We prioritized 44 *Alu* variants in strong LD ($r^2 > 0.8$) with a GWAS trait-associated SNP and sought to determine the mechanism by which they may alter gene regulation. We hypothesized that polymorphic *Alu* elements may alter transcript level or structure. To evaluate the first model, we used standard luciferase assays to test 44 *Alu* polymorphisms for their potential to alter transcript levels, including ~300 bp of their genomic context. We observe a continuum of effects. Most notably, 4 *Alu* displaying significant effects were associated with breast cancer risk. To validate, we modified (CRISPR) the endogenous sequences in HEK293T cells establishing isogenic lines varying only in the *Alu* genotype. Consistent with our luciferase assays, these data reveal modulation of transcript levels (RT-qPCR) of adjacent genes, including *RANBP9* and *MYC*. To address the second model, evaluating the disruptive potential of intronic *Alu* variants, we used a minigene reporter assay to measure exon incorporation. We tested 5 polymorphic *Alu* elements that map near exon-intron boundaries (≤ 100 bp), of which 3 alter exon incorporation rates (Genes *SLC2A9*, *NUP160*, *CD58*). The latter is of particular interest due to its hypothesized role in multiple sclerosis (MS) risk. This *Alu* element promotes skipping of *CD58* exon 3 leading to a frameshift and predicted nonsense mediated decay, consistent with the reduced *CD58* expression associated with MS risk. Collectively, these data demonstrate that, as a class, polymorphic *Alu* elements commonly alter transcript levels and content, highlighting their potential contribution to disease risk.

2657F

The NHGRI genome sequencing program. T. Matise, NHGRI Genome Sequencing Program. Genetics, Rutgers University, Piscataway, NJ.

The NHGRI Genome Sequencing Program (GSP; <http://gsp-hg.org>) uses genome sequencing to identify genes and genomic variants underlying the full spectrum of human inherited disease from Mendelian diseases to common, genetically complex diseases. The GSP comprises four components: The Centers for Mendelian Genomics (CMG; Baylor-Hopkins, Broad, U Washington, Yale) use whole exome sequencing and other genomic approaches to discover the genetic basis underlying as many Mendelian traits as possible and accelerate discoveries by disseminating the new knowledge and effective approaches, by reaching out to individual investigators, and by coordinating with other rare disease programs. The CMG have studied over 22.7K families and have discovered over 1.2K novel unique gene-phenotype associations and 400 cases with phenotypic features that expand those previously reported for a Mendelian condition. The Centers for Common Disease Genomics (CCDG; Baylor, Broad, New York Genome Center, Washington U) collaborate to comprehensively identify rare risk and protective variants that contribute to multiple common disease phenotypes. The CCDG aim to improve understanding of the general principles of genomic architecture, to understand how best to design rare variant studies for common disease, and to develop resources, informatics tools, innovative approaches, technologies. They have so far produced over 56K WGS and 46K WES (using equivalent pipelines) from samples from over 90 common disease cohorts focused on immune-mediated disease (asthma, inflammatory bowel disease, type I diabetes), cardiovascular disease (early onset coronary disease, atrial fibrillation, hemorrhagic stroke), and neuropsychiatric disease (Alzheimer disease, autism, epilepsy). Where consents allow, these data are also being used for haplotype imputation, genome annotation, population genetics studies, and as common controls. Three GSP Analysis Centers are developing improved and novel analyses and methods to maximize the value of the CMG and CCDG studies, with a focus on genome annotation. The GSP Coordinating Center provides scientific leadership and expertise for specific cross-program objectives and coordination for administrative, logistical, and outreach activities. The GSP shares raw and processed data, analysis results, and novel methods and analysis pipelines. These products are available through dbGaP, ClinVar, GeneMatcher, MyGene2, Matchmaker Exchange, ClinVar, gnomAD, Bravo, and GitHub.

2658W

Large scale implementation of clinical preemptive pharmacogenomic testing: The Mayo-Baylor RIGHT10K project. S.E. Scherer¹, S.J. Bielinski², X. Qin¹, J.L. Black², E. Venner¹, C. Kovar¹, J. Hu¹, K. Walker¹, V. Korchina¹, A. Moyer², T. McAllister², L. Jones², C.R.R. Vitek², J. St. Sauver², J.E. Olson², M.E. Karow², A. Staj², M. Wiepert², W.T. Nicholson², E. Matey², J. Wright², J. Giri², P. Caraballo², R. Sharp², T.B. Curry², D. Muzny¹, E. Boerwinkle¹, L. Wang², R. Gibbs¹, R. Weinshilbaum². 1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 2) Center for Individualized Medicine, Mayo Clinic, Rochester, MN.

A major opportunity for the application of genome technologies for individualized therapies is informing drug prescribing practices by sequencing and analysis of variants found in genes encoding proteins involved in drug transport, metabolism and elimination. The Human Genome Sequencing Center at Baylor College of Medicine and Mayo Clinic's Center for Individualized Medicine have collaborated to institute clinical pharmacogenomic sequencing of 10,085 research participants as part of Mayo's RIGHT Study. The capture sequencing panel targets 77 genes and 2140 SNPs of pharmacogenomic relevance. Participant sequence generation was performed in a CAP/CLIA environment with results returned for an initial list of 13 genes and associated variants that correspond to clinically actionable drug interactions. To further assess the participant's relevant HLA genotypes, we incorporated Omixon's *HLA Explore* software. For *CYP2D6*, bam files were analyzed via local software to identify novel variants, including novel copy number/structural variations, for further analysis. Discrete clinical genotype and phenotype results were entered into the electronic health record to be utilized by electronic clinical decision support tools to inform future prescribing events. An annotated test report with genomic results and phenotype descriptions as well as clinical recommendations offered by pharmacists for results deemed clinically actionable were also added to the EHR for immediate and future use by providers. Additionally, test results and the test report were made available to participants by mail and through an electronic health portal. To support the use of results, an education and communication strategy was implemented specifically targeting 849 providers and 39 pharmacists as well as a broader awareness initiative for providers and allied health staff. In consultation with the Mayo-CIM community advisory board, the annotated report format was modified to increase patient comprehension and facilitate appropriate clinical response. Preliminary results within the 13 gene subset suggest an actionable variant rate of greater than 90%, a novel variant discovery rate of 13%, and a *CYP2D6* reflex rate of 2.3%. Further analyses are underway, including economic and overall changes to drug prescribing practices. This project represents the first large-scale comprehensive effort of this type and provides a model for broader implementation of precision medicine applied to drug response.

2659T

Improving drug discovery with PheWAS across >400,000 UK Biobank participants. E. Wong¹, D. Ficenece², M. Sabripour¹, D. Merberg¹, B. Li¹, Y. Ge¹, S. Szalma¹. 1) Data Science Institute, Takeda Pharmaceuticals, USA; 2) Information Technology, Takeda Pharmaceuticals, USA.

As high-quality phenotype data becomes available, there is increasing interest to conduct phenome-wide association study (PheWAS) using those data, which assesses the impact of one or more genetic variants across multiple phenotypes. PheWAS enables us to better understand the landscape of associations across human diseases and traits. This would help validate disease-target association reported previously, identify pleiotropy, and thus help drug development through elucidating mechanisms of action, suggesting alternative indication hypotheses and predicting adverse drug events. UK Biobank collects vast amounts of phenotype data through electronic health records, online questionnaires, imaging, blood biomarker measurements, actigraphy, etc. This provides valuable resources to characterize the complex relationship between human genome and phenome. We therefore conducted a PheWAS using genotype data and extensively curated phenotype data from >400,000 UK Biobank participants with Caucasian ancestry. Variants of interest were identified based on three major therapeutic areas – gastroenterology, neuroscience and oncology. The phenotype data was pre-processed and harmonized to streamline subsequent analyses. We performed the PheWAS with a Spark and Hail framework on Amazon Databricks. The framework can be extended to the full exome sequencing dataset from UK Biobank when the data becomes available.

2660F

The influence of population structure on complex trait architecture:

Insights from the PAGE study. G.L. Wojcik¹, M. Graff², S. Cullina³, G.M. Belbin³, H.M. Highland³, C.D. Bustamante¹, B. Hailu⁴, L.A. Hindorf⁵, S. Buyske⁶, C. Haiman⁷, C. Kooperberg⁸, L. Le Marchand⁹, R.J.F. Loos³, T.C. Matise⁶, K.E. North², U. Peters⁸, C.R. Gignoux¹⁰, C. Carlson⁸, E.E. Kenny⁸. 1) Stanford University School of Medicine, Stanford, CA; 2) University of North Carolina at Chapel Hill, Chapel Hill, NC; 3) Icahn School of Medicine at Mount Sinai, New York, NY; 4) NIH National Institute on Minority Health and Health Disparities, Bethesda, MD; 5) NIH National Human Genome Research Institute, Bethesda, MD; 6) Rutgers University, New Brunswick, NJ; 7) Keck School of Medicine, University of Southern California, Los Angeles, CA; 8) Fred Hutchinson Cancer Research Center, Seattle, WA; 9) University of Hawaii, Honolulu, HI; 10) University of Colorado, Anschutz Medical Campus, Aurora, CO.

Over the past several years, genome-wide association studies (GWAS) have come of age with sample sizes reaching the hundreds of thousands, providing unprecedented statistical power for discovery. However, the majority of this progress has taken place in European descent populations, exacerbating pre-existing disparities in genomic knowledge. The Population Architecture using Genomics and Epidemiology (PAGE) Study was developed to address the lack of non-European descent groups in large-scale genetic studies, comprising ~50,000 individuals self-identified as African American, Hispanic/Latino, Native American, Native Hawaiian, and Asian. While the focus of PAGE has been on discovery, it is important to consider its value within the context of the current state of the field. We empirically assessed the value of multi-ethnic studies by examining discovery, fine-mapping, and replication trends. To directly compare study findings of height to our multi-ethnic PAGE results, we subset a random sample of 50k 'White British' from UK Biobank (UKB50k) and meta-analyzed each stratified GWAS with GIANT, an existing large-scale study of anthropometric traits in European descent individuals. While UKB50k+GIANT yielded more novel significant loci, these signals explained a smaller proportion of variance in PAGE, reflecting differences in allele frequencies (AF) and exacerbating inequity. To assess the impact of linkage disequilibrium (LD), we compared the ability to finemap top GIANT loci. The mean 95% credible set size decreased with the addition of PAGE to GIANT ($P=0.01$), while there was no decrease with UKB50k+GIANT ($P>0.05$). Building on the unique multi-ethnic dataset in PAGE, we then interrogated the role of genetic ancestry to predict replication of known GWAS loci from the NCBI/EBI GWAS catalog across 26 traits. While there was no homogeneous relationship between African, European, or Native American ancestry and replication, we found several trait-specific trends ($P<0.05$), reflecting both heterogeneity in genetic architecture and biases in existing literature. To further explore the role of heterogeneous LD and AF, we characterize genetic correlation between groups within PAGE. Together, these trends demonstrate the importance of leveraging ancestral diversity for both discovery and refinement of GWAS signals, as well as directly addressing the disparity in genomic knowledge between populations and the unique opportunities provided by multi-ethnic studies..

2661W

Analysis of densely imputed UK Biobank genetic data reveals disease-associated rare loss of function variation.

S.A. Gagliano¹, W. Zhou¹, D. Taliun¹, J. Nielsen², J. LeFaivre¹, R. Dey¹, S. Das¹, G.R. Abecasis¹, The NHLBI Trans-Omics for Precision Medicine (TOPMed) Program. 1) Center for Statistical Genetics, University of Michigan School of Public Health, Ann Arbor, MI; 2) Division of Cardiovascular Medicine, Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, MI.

Loss of function (LoF) variants, such as those that introduce a premature stop codon or shift the reading frame for translation machinery, alter protein structure to the extent of eliminating or greatly diminishing protein action. This biological interpretability makes LoF variants a particularly important subset of genetic variation to study. To pinpoint disease-associated LoF, we imputed additional variants into the UK Biobank cohort of half a million genotyped individuals using 60,039 deeply whole genome sequenced individuals from the multi-ethnic TOPMed project. This allowed us to expand the number of genotyped or imputed variants characterized in the UK Biobank from 39,131,578 to 177,895,992. The vast majority (94%; 167,502,731) of the imputed variants are rare (alternate allele frequency $<0.5\%$), of which 0.03% (49,892) are high-confidence predicted LoF. This dramatic increase in the number of variants in a large well-characterized cohort provides an ideal setting to assess the impact of LoF variation. To identify disease-associated LoF variants, we conducted single-variant and gene-burden association tests for $>1,400$ binary traits constructed from health record billing codes. In the single-variant analyses, we identified five rare LoF variants with frequency $<0.5\%$ (not found in the 39M dataset) to be associated with disease. This included two variants associated with breast cancer: a frameshift indel in *CHEK2* (chr22:28695868:AG:A; build38; $p=7.0E-22$) and a SNP resulting in a gained stop in *PALB2* (chr16:23621362:C:T; $p=6.9E-14$). Both are present in the ClinVar database as potentially pathogenic for familial breast cancer, but this is the first time these variants have been identified by GWAS, and characterized in a population of generally healthy adults. Beyond single variant analyses, we found significant burden signals in genes previously implicated in familial disease, but for which no LoF variants were significantly associated in the single-variant analyses. This included an association between *USH2A* LoF and hereditary retinal dystrophies, *IFT140* LoF and kidney cysts, and *BRCA2* LoF and breast cancer. We demonstrate that association studies in large-scale biobanks, even with a relatively small number of cases, are capable of yielding pathogenic findings that previously were only detected in clinical cases or difficult-to-collect family cohorts, in which it can be challenging to obtain robust associations.

2662T

Genomic integrity of human induced pluripotent stem cells across nine studies in the NHLBI NextGen Program. K. Kanchan¹, K. Baldwin², L.C. Becker³, U. Broeckel⁴, L. Cheng⁵, C. Cowan⁶, K. Frazer⁷, C.E. Jaquish⁸, J.W. Knowles⁹, G. Mostoslavsky¹⁰, G. Murphy¹¹, T. Quertermous¹², M. Rabinovitch¹³, D.J. Rader¹⁴, M.H. Steinberg¹⁵, E. Topol¹⁶, W. Yang¹⁷, I. Ruczinski¹⁸, R.A. Mathias¹. 1) Department of Medicine, School of Medicine, Johns Hopkins University, Baltimore, MD; 2) Department of Molecular and Cellular Neuroscience, Dorris Neuroscience Center, The Scripps Research Institute, La Jolla, CA, USA; 3) Department of Pediatrics, Medical College of Wisconsin, Milwaukee, WI, USA; 4) Division of Cardiovascular Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA; 5) Department of Pediatrics and Rady Children's Hospital, University of California, San Diego, CA, USA; 6) National Heart, Lung, and Blood Institute, NIH, Bethesda, MD, USA; 7) Stanford University School of Medicine, Stanford Center for Inherited Cardiovascular Disease, Stanford University, Stanford, CA, USA; 8) The Center for Regenerative Medicine, Boston Medical Center, Boston University School of Medicine, Boston, MA, USA; 9) Department of Medicine, Division of Cardiovascular Medicine, Stanford University, Stanford, CA, USA; 10) Department of Pediatrics, Stanford University, Stanford, CA, USA; 11) Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; 12) Department of Medicine, Section of Hematology-Oncology, Boston University School of Medicine, Boston, MA, USA; 13) Department of Molecular Medicine, The Scripps Research Institute, La Jolla, CA, USA; 14) Penn Center for Pulmonary Biology and Institute for Regenerative Medicine, University of Pennsylvania, Philadelphia, PA, USA; 15) Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA.

Prior studies suggest that human induced pluripotent stem cell (hiPSC) lines suffer from genomic instabilities (e.g. karyotypic abnormalities, chromosomal aberrations). Within the NHLBI sponsored NextGen program hiPSC lines were generated by 9 consortia to study the complex genetics of metabolic, cardiovascular and respiratory diseases. The objective of this work is to assess the genomic integrity of the NextGen hiPSCs. Genomic integrity of 1,226 hiPSC lines generated from a range of sources (PBMCs, fibroblasts and lung epithelium) was compared to their matched donor DNA from blood/PBMCs leveraging 1.3 million genetic variants from the Illumina Multiethnic Genotyping (MEGA) array. We investigated two levels of integrity: genotype discordance between parent donor DNA and hiPSC lines and structural variability using copy number variants (CNVs). CNVs were called using the automated *VanillaICE* package, and then subsequently manually validated by visual examination of all plotted data. We observed low rates of genotype discordance with average rates of 0.01% between the donor and hiPSC lines. There was a mean of 8 deletions and 4 amplifications per hiPSC line. Manual inspection of these CNVs was performed to identify which of these were acquired in the hiPSC (observed in the hiPSC but not the donor DNA). We found 151 hiPSC lines acquired 266 CNVs relative to the parent donor DNA. Out of 151 hiPSC lines, 91 hiPSC lines had 1 CNV, 33 lines had 2 CNVs and 27 had ≥ 3 CNVs. These CNVs were generally small (83% were less than 1Mb in size) and included 152 deletions and 114 amplifications. The cumulative impact per hiPSC line was small; 85% showing less than 2 Mb of cumulative CNV coverage. We observed a previously reported amplification on Chr20q in 15 hiPSC lines and the overlap (chr20:29,810,887-30,412,182) across these lines is limited to 15 genes including known cancer loci *ID1*, *BCL2L1* and *TPX2*. In conclusion, a low level of genomic instability was observed in the NextGen hiPSC lines compared to prior reports. The highest number of CNVs were observed in lines generated from fibroblasts, but it should be noted that these lines were compared to genomic DNA from blood/PBMC sources and not donor fibroblasts. Our results substantiate the importance of systematic evaluation of genetic variations in hiPSCs prior to their use as disease models and confirm the preferential amplification previously reported on Chr20q within hiPSCs.

2663F

Overexpression of *RUNX2* in mesenchymal stem cells derived from fused cranial sutures of individuals with metopic non-syndromic craniosynostosis. K. Bala¹, M. Barba², L. Di Pietro², W. Lattanzi², A. Cuellar¹, S. Boyadjiev¹. 1) Section of Genetics, Department of Pediatrics, University of California Davis, Sacramento, CA; 2) Institute of Anatomy and Cell Biology, Università Cattolica S. Cuore, Rome, Italy.

Craniosynostosis is a genetic birth defect occurring in 3–5 per 10,000 live births; causing the premature fusion of one or more cranial sutures. It presents as a feature of over 200 syndromes, but approximately 80% of cases are non-syndromic (NCS). The development and maintenance of cranial sutures is dependent on the activity of osteoblasts and osteocytes, which are controlled by several regulatory factors including those involved in the bone morphogenic protein (BMP)/transforming growth factor beta (TGF β), fibroblast growth factor (FGF), hedgehog (HH), eph-ephrin, and Wnt signaling pathways. These pathways all converge on the same downstream effector Runt-related transcription factor 2 (*RUNX2*), a major regulator of osteogenesis. In 2013, Lattanzi et al demonstrated overexpression of *RUNX2* in cells derived from fused sutures when compared to those from open sutures, but focused on the role of LMP (a target of *RUNX2*). The fused suture samples were derived from sagittal, coronal, and metopic sutures. Metopic NCS (mNCS), occurs between 1/7,000 and 1/15,000 births. To focus on *RUNX2* expression in mNCS patients, calvarial tissue fragments were collected from sixteen patients undergoing cranial surgery. Mesenchymal stem cells were isolated from both open and fused sutures of the same individual and RNA was isolated from these cells. qPCR was performed on these 16 pairs of samples to measure *RUNX2* mRNA expression. To evaluate statistical significance, a paired t-test on Δ CT values with a p-value cut-off of 0.05 was used. When comparing within each pair, *RUNX2* was significantly upregulated in 62.5% of the cells derived from the fused suture. When comparing all cells derived from the fused sutures to those from the open sutures, *RUNX2* was significantly upregulated in cells derived from the fused suture. Also, differential expression patterns for the different splice variants of the gene were found. *RUNX2* overexpression may cause accelerated differentiation of the mesenchymal stem cells present in the fused sutures, leading to the premature suture ossification.

2664W

USF1 modulates chronotype and consolidates sleep. *H.M. Ollila*^{1,2,3}, *N. Sinnot-Armstrong*⁴, *K. Zitting*⁵, *T. Paunio*³, *R. Saxena*¹, *T. Porkka-Heiskanen*⁶, *J.K. Pritchard*⁴. 1) Department of Medical Genetics, Massachusetts General Hospital, Boston USA; 2) Department of Psychiatry and Behavioral Sciences, Stanford University, Palo Alto CA, USA; 3) Institute for Molecular Medicine Finland, FIMM, University of Helsinki, Helsinki, Finland; 4) Department of Genetics, Stanford University, Stanford, CA 94305, USA; 5) Division of Sleep Medicine, Harvard Medical School, 221 Longwood Avenue, BLI438, Boston, MA, USA; 6) Department of Physiology, Faculty of Medicine, University of Helsinki Helsinki, Finland.

Timing and amount of sleep are controlled by a set of master regulators in the brain. *USF1* has been shown to modify sleep traits in mammals. Here we aimed to understand how *USF1* affects sleep. We show that *Usf1* knock-out mice have lower activity, frequent napping, extreme sleepiness and severe fragmentation of sleep. The sleep defects were compensated by increase in total sleep amount, which was specific for NREM sleep. To understand the effect on human sleep we then examined whether *USF1* transcription factors affect sleep traits in a sample of 453,000 European individuals from the UK Biobank. While modest association was seen at *USF1* locus itself ($p=3.7E-7$ with insomnia), we saw a strong enrichment of *USF1* and *USF2* binding sites, which explained up to 20% of SNP-based heritability for morningness-eveningness, insomnia, excessive daytime sleepiness and napping ($p\text{-value} < 0.0001$). These findings replicated the phenotypes seen with *Usf1* knock-out mice. Furthermore, 28 sites at *USF1* binding sites were both GWAS-significant for sleep traits and showed a statistically significant allele specific DNase directly on *USF1* binding site ($p<0.05$) including an allele specific expression at *USF1* binding site near a key circadian pace maker *PER1* ($p<0.05$). These findings suggest that *USF1* transcription factor is a crucial master regulator in shaping sleep. These effects of *USF1* are likely conveyed through small cumulative effects on a number of circadian downstream targets.

3250T

Genetic associations between obesity and asthma: A genome-wide cross trait analysis of 450,000 individuals from UK Biobank. Z. Zhu^{1,2}, Y. Guo¹, R. Pangniban³, W. Chung⁴, L. O'Connor¹, B. Himes⁴, S. Gazal¹, K. Hasegawa⁵, C. Camargo Jr.⁶, L. Qi⁶, M. Moffatt⁷, Q. Lu⁸, W. Cookson⁷, L. Liang^{1,8}. 1) Epidemiology, Harvard University T H Chan School of Public Health, Boston, MA; 2) Environmental Health, Harvard University T H Chan School of Public Health, Boston, MA; 3) Program in Molecular and Integrative Physiological Sciences, Departments of Environmental Health and Genetics & Complex Diseases, Harvard T.H. Chan School of Public Health, Boston, MA; 4) Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, Philadelphia, PA; 5) Department of Emergency Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA; 6) Department of Epidemiology, School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA; 7) Section of Genomic Medicine, National Heart and Lung Institute, London, UK; 8) Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA.

Clinical and epidemiological studies showed that obesity and other metabolic disorders are associated with asthma. Recent observational studies suggest that these associations may differ by asthma phenotype. In the current study, we investigated the shared genetic components between metabolic traits and asthma, and specifically examined whether these shared genetic components differ among childhood-onset, later-onset, atopic and non-atopic asthma. We conducted a large-scale GWAS cross-trait analysis between asthma (childhood-onset, later-onset, atopic, and non-atopic asthma) and body mass index (BMI) in UK Biobank (N=459,306) and publically available GWAS data for 8 obesity related metabolic traits. We found a strong positive genetic correlation of obesity (evaluated by body mass index [BMI] and waist to hip ratio [WHR]) with later-onset asthma ($r_g = 0.25$, $P = 4.21 \times 10^{-22}$ for BMI; and $r_g = 0.23$, $P = 3.29 \times 10^{-10}$ for WHR), and non-atopic asthma ($r_g = 0.23$, $P = 1.32 \times 10^{-10}$ for BMI; and $r_g = 0.17$, $P = 5.03 \times 10^{-5}$ for WHR). Cross phenotype meta-analysis identified 11 independent loci that reached genome-wide level of significances ($P_{meta} < 5 \times 10^{-8}$) for BMI and later-onset asthma, 9 loci for BMI and non-atopic asthma. Colocalization analysis found that most of these loci were colocalized between BMI and asthma either at a single causal variant or with different causal variants. Additionally, we found causal relationship between the association of BMI with later-onset asthma and eQTLs of 10 relevant GTEx tissues. Pathway analysis highlighted immune and antigen related pathways shared by BMI and asthma. Finally, we conducted RNA-seq gene expression experiment of lung tissues from two sets of obese mouse and identified 8 out of the 17 independent loci from GWAS analysis were enriched in mouse data. Our work identifies shared genetic components between specific types of obesity and asthma, and will help to advance our understanding of the molecular mechanisms underlying co-morbid obesity and asthma.

3251F

Polygenic risk stratified by eQTL influence for obesity and related comorbidities for patients from the Geisinger MyCode Community Health Initiative. M. Butkiewicz, S.A. Pendergrass. Department of Biomedical and Translational Informatics, Geisinger, Rockville, MD.

Background: Obesity is a major public health burden in the United States and is associated with other conditions, such as coronary artery disease, stroke, diabetes, and sleep apnea. There is substantial heritability of many obesity related comorbidities, and quantitative measures of genetic risk may be useful towards this end. Polygenic risk scores (PRS), in particular, can be calculated well before illness onset and have the potential for clinical use in health care. **Methods:** We used summary genetic association statistics for multiple obesity-comorbidity diagnoses available across individuals from the MyCode Community Health Initiative of Geisinger, and calculated PRS for 59,499 individuals of European-American ancestry from MyCode using all genetic variants with minor allele frequency > 1%. We also calculated PRS using the software PRSice (v2.1.2.beta) after filtering genetic variants only for those identified as expression quantitative trait loci (eQTLs) from The Genotype-Tissue Expression (GTEx) project. Next, associations between these PRS and comprehensive patient diagnoses were assessed using univariate logistic regression in a phenome-wide association study (PheWAS). Diagnoses were defined using 'phecodes': a hierarchical grouping of ICD-9 diagnostic codes. Cases for a given phecode had three or more relevant ICD-9 codes, and phecodes with 100 cases or fewer were excluded. Further, we created an obesity-based comorbidity network, linking obesity related ICD-9 codes to associated phecodes. **Results:** As expected, the obesity co-morbidity based PRS were associated with Type-1 ($p < 1 \times 10^{-35}$) and Type-2 diabetes ($p < 1 \times 10^{-10}$). However, when excluding SNPs that are eQTLs in the PRS calculations, we see an increase in statistical signal for obesity related PRS associations, like hypothyroidism (resultant $p = 1 \times 10^{-7}$) as well as a selection of known severe diabetes related comorbidities, such as foot and leg ulcers (resultant $p = 2 \times 10^{-4}$). The resulting network is an undirected graph representing disease traits as nodes and their associations as edges. **Discussion:** This study suggests a substantial difference in significance when using eQTLs for PRS calculation, showing the importance of filtering on functionally impactful SNPs. Next, we will investigate the impact of tissue specificity on PRS. In general, such scores could be applied in a hospital setting, elevating their potential for risk stratification efforts in the future.

3252W

Genome-wide analyses of secondary outcomes with electronic health records. C. Gao¹, A. Li¹, A. Marcketta¹, B. Ye¹, D. Carey², D. Ledbetter², M. Ritchie², C. O'Dushlaine¹, J. Overton¹, J. Reid¹, A. Baras¹, A. Shuldiner¹, T. Teslovich¹, F. Dewey¹, S. McCarthy¹. 1) Regeneron Genetics Center, Tarrytown, NY; 2) Geisinger Health System, Danville, PA.

Many insights into the genetic risk of disease have been gleaned through genome-wide association studies (GWAS), shedding light on novel biological mechanisms and opening avenues for therapeutic development. By extending GWAS to longitudinal electronic health records (EHR), new questions about disease risk and therapeutic outcomes can be assessed e.g. what are the genetic risks for disease co-morbidities, are treatment side-effects predictable, and can existing drugs be repurposed? In this study, we aim to investigate the genetic basis of secondary outcomes (co-morbidities, side-effects, repurposing) in DiscovEHR. Exomes from 88,584 European individuals were analyzed for genetic factors that impact the risk of Coronary Artery Disease (CAD) in patients with Type 2 Diabetes (T2D) using two-degree-of-freedom (2DF) joint analysis, 1DF interaction tests and sample-stratified analyses. The genetic risk for CAD and effects on four serum lipids measures were evaluated given T2D status and serum glycated hemoglobin (HbA1C) as interaction variables. Overall, 505 variants from 184 genes were significantly associated ($P < 5 \times 10^{-8}$) with serum lipid levels in 2DF joint test (156 variants and 47 genes located in HLA region). The E40K loss-of-function (LOF) variant (rs116843064) in Angiopoietin-like 4 (*ANGPTL4*) was strongly associated with reduced triglycerides (TG) and elevated high-density lipoprotein cholesterol (HDL-c) including potential interaction effects with HbA1C ($P_{2DF\_TG} = 1.32 \times 10^{-37}$, $P_{1DF\_TG} = 3.35 \times 10^{-63}$, $P_{2DF\_HDL} = 1.59 \times 10^{-26}$, $P_{1DF\_HDL} = 6.40 \times 10^{-64}$). On average, carriers of the LOF allele had 15.78mg/dl lower TG compared to non-carriers in the normal HbA1C group ($\leq 5.6\%$). However, in the high HbA1C group ($> 8\%$), the LOF allele had twice of the effect size (-37.56mg/dl per allele). Similar pattern was also observed for HDL-c (2.76mg/dl and 5.03mg/dl change per allele in normal and high HbA1C group, respectively). Consistently, a LOF variant (rs268) in the *ANGPTL4* target gene, lipoprotein lipase (*LPL*), also showed a moderate HbA1C dependent effect on TG and HDL ($P_{1DF\_TG} = 3.54 \times 10^{-63}$, $P_{1DF\_HDL} = 2.97 \times 10^{-62}$). In summary, exome-wide stratified and interaction analyses were performed and *ANGPTL4*-*LPL* pathway was identified as a potential target to reduce the risk of CAD in diabetics. Results from the secondary effect analysis may shed light on novel biological mechanisms underlying disease risk and inform drug discovery to develop medications for specific diseases complications.

3253T

Effect of adiponectin level and genetic variation of *ADIPOQ* and *CDH13* on diabetic retinopathy. W.L. Liao^{1,2}, Y.C. Huang^{3,4}, C.C. Chen^{2,5}, Y.H. Chen^{1,5,7,8}, Y.T. Chen⁹, C.M. Wu⁴, Y.W. Chang⁴, F.J. Tsa^{3,4,10}. 1) Graduate Institute of Integrated Medicine, China Medical University, Taichung, Taiwan; 2) Center for Personalized Medicine, China Medical University Hospital, Taichung, Taiwan; 3) School of Chinese Medicine, China Medical University, Taichung, Taiwan; 4) Human Genetic Center, Department of Medical Research, China Medical University Hospital, China Medical University, Taichung, Taiwan; 5) Division of Endocrinology and Metabolism, Department of Medicine, China Medical University Hospital, Taichung, Taiwan; 6) Departments of Chinese Medicine, China Medical University Hospital, Taichung, Taiwan; 7) Chinese Medicine Research Center, Research Center for Chinese Medicine & Acupuncture, China Medical University, Taichung, Taiwan; 8) Department of Psychology, College of Medical and Health Science, Asia University, Taichung, Taiwan; 9) Graduate Institute of Food Safety, National Chung Hsing University, Taichung, Taiwan; 10) Division of Pediatrics Genetics, China Medical University Children's Hospital, Taichung, Taiwan.

Background/Objective: The effects of adiponectin (APN) level and related APN receptors on diabetic retinopathy (DR) have been raised but the results were not consistence. In present study, we investigate the association between the plasma ANP level and variations in APN-related genes, including *ADIPOQ* and *CDH13*, individually and in combination, and DR. We also investigated the effect of genetic polymorphism and epigenetic modification on the expression of APN receptors. **Subjects/Methods:** Type 2 diabetes (T2D) patients (N = 1,604), above 20 years of age in Taiwanese population. Demographic information was collected from questionnaires, and blood pressure and serological markers were measured at enrollment. Genomic DNA was isolated from blood and genotyping. The relative mRNA levels were measured by qPCR and the DNA methylation level was examined using methylation-specific PCR method. The plasma APN level was assessed by enzyme-linked immunosorbent assay. **Results:** T2D patients with DR (N = 632) had diabetes for a longer duration, and had higher HbA1c, and systolic and diastolic blood pressure compared to those without DR (N = 972) ($P < 0.001$, for all the parameters). Overall, 10 SNPs in *ADIPOQ* and *CDH13* susceptibility loci were associated with DR. Furthermore, the genetic risk score (GRS) was calculated based on these 10 SNPs for each subject and the cumulative effect of genes was observed. Among the subjects with plasma APN level (N = 518), *natural logarithm* (LN) [APN] [OR was 1.63 (95% C.I. = 1.19–2.25)] and GRS [ORs (95% CI) for those in the middle and high range of GRS were 1.90 (1.11–3.26) and 2.61 (1.48–4.59) compared to those in the bottom range of GRS, respectively] were independent risk factors for DR after adjustment for age, duration of diabetes, HbA1c, and SBP in multivariate logistic regression model. Compared to non-DR subjects in the lowest range of GRS, the mRNA levels of T cadherin which encoded by *CDH13* were higher in non-DR subjects with middle/high GRS, DR subjects with low GRS, and DR subjects with middle/high. **Conclusions:** The plasma APN level and the genetic variations in adiponectin receptors were associated with DR. And the higher expression mRNA level of T cadherin among DR subjects with high GRS was observed.

3254F

Genome-transcriptome-wide association of renal and morphometric traits across glomerular and tubular tissues in diabetic Pima Native Americans. A. Liu¹, V. Nair¹, R. Nelson², R. Hanson², H. Looker², C. Boustany³, P. Guarnieri³, J. Hill³, M. Mauer⁴, M. Kretzler⁴, H.M. Kang¹. 1) University of Michigan, Ann Arbor, MI; 2) National Institute of Diabetes and Digestive and Kidney Diseases, Phoenix, AZ; 3) Boehringer Ingelheim, Ridgefield, CT; 4) University of Minnesota, Minneapolis, MN.

Diabetic kidney disease (DKD) is a global health concern and a leading cause of chronic kidney disease (CKD) and end-stage kidney disease. Understanding the etiology of DKD is challenging due to the lack of high quality and population-scale omics data from glomerular and tubular tissues from persons with type 2 diabetes (T2D). We deeply sequenced genomes of 97 diabetic Pima Indians with >10 years of exposures to T2D and assayed their transcriptomic profiles (using expression arrays and/or RNA-seq) from micro-dissected kidney biopsies [NR([1] collected at two time points 7-10 years apart. We then performed systems genetic analysis with clinical and morphometric phenotypes. We detected and genotyped 10.2 million high quality genetic variants from WGS and performed eQTL association analysis with glomerular and tubular transcriptomic profiles using linear mixed models, after accounting for reference bias by removing variant-overlapping probes and correcting for systematic confounding factors using Singular Value Decomposition. We identified 2,018 eGenes that contain cis-eQTLs in at least one pair of biopsy-tissues, revealing numerous tissue-specific cis-eQTL genes, including *VPS33B* from glomerular tissue and *FERMT1* from tubular tissue. We identified 5-fold larger cis-eQTL genes with RNA-seq than with expression arrays. These eQTL variants are highly enriched for significant previously reported GWAS associations relevant to CKD and ESRD. We then developed and applied methods that integrate information from external transcriptome association studies (e.g. GTEx) into GWAS, providing further insights and combined p-values for top GWAS signals. Genome-wide and transcriptome-wide association analysis identified significant associations with clinically important kidney phenotypes, including albuminuria (urine albumin-creatinine ratio) and measured glomerular filtration rate (iothalamate). Associations with quantitative morphometric traits identified many significantly associated genes that are also implicated as clinically relevant from independent studies. [NR([1]Consider specifying the number of paired kidney biopsies here.

3255W

Genome-wide association study identifies susceptibility loci for proliferative diabetic retinopathy in Africans. C. Liu^{1,2}, G.J. Chen¹, A.R. Bentley¹, D. Shriner¹, A.P. Doumatey¹, J. Zhou¹, J.K. Yang², A. Adeyemo², C.N. Rotimi¹. 1) Center for Research on Genomics and Global Health, National Human Genome Research Institute, Bethesda, MD. Bethesda, Maryland 20892 USA; 2) Beijing Key Laboratory of Diabetes Research and Care, Beijing Diabetes Institute, Beijing Tongren Hospital, Capital Medical University, Beijing 10730, China.

Proliferative diabetic retinopathy (PDR) is a sight-threatening complication of diabetes that is associated with longer duration of diabetes and poor glycemic control under a genetic susceptibility background. Although genome-wide association studies (GWAS) of PDR have been conducted in European and Asian ancestry individuals, none have been conducted in Africans, a population of interest given higher burden of undiagnosed and untreated type 2 diabetes (T2D), which may increase PDR risk. Here, we conduct a GWAS of PDR among Africans drawn from the AADM study. PDR cases (n = 44) were T2D patients with new vessel growth in the retina and/or retinal detachment. Controls (n = 138) were T2D patients without listed eye complications despite high risk (T2D duration ≥ 10 years and fasting blood glucose > 169 mg/dl; i.e., A1C > 7.5). Participants were genotyped on the MEGA array and imputation was performed using the African Genome Resources Panel through Sanger Imputation Service. EPACTs was used to perform association testing with adjustment for covariates (sex, age, and three significant PCs) and a random effect for cryptic relatedness. We identified 4 significant loci: *WDR72*, *TMEM132B*, *KLHL3*, and *AMFR*. The T allele for rs12906891 (*WDR72*) was associated with increased risk of PDR (freq = 0.08 p = 1.1×10^{-8} , OR = 1.60, 95%CI [1.38, 1.84]); variants in this gene have been associated with glycemic control. At *TMEM132B*, the lowest p-value was for rs4073534 (T allele freq = 0.04; p = 2.6×10^{-8} , OR = 0.48, 95%CI [0.38, 0.6]). *KLHL3* variant rs72800063 was associated with increased PDR risk (G allele freq = 0.03, p = 1.2×10^{-8} , OR = 2.0, 95%CI [1.6, 2.5]); *KLHL3* is a known risk variant for hypertension, a condition that increases PDR risk. African ancestry-specific variant rs148177825 (*AMFR*) was associated with increased PDR risk (G allele freq = 0.03, p = 2.7×10^{-8} , OR = 2.1, 95%CI [1.63, 2.69]). *AMFR* has been associated with traits relevant for PDR, including insulin resistance, vascular endothelial growth factor, hypoxia, and cell proliferation. Despite the small sample size, this first GWAS of PDR in Africans identified 4 susceptibility loci, perhaps reflecting our strict definition of controls to include only those at high risk but did not develop PDR, thus minimizing inclusion as controls those who would eventually progress to PDR. Known associations of these loci with relevant traits highlight expected pathways, including glycemic control, hypertension, and revascularization.

3256T

Rare variants in *SLC5A10* are associated with serum 1,5-anhydroglucitol (1,5-AG) in the Atherosclerosis Risk in Communities (ARIC) Study. S. Loomis¹, A. Köttgen^{1,2}, M. Li³, A. Tin¹, J. Coresh^{1,4}, E. Boerwinkle⁵, R. Gibbs⁶, D. Muzny⁶, J. Pankow⁷, E. Selvin^{1,4}, P. Duggal¹. 1) Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; 2) Institute of Genetic Epidemiology, Medical Center and Faculty of Medicine - University of Freiburg, Freiburg, Germany; 3) Division of Nephrology and Department of Human Genetics, University of Utah, Salt Lake City, Utah; 4) Welch Center for Prevention, Epidemiology, & Clinical Research, The Johns Hopkins University, Baltimore MD; 5) Department of Epidemiology, The University of Texas Health Science Center at Houston School of Public Health at Houston, Houston, TX; 6) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 7) Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN.

Introduction: Serum 1,5-anhydroglucitol (1,5-AG) is a measure of hyperglycemic excursions that is available as a clinical test to aid monitoring of glycemic control in persons with diabetes. Recent genome-wide association studies (GWAS) have identified several 1,5-AG associated common variants. Rare genetic variants may have large effects on the concentrations of biomarkers that are in clinical use. The association between rare coding variants and 1,5-AG concentrations measured with a targeted assay of 1,5-AG was investigated in individuals of European and African ancestries. **Methods:** Whole exome sequencing association analysis was performed on 1,5-AG among European ancestry (N=6,589) and African ancestry (N=2,309) participants without diagnosed diabetes in the Atherosclerosis Risk in Communities (ARIC) Study. Both single variant and gene-based tests were conducted. **Results:** Five variants representing 3 independent signals on chromosome 17 in *SLC5A10* were associated with lower 1,5-AG levels of up to 10.38 µg/mL per allele (1,5-AG range 3.4–32.8 µg/mL). Two of these variants (rs61741107, $p=8.85E-56$, amino acid change: G>E; rs148178887, $p=1.13E-36$, amino acid change: N>I) with similar allele frequencies in the European (MAF=0.002 to 0.007) and African (MAF=0.0005 to 0.004) ancestry samples were rare, nonsynonymous, and predicted to be damaging or deleterious by multiple algorithms that predict deleteriousness. Taken together, these three signals explained 6% of the variance in 1,5-AG. Gene-based SKAT-O tests were significant for *SLC5A10* in the European ancestry ($p=5.13 \times 10^{-64}$) and validated in the African ancestry ($p=0.006$) samples. Significant common and low frequency variants were identified at five other loci in or near *MUC1*, *LCT*, *SI*, *MGAM*, and *SLC5A1* that also contain common 1,5-AG associated variants. **Conclusions:** Several rare, protein-altering variants in individuals of European ancestry were associated with 1,5-AG concentrations and were validated in an African ancestry sample. The large effect sizes for *SLC5A10* variants and similar allele frequencies across populations without diabetes, along with the multiple independent signals are evidence of the important impact of *SLC5A10* on serum 1,5-AG concentrations, and suggest *SLC5A10* may code for an important transporter of 1,5-AG in the kidney. This study helps to characterize the genetic architecture of 1,5-AG, an emerging diabetes biomarker, and may have implications for its clinical utility.

3257F

Undiscovered novel polymorphisms in the SOWAHC-RGPD5 intergenic region on chromosome 2q are associated with diabetic retinopathy. S. Y. Rhee¹, J. Y. Lee², D. W. Lee², S. Chon¹, J. T. Woo², S. J. Kim². 1) Endocrinology and Metabolism, Kyung Hee University School of Medicine, Seoul, Korea, Republic of; 2) Theragen Etx, Inc., Suwon, Korea, Republic of.

Diabetic retinopathy (DR) is a significant diabetic microvascular complication that causes visual disturbance and blindness in adults. We conducted a genome-wide association study of 426,593 single nucleotide polymorphisms (SNPs) in 1,456 subjects (741 cases and 715 controls). We followed up with an imputation analysis using approximately 6,198,263 SNPs and then validated the results with an independent cohort that included 974 subjects (553 cases and 421 controls). We identified significant genome-wide novel variants, such as imputed rs34229733 ($P = 3.62 \times 10^{-9}$) and genotyped rs17197569 ($P = 7.32 \times 10^{-6}$), in the intergenic region of *SOWAHC-RGPD5* (Sosondowah Ankyrin Repeat Domain Family Member C, RANBP2-Like and GRIP Domain Containing 5) on chromosome 2. Furthermore, we observed marginally associated rs17741348 and rs8077896 (*STX8*, $P = 1.45 \times 10^{-7}$ and $P = 4.40 \times 10^{-6}$), rs35774187 (*SASH1*, $P = 2.52 \times 10^{-6}$), rs2734363 (*FHIT*, $P = 9.56 \times 10^{-6}$), rs7730344 (*ANKRD31-HMGCR*, $P = 9.66 \times 10^{-6}$), and 26 loci with statistical significance. Further imputation and validation analyses revealed that *AREL1* (rs9671386, $P = 1.42 \times 10^{-5}$), *MED13L* (rs8181760, $P = 4.52 \times 10^{-5}$), *PLCB1* (rs6108152, $P = 1.36 \times 10^{-5}$), *SASH1* (rs35774187, $P = 4.03 \times 10^{-5}$), and *SLC24A3* (rs6035258, $P = 6.79 \times 10^{-5}$) were associated with DR. Moreover, after a comparison to previously reported genes, the matched SNPs did not show statistical significance; however, *GLIS1*, *MCPH1*, *PLXDC2*, and *VEGFA* were associated with DR. These findings suggest that genetic variation in the *SOWAHC-RGPD5* intergenic region plays an important role in the pathogenesis and development of severe DR, and may explain why some patients with type 2 diabetes are less susceptible to microvascular complications than patients with DR.

3258W

Genome-wide association study of type 2 diabetes phenotypes in

Sāmoans. E.M. Russell¹, J.C. Carlson¹, N.L. Hawley², G. Sun³, H. Cheng³, T. Naseri⁴, M. Sefuiva Reupena⁵, R. Deka⁶, D.E. Weeks^{1,6}, S.T. McGarvey^{7,8}, R.L. Minster¹, TOPMed Diabetes Working Group. 1) Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA; 2) Department of Epidemiology (Chronic Disease), School of Public Health, Yale University, New Haven, CT; 3) Department of Environmental Health, College of Medicine, University of Cincinnati, Cincinnati, OH; 4) Ministry of Health, Government of Samoa, Apia, Samoa; 5) Bureau of Statistics, Government of Samoa, Apia, Samoa; 6) Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA; 7) International Health Institute, Department of Epidemiology, School of Public Health, Brown University, Providence, RI; 8) Department of Anthropology, Brown University, Providence, RI.

Type 2 diabetes (T2D) is prevalent in the Independent State of Sāmoa in Polynesia with an age-standardized prevalence of 22.7% for men and 26.6% for women. Unique genetic factors may exist that contribute to this high prevalence. In addition, the effect of genes on susceptibility to T2D can differ between obese and lean cases. The Sāmoan population is genetically isolated, making it a prime candidate for genetics studies. We used whole-genome sequence data generated through NHLBI's TOPMed Program for 1,195 Sāmoans to create a Sāmoan-specific reference panel to impute genotypes for an additional 1,897 Sāmoans. To identify genes and variants associated with T2D phenotypes, we completed genome-wide association studies (GWAS) of (1) T2D, (2) T2D in obese (Polynesian cutoff of BMI $\geq$ 32 kg/m²) and (3) T2D in non-obese with 9,609,170 sequenced and imputed variants, adjusting for age, sex, principal components of ancestry and empirical kinship (model 1). We also adjusted for BMI in model 2. In the T2D GWAS ($n=478$ cases, $n=2,389$ controls), a variant in a known T2D gene, *PPARG-C1A*, was genome-wide significant in both models 1 and 2 ($p=1.30\times 10^{-9}$ and $p=8.32\times 10^{-9}$, respectively). This variant is common both in other TOPMed populations (minor allele frequency (MAF)=0.376) and in Sāmoans (MAF=0.250). In the non-obese T2D GWAS ($n=148$ cases, $n=1,119$ controls), variants in or near three BMI-associated genes were significantly associated with T2D in models 1 and 2 (in *ADAM23*, $p=6.67\times 10^{-11}$ and $p=2.08\times 10^{-11}$, respectively; near *SATB2*, $p=1.37\times 10^{-10}$ and $p=6.68\times 10^{-10}$; and near *CREB1*, $p=1.17\times 10^{-9}$ and $p=7.02\times 10^{-10}$), and a variant in *MAP2*, which has not been previously associated with T2D or related traits, was genome-wide significant in models 1 and 2 ($p=4.80\times 10^{-10}$ and $p=5.22\times 10^{-11}$, respectively). These four variants are common in other TOPMed populations (MAF:0.094–0.179), but rare in Sāmoans (MAF<0.009). In the obese T2D GWAS ($n=330$ cases, $n=1,270$ controls), no variants were genome-wide significant. However, a suggestive (model 1: $p=5.38\times 10^{-8}$) variant near known BMI gene *TMEM182* is notable because this variant is rare in other TOPMed populations (MAF=0.0003), but common in Sāmoans (MAF=0.151). Additional exploration is necessary to further replicate these novel associations and characterize these variants. This highlights the importance of including diverse populations in genetics research to further characterize a map of disease-associated variation.

3259T

Genome-wide association study identifies novel associations with serum cytokine levels in southwestern native americans. L.E. Wedekind¹, M.F. Walter², S. Kobes¹, P. Chen¹, W.C. Hsueh¹, R.G. Nelson¹, L.J. Baier¹, W.C. Knowler¹, R.L. Hanson¹. 1) National Institute of Diabetes and Digestive and Kidney Disorders; Phoenix, AZ; 2) National Institute of Diabetes and Digestive and Kidney Disorders; Bethesda, MD.

Circulating cytokine levels may reflect important molecular processes leading to type 2 diabetes, but there is limited information about genetic associations with levels of circulating cytokines. We performed a genome-wide association study (GWAS) of serum cytokine levels in 1061 non-diabetic Southwestern Native Americans (500 female; 561 male). We measured fasting levels of 6 traits—adiponectin, adipsin, amylin, glucose-dependent insulinotropic peptide (GIP), leptin and resistin—using EMD Millipore assays (Sigma-Aldrich, St Louis, MO). We conducted a GWAS involving 496,190 single-nucleotide polymorphisms (SNPs) with minor allele frequency >1% on a custom Axiom array designed to capture common variation in Pima Indians (Affymetrix, Santa Clara, CA). Imputation was performed based on whole-genome sequence data from 266 Pima Indians; a total of 4.8M variants were analyzed. Data were normalized by inverse Gaussian transformation and analyzed for association using a mixed model accounting for familial relationships; adjusted for age, sex and genetic principal components 1-5. A genomic control procedure was performed to account for residual population stratification. The strongest genetic associations with respect to serum levels of each trait were as follows: adiponectin with rs188692736 near *KDR* on chromosome 4 ($\beta=0.32$ SD per copy of the A allele, $p=6.21\times 10^{-7}$), adipsin with rs146930006 near *LOC105372762* on chromosome 21 ($\beta=0.26$ SD per G allele, $p=1.47\times 10^{-6}$), amylin with rs117681391 in *LINC00290* on chromosome 4 ($\beta=0.53$ SD, $p=2.05\times 10^{-6}$), GIP with rs1975618 in *MCPH1* on chromosome 8 ($\beta=0.22$ SD per A allele, $p=1.03\times 10^{-6}$), leptin with rs61801876 on chromosome 1 ($p=1.08\times 10^{-6}$), and resistin with rs190461045 in *B3GALT1* on chromosome 2 ($\beta=0.30$ SD, $p=1.60\times 10^{-6}$). In summary, these studies have identified several variants with suggestive genetic associations with each of six different cytokine levels. For some of these variants minor allele frequencies are low for populations outside of Native Americans—suggesting possibly unique signals. These associations need to be confirmed in additional cohorts, but may yield important information about molecular processes contributing to complex traits.

3260F

Rare variation in and near *CREBRF* and association with fasting glucose in non-Polynesian participants in the TOPMed Program. R.L. Minster¹, E.M. Russell¹, N.L. Hawley², G. Sun³, H. Cheng³, T. Naseri⁴, M.S. Reupena⁵, R. Deka³, D.E. Weeks^{1,6}, S.T. McGarvey^{7,8}, TOPMed Diabetes Working Group. 1) Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA; 2) Epidemiology (Chronic Disease), School of Public Health, Yale University, New Haven, CT, USA; 3) Environmental Health, College of Medicine, University of Cincinnati, Cincinnati, OH, USA; 4) Ministry of Health, Government of Samoa, Apia, Samoa; 5) Bureau of Statistics, Government of Samoa, Apia, Samoa; 6) Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA; 7) International Health Institute, Epidemiology, School of Public Health, Brown University, Providence, RI, USA; 8) Anthropology, Brown University, Providence, RI, USA.

The minor allele of rs373863828, a missense variant in *CREBRF*, has been associated with lower fasting glucose (FG) in Samoans. This variant is common in Polynesians (reported minor allele frequency [MAF] range: 0.096–0.259) but very rare in non-Polynesian populations (MAF < 0.001). Other variation in the region may affect FG in other populations. We explore here whether rare variation in the *CREBRF* region is associated with FG in non-Polynesians in NHLBI's TOPMed Program. We used whole-genome sequencing data from 11 TOPMed studies with reported FG and excluded individuals with type 2 diabetes. We divided the sample into four groups based on self-reported ancestry (African, $n = 6,551$; Asian, $n = 2,309$; European, $n = 18,013$; and Hispanic, $n = 1,054$). We used SKAT-O to test rare variation for association with FG, adjusting for age, sex, BMI, and ancestry. We selected rare SNPs in the *CREBRF* region (gene boundaries ± 50 bp) with a TOPMed-wide minor allele count ≤ 10 . Within each ancestry group, we excluded SNPs with a MAF > 0.05. First, we used a sliding window approach covering the region with 41 windows (4 kbp with 2 kbp overlap) with 28 to 348 variants per window and a median of 106. We set a conservative significance threshold of $p < 6 \times 10^{-4}$ based on the number of non-overlapping windows and ancestry groups ($0.05 / (21 \times 4)$). We found a significant association between a European window of 263 variants and FG ($p = 2.9 \times 10^{-4}$). The window comprises the last 3.3 kbp of intron 8 and the first 0.7 kbp of exon 9, including the stop codon. However, we did not see significant evidence of association in other ancestries, and the location of the peak association was not consistent between ancestry groups (African: $p = 0.014$, 185 variants; Asian: $p = 0.023$, 76 variants; Hispanic: $p = 1.7 \times 10^{-3}$, 62 variants). Second, we completed a gene-based analysis of the same *CREBRF* region. Our significance threshold was $p < 0.013$ given the number of ancestry groups tested ($0.05 / 4$). In this analysis, we did not see significant evidence of association of *CREBRF* with FG in any ancestry (African: $p = 0.097$, 3,319 variants; Asian: $p = 0.261$, 1,671 variants; European: $p = 0.027$, 4,643 variants; Hispanic: $p = 0.018$, 1,336 variants). These results suggest that rare variation in the *CREBRF* region could be affecting the variance of FG in Europeans. It is possible that with more power association between rare variation and FG could be detected in other ancestries.

3261W

Rare variants explain a substantial portion of the missing heritability of the uric acid level. K. Misawa^{1,2}, E. Mishima³, M. Ouchi⁴, T. Hasegawa⁵, K. Kojima^{1,2}, Y. Kawai⁶, M. Nagasaki^{1,2}, N. Anzai⁷, T. Abe⁸, M. Yamamoto^{1,2}. 1) Tohoku Medical Megabank Organization, Tohoku University, Sendai, Miyagi, Japan; 2) Graduate School of Medicine, Tohoku University, Sendai, Miyagi, Japan; 3) Division of Nephrology, Endocrinology and Vascular Medicine, Tohoku University Hospital, Sendai, Miyagi, Japan; 4) Department of Pharmacology and Toxicology, Dokkyo Medical University School of Medicine, Tochigi, Japan; 5) The Institute of Medical Science, The University of Tokyo, Tokyo, Japan; 6) Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; 7) Department of Pharmacology, Chiba University Graduate School of Medicine, Chiba, Japan.

Gout is one of the most common inflammatory arthritides, caused by hyperuricemia. Hyperuricemia is an abnormally high level of uric acid in the blood. The prevalence of hyperuricemia in Japanese adult men is estimated to be nearly 30%. A twin study showed that the heritability of urate levels is estimated to be approximately 60%. However, the proportion of variance in urate levels explained by common genetic variants detected in the previous GWAS was about 6.0%, accounting for only a small proportion of the genetic component. This problem has been known as the "missing heritability". In this study, we focused on the rare variants as the causes of the missing heritability of SUL. We compared the serum uric acid level (SUL) of 2,049 participants of ToMMo cohort study to their whole genome sequences. By using SKAT, we detected one transporter gene whose rare variants might be associated with SUL. Then we conducted the functional analyses the variants identified the rare-variant analysis experimentally on splicing by using Human Embryonic Kidney cells 293 (HEK293) as well as on uric acid uptake by using the *Xenopus* oocytes. These experimental studies showed that the rare variants are affecting on urate uptake. We estimated the effects of the rare variants found in this study on SUL by using the regression analyses that include age, gender, BMI, drug, and food intake as covariates of SUL. We also used eGFR as the measure of the level of kidney function and HbA1s as the measure of diabetic status. The regression analyses showed that they account for more than 10% of the variation of SUL. The result suggests that rare variants explain a substantial portion of the missing heritability of the uric acid level.

3262T

Replicating type 2 diabetes GWAS results using self-reported data. A. Annis, A. Pandit, S. Gagliano, G. Abecasis. University of Michigan, Ann Arbor, MI.

The efficacy of physician- vs. self-diagnosed phenotypes in GWAS is disputable, particularly for Type 2 Diabetes (T2D) where much of the affected population is undiagnosed. Some researchers fear that self-diagnosis leads to underreporting and failure to replicate associated SNPs, while others hold that self-diagnosis is an effective means of data collection. We address this issue by comparing T2D GWAS results from a primarily physician-diagnosed meta-analysis (DIAGRAM) with results from a smaller, self-reported dataset (Genes for Good). DIAGRAM's most recent publication, Mahajan et al (2018), gives a meta-analysis of 74,124 T2D cases and 824,006 controls, largely drawn from physician-diagnosed studies. Genes for Good (GfG), a University of Michigan study, relies on questionnaires administered through a social-media-based web application to gather individuals' self-reported phenotypes. GfG currently has 848 self-reported T2D cases and 16,898 controls. We identified the ten most strongly associated independent SNPs from DIAGRAM's (BMI-unadjusted) GWAS and then found the same SNPs in the GfG dataset for comparison. For our analysis, we first used the online GAS Power Calculator to find GfG's statistical power for replicating the DIAGRAM associations at a 0.05 significance level. We then compared p-values, odds ratios (ORs), and confidence intervals (CIs) from DIAGRAM with those calculated in GfG at the same SNPs. Using DIAGRAM's disease prevalence (10%), disease allele frequencies, and genotype relative risks, we found that GfG averaged 45% power for replicating the DIAGRAM associations at the 0.05 significance level. Five of these 10 SNPs had p-values below the 0.05 significance level in GfG, which is consistent with our power calculation. We also found all ten effect sizes in GfG and DIAGRAM to have consistent direction, and nine out of the ten ORs from DIAGRAM fell within the GfG CIs. As we would expect our CIs to contain the true effect size 95% of the time, our results coincide with our expectations. The consistency of our GfG findings with the DIAGRAM results suggests that self-reported T2D diagnoses are an effective means of collecting data for GWAS. Although GfG remains underpowered due to sample size, we were able to replicate five of the ten associations identified in the much larger DIAGRAM meta-analysis. This shows that GWAS with self-reported data has potential to produce meaningful results and contribute to our understanding of human genetics.

3263F

A genome-wide and phenome-wide association study in individuals of African Ancestry identified novel loci of body mass index and obesity-related diseases. Y. Huang¹, J. Huang², T. Assimes^{3,4}, K. Cho², Q. Hui¹, Y. Ho², S.M. Damrauer^{5,6}, J. Honerlaw⁷, D. Miller^{7,8}, L. Phillips^{9,10}, P. Reaven^{11,12}, H. Tang¹³, M.C.Y. Ng^{14,15}, M. Graff¹⁶, C.A. Haiman^{17,18}, R.J.F. Loos^{19,20}, K.E. North¹⁶, S. Pyarajan^{2,21}, M. Gaziano^{2,21}, P. Tsao^{3,4}, K. Chang^{2,22}, P. Wilson^{2,23}, D. Saleheen²⁴, Y.V. Sun^{1,25}, C.J. O'Donnell², *African Ancestry Anthropometry Genetics Consortium, The VA Million Veteran Program.* 1) Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA; 2) Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC), VA Boston Healthcare System, Harvard Medical School, Boston, MA; 3) Department of Medicine, Stanford University School of Medicine, Stanford, CA; 4) VA Palo Alto Health Care System, Palo Alto, CA; 5) Department of Surgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 6) Corporal Michael Crescenz VA Medical Center, Philadelphia, PA; 7) Center for Healthcare Organization and Implementation Research, Edith Nourse Rogers Memorial Veterans Hospital, Bedford, MA; 8) Boston University School of Public Health, Boston, MA; 9) Atlanta VA Medical Center, Decatur, GA; 10) Division of Endocrinology, Emory University, Atlanta, GA; 11) Phoenix VA Health Care System, Phoenix, AZ; 12) Department of Medicine, University of Arizona, Phoenix, AZ; 13) Department of Genetics, Stanford University School of Medicine, Stanford, CA; 14) Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, NC; 15) Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, NC; 16) Department of Epidemiology, University of North Carolina, Chapel Hill, NC; 17) Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA; 18) Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA; 19) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY; 20) The Mindich Child Health and Development Institute, Ichan School of Medicine at Mount Sinai, New York, NY; 21) Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 22) Institute for Translational Medicine and Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 23) Emory Clinical Cardiovascular Research Institute, Atlanta, GA; 24) Department of Biostatistics and Epidemiology, Perelman School of Medicine, Philadelphia, PA; 25) Department of Biomedical Informatics, School of Medicine, Emory University, Atlanta, GA.

Body mass index (BMI), a commonly used estimate of obesity, is highly heritable and a leading risk factor for type 2 diabetes, cardiovascular disease, some cancers, and premature death. Genome-wide association studies (GWAS) have identified many genetic variants that influence BMI mostly in populations of European descent, but only a few loci have been established in populations of African ancestry. We performed a large GWAS of BMI in up to 98,277 individuals of African ancestry. The discovery GWAS of BMI was conducted among 55,525 non-Hispanic African American participants in the Million Veteran Program (MVP) with over 47 million 1000 Genome phase 3 imputed SNPs. BMI-associated SNPs ($p < 5e-8$) from the MVP were replicated among 42,752 participants from the African Ancestry Anthropometry Genetics Consortium (AAAGC). Genome-wide significant ($p < 5e-8$) SNPs were identified in a fixed effect meta-analysis combining all participants from the MVP and the AAAGC, followed by functional annotation and bioinformatic analyses. In the phenome-wide association study (PheWAS), two BMI-polygenic risk scores (PRS) were developed based on: 1. Ten variants from established BMI loci of African ancestry by the AAAGC; 2. 100 variants from 41 BMI loci identified by current meta-analysis. We tested associations of BMI-PRS with >1,800 ICD-9 diagnosis codes from 16 disease groups among non-Hispanic African American participants in the MVP. Among 19 BMI loci identified in the MVP, 12 were replicated in the AAAGC (consistent direction and $p < 0.05$). In the combined meta-analysis, 985 SNPs from 41 loci reached genome-wide significance, of which 63 SNPs were analyzed only in the MVP. By comparing to all previous GWAS findings from populations of European or African ancestry, 10 loci were novel (distance to lead SNP of published loci >500 kb). Genes mapped by the novel BMI loci include *MOB1B*, *KCNQ5*, *SMCR5* and *CCDC85A*. PheWAS using established BMI loci from AAAGC identified associations with type 2 diabetes and heart disease, while the expanded BMI-PRS identified further associations with sleep apnea and orthopnea. Using the largest sample of African ancestry, we identified 41 BMI loci, including 10 novel loci not reported in previous GWAS in populations predominantly of European ancestry. PheWAS of BMI-PRS can explore novel associations between BMI and disease outcomes, and generate hypotheses for causal role of BMI for testing in future Mendelian Randomization analysis.

3264W

Evidence for bidirectionally-causal relationships between dental disease and metabolic health. S. Haworth¹, I. Johansson², P.W. Franks^{3,4,5}, E. Ingels-son^{6,7}, N.J. Timpson¹, GLIDE consortium. 1) MRC Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom; 2) Department of Odontology, Umeå University, Umeå, Sweden; 3) Department of Clinical Sciences, Genetic and Molecular Epidemiology Unit, Lund University, Malmö, Sweden; 4) Department of Public Health and Clinical Medicine, Umeå University, Umeå, Sweden; 5) Department of Nutrition, Harvard T. H. Chan School of Public Health, Boston, MA; 6) Department of Medicine, Stanford University School of Medicine, Stanford, CA; 7) Stanford Cardiovascular Institute, Stanford University, Stanford, CA.

Background. Dental caries and periodontal disease are associated with many systemic diseases in epidemiological studies, potentially through metabolic effects. It has been argued that dental diseases increase risk of systemic metabolic disturbances through systemic spread of oral bacteria or inflammatory by-products. Conversely, it is also believed that systemic metabolic changes affect homeostasis of the oral environment, leading to increased risk of dental disease. Recent analysis within the Gene Lifestyle Interactions and Dental Endpoints (GLIDE) consortium identified genetic correlation between dental disease and metabolic health traits. The aim of the study was to test these relationships for causal effects. **Methods.** All analyses were undertaken using generalized summary Mendelian randomization (GSMR), implemented in GCTA. HEIDI-outlier analysis was used to assess for horizontal pleiotropy. Summary statistics for 6 metabolic disease traits were taken from published consortia. Summary statistics for dental disease (n=487,898) were taken from combined analysis of a dental disease index in GLIDE and a dental disease proxy endpoint in UK Biobank (dentures). **Results.** In GSMR analyses, dental disease was seen to have bidirectional causal relationships with metabolic traits. Dental disease increased risk of type 2 diabetes (OR 1.37 per 20 diseased tooth surfaces, p=0.00086), increased body mass index (BMI, p=1.4e-12) and decreased HDL-cholesterol (p=2.2e-05). In reciprocal analysis, metabolic traits were seen to influence dental disease. Each 1 mM increase in fasting glucose resulted in approximately 1.2 additional diseased tooth surfaces (p=0.012) and increasing BMI also had detrimental effects on dental disease (p=6.5e-11). There was evidence for pleiotropic genetic effects. **Discussion.** This study supports prior observational studies and clinical belief that dental and metabolic disease are closely intertwined, through a combination of forward-causal, reverse-causal and pleiotropic mechanisms. Analyses are therefore planned to characterize these relationships further and investigate whether dental diseases causally influence health outcomes including cardiovascular disease endpoints.

3265T

Mendelian randomization analysis provides evidence for a casual role of serum triglyceride leads to increased glycated hemoglobin in the population of Taiwan. C. Hsiung, C. Lin, C. Chang, C. Shen. Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan.

The relationships between elevated circulating triglyceride (TG) levels and elevated glucose level/HbA1c/type 2 diabetes (T2D) have been reported in many studies. However, the causal relationship of these associations is ambiguous. This is because complex interactions among fat, muscle, liver, pancreas, and the intestine. Furthermore, these interactions may not be the same among populations because differences in diet, dietary habits, and more importantly genetic predisposition. Mechanistically, both sides of relationships are possible. To clarify this relationship, we used a Mendelian randomization (MR) analysis to test the hypothesis that elevated circulating TG levels causally results in elevated circulating HbA1c level. To employ a two-sample MR approach, we used summary results from GWAS catalog database, and seven SNPs reaching the GWAS significant ($P < 5 \times 10^{-8}$) specially in East-Asian population, being genomically independent to each other, and without pleiotropic effect were selected as instrumental variable (IV) to represent TG. We used Taiwan Biobank database in order to identify the casual effects between serum TG and HbA1c in the population of Taiwan. The Taiwan biobank database has completed whole-genome genotyping typing in more than 16,000 individuals. Associations using the MR estimate and the MR-inverse variance weighted method analysis show a unit increase in TG (measured by a log-transformed unit, mg/dl) was associated with a significant increase of 10.86 units of HbA1c (indicated by the beta value of regression model, 95% CI, 1.16-20.56; $P=0.028$). Sensitivity analysis, including MR-egger regression and the weighted-median approach provides no evidence to pleiotropic effect of IV. To confirm our results, we performed bidirectional MR analysis. Our results show that the genetic determinants of HbA1c don't contribute to TG, suggested by a non-significant beta value. (beta: 1.75, 95% CI: -11.50-14.90). These results suggest that elevated circulating triglyceride level casually affects circulating elevated level of glycated hemoglobin.

3266F

Non-alcoholic Fatty Liver Disease (NAFLD) increases risk for Type 2 Diabetes (T2D) but decreases susceptibility to obesity: A Mendelian randomization study. Z. Liu¹, R. Pique-Regi^{2,3}, W. Liu^{1,4,5,6}. 1) Department of Medicinal Chemistry and Molecular Pharmacology, College of Pharmacy, Purdue University, West Lafayette, IN; 2) Center for Molecular Medicine and Genetics, School of Medicine, Wayne State University, 540 E. Canfield, Detroit, MI; 3) Department of Obstetrics and Gynecology, School of Medicine, Wayne State University, Detroit, MI; 4) Department of Pharmaceutical Sciences, Eugene Applebaum College of Pharmacy and Health Sciences, Wayne State University, Detroit, MI; 5) Department of Pharmacology, School of Medicine, Wayne State University, Detroit, MI; 6) Integrative Biosciences Center, Detroit, MI.

Background Non-alcoholic fatty liver disease (NAFLD) is highly correlated with type 2 diabetes (T2D) and obesity in many observational studies. However, the causal relationship between them has not been completely investigated yet. **Aim** We aim to test whether genetic factors conferring risks to NAFLD increase the susceptibility to T2D and obesity. **Design** Two-sample Mendelian Randomization using published data **Methods** The genome-wide associated variants for computerized tomography (CT) measured hepatic steatosis were searched in a meta-analysis of 7,176 individuals of European descent from the GOLD (Genetics of Obesity-related Liver Disease) consortium. We considered two of the top associated variants, *PNPLA3* rs738409 and *NCAN* rs2228603, and tested their associations with T2D and Obesity in individuals from the DIAGRAM (Diabetes Genetics Replication And Meta-analysis) consortium, MAGIC (the Meta-Analyses of Glucose and Insulin-related traits) consortium, and the Genetic Investigation of Anthropometric Traits (GIANT) consortium. We conducted the two-sample MR analysis using each of the two variants as instrumental variable (IV). We also performed the MR analysis using the combined effects of the two variants and compared the results. **Results** MR analysis using either *PNPLA3* rs738409 or *NCAN* rs2228603 as instrumental variable gave similar results. In a cohort consists of 12,171 T2D cases and 56,862 controls of European decent, we found that genetically instrumented NAFLD was significantly associated with a higher risk of T2D (odds ratio [OR] per 1-SD higher fatty liver, 1.29, 95%CI [1.12, 1.49], $p = 0.001$). Statistically significant associations were also found for genetically instrumented hepatic steatosis and increased insulin resistance ($p = 0.016$), fasting glucose ($p = 0.027$) and fasting insulin ($p = 0.02$) levels. With regards to the relationship between NAFLD and obesity, surprisingly, we also found a moderate, but significant correlation between genetically increased hepatic steatosis and decreased BMI ($b = -0.046$ kg/m², 95%CI [-0.078, -0.014] kg/m², $p = 0.005$) in a cohort consists of 209,895 individuals of European decent. Similar negative relationship ($b = -0.028$ kg/m², 95%CI [-0.046, -0.011] kg/m², $p = 0.002$) was estimated using another sample of GWAS for BMI with a larger sample size of 420,805 participants. **Conclusion** Lifelong, genetically contributed NAFLD is a causal risk factor of T2D, but plays a protective role for obesity.

3267W

The human gut microbiome and obesity – causal or confounded companionship? K.H. Wade¹, D. Hughes¹, L. Corbin¹, J. Wang², R. Bacigalupe³, N.J. Timpson¹, J. Raes³. 1) MRC Integrative Epidemiology Unit, Population Health Sciences, Bristol Medical School, Bristol, United Kingdom, BS8 2BN; 2) Institute of Microbiology, Chinese Academy of Sciences, Chaoyang District, Beijing, 100101; 3) VIB-KU Leuven Center for Microbiology, Gasthuisberg Campus, Rega Instituut, 3000 Leuven.

Despite the exceptional growth in studies implicating the gut microbiome in health, there has been very little causal assessment within humans. We aimed to assess causality in the association between gut microbial taxa abundance and body mass index (BMI) using bi-directional, two-sample Mendelian randomization (MR). This work is based on data from the Flemish Gut Flora Project (FGFP), a population-level microbiome collection in Flanders, Belgium (primary analysis, $N=2259$). Microbiome phylogenetic profiling was characterised with 16S ribosomal RNA gene amplicon sequencing of the V4 hypervariable region. Given the equilibrated range of age, sex, health and life-style, FGFP is likely to represent the average gut microbiota composition in a Western European population. Observational associations between BMI and microbial taxa abundance were assessed using Spearman's correlation. Of the microbial taxa with evidence supporting an observational relationship, MR analyses were conducted to assess causality in a bi-directional framework. For MR analyses assessing the impact of the microbiome on BMI, the best proxies for specific bacterial taxa were obtained from the genome-wide association study (GWAS) of the rank-normalised raw counts of microbial taxa abundance conducted in FGFP. For MR analyses assessing the reverse direction (i.e., impact of BMI on the microbiome), 97 genetic variants robustly associated with BMI from the GIANT GWAS were used as instruments. All analyses were conducted using the MR-Base platform (www.mrbase.org), a large-scale database of GWAS summary-level data, enabling automated and efficient MR analyses. The inverse-variance weighted (IVW), weighted median, weighted mode and MR-Egger methods were applied. Observational analyses provided evidence that multiple taxa were correlated with BMI to a varying degree (range of correlation coefficients = -0.16, 0.14; $P < 1 \times 10^{-66}$). Follow-up MR analyses provided evidence that a higher abundance of certain microbial taxa may influence BMI (e.g., maximum effect = -0.02 kg/m²; $P = 0.05$ per SD higher level of count-derived abundance). In contrast, greater BMI may lead to a variation in the count-derived abundance of gut microbial taxa (e.g., maximum effect = 0.33 SD; $P = 0.02$ per kg/m² higher BMI). Refinement of the current GWAS of microbiome taxa abundance within the FGFP and consortium expansion will allow systematic MR analyses to improve causal inference within the relationship between the human gut microbiome and BMI.

3268T

Modeling with semi-continuous predictors: Testing and effect estimation in the presence of zero-inflated metabolites. *S.H. Chu^{1,2}, J. Lasky-Su^{1,2}, P. Kraft^{3,4}*. 1) Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA; 2) Department of Medicine, Harvard Medical School, Boston, MA; 3) Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA; 4) Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, MA.

Zero-inflated metabolite levels are very common when analyzing metabolomic data, whether due to true absence of the metabolite (such as in the case of xenobiotics), or due to measurements falling below the limit of detection. While a great deal of statistical literature has explored best practices for modeling zero-inflated outcomes, less attention has been paid to the context in which the exposure of interest is semi-continuous. We conducted simulation studies to compare power and bias of effect estimates for three methods: two typical approaches in case-control study designs, naïve logistic regression (n-LR) and logistic regression with log-transformed metabolite levels (ln-LR); and a new and simple alternative employing logistic regression with interaction terms (2df-LR). Using a factorial design, we explored the effect of altering limit of detection thresholds, mean and SD of distributions for observed metabolite levels, metabolite effect size, and population proportion of xenobiotic ingestion across 7 different distributions of metabolite sampling probabilities, with 1000 iterations each. Simulations were further conducted to investigate differences in power and bias estimates with respect to chronicity of the metabolite-disease association (i.e. acute vs chronic outcomes). All approaches demonstrated appropriate Type-I error rates across all 7 sampling probability simulations for the chronic (α_{n-LR} : 0.043-0.063, α_{ln-LR} : 0.035-0.060, α_{2df-LR} : 0.047-0.063) and acute (α_{n-LR} : 0.034-0.059, α_{ln-LR} : 0.033-0.058, α_{2df-LR} : 0.044-0.061) association settings. In general, n-LR demonstrated good power performance across a wide range of simulation settings. However, 2df-LR performed best in simulations with higher population proportions of xenobiotic ingestion in acute outcome settings. Overall, all methods were biased toward the null in terms of effect estimates, but best performance with respect to the variability and degree of bias varied by specific simulation settings.

3269F

Metabolite-GWAS of obesity in the Atherosclerosis Risk in Communities (ARIC) study. *K.L. Young¹, H.M. Highland¹, B. Yu², M. Grove-Gaona², C. Avery¹, P. Gordon-Larsen¹, E. Boerwinkle², K.E. North¹*. 1) University of North Carolina, Chapel Hill, NC; 2) University of Texas Health Science Center at Houston, Houston, TX.

Over the past 30 years, the US prevalence of obesity (measured as body mass index (BMI) >30) has more than doubled. However, not all obese individuals have the same risk for adverse health consequences. As obesity can be considered a dysregulation of energy balance, metabolites are a logical means to refine obesity phenotypes and may provide insights into mechanisms leading to disease. GWAS with metabolomics (mGWAS) can identify SNP-metabolite associations that can often be linked to metabolic transporters, regulators, or enzyme coding genes, providing an excellent opportunity to discover the best biological candidates in known GWAS regions, as well as identify novel regions associated with obesity. Our study leverages existing data in the Atherosclerosis Risk in Communities (ARIC) Study (N=4,032), a prospective, longitudinal cohort study of cardiovascular disease in African and European Americans, to identify metabolic signatures of obesity. We estimated associations between baseline natural log transformed BMI (lnBMI) and 245 baseline standardized plasma metabolites, using ancestry and sex-stratified additive linear models adjusting for age, center, smoking status, estimated glomerular filtration rate, and batch. The top 10 metabolites positively associated with lnBMI in at least one stratum ($p=2.04E-4$ (0.05/245)) were carried forward for mGWAS ($n=21$). Ancestry and sex-stratified mGWAS were conducted using additive linear models adjusted for prior covariates plus PCs in EPACTS, followed by meta-analysis in METAL. Sixty-two percent of metabolites positively associated with lnBMI in this preliminary analysis were involved in amino acid/peptide pathways, replicating previous findings implicating amino acid catabolism in adipose tissue formation. We replicated mGWAS signals for kynurenine, urate, tyrosine, and mannose, and found interesting novel signals for dihomolinolenate. An intronic variant in *MYRF1* is associated with metabolite level in all strata ($\beta=0.32$, $SE=0.04$, $p=3.20E-13$, $MAF=0.14$); this gene was previously associated with plasma fatty acid levels. A variant near *CDKAL1* is associated with dihomolinolenate levels in African Americans only ($\beta=0.65$, $SE=0.12$, $p=1.33E-8$, $MAF=0.03$); the gene was previously associated with BMI and T2D. These results highlight the potential for metabolomics to refine obesity signals in GWAS. Future studies will examine causal inference and improved obesity risk prediction using metabolomics.

3270W

A Mendelian randomisation study shows a causal role for the gut microbiome in insulin response, obesity and type 2 diabetes. N.R. van Zuydam^{1,2}, S. Sanna³, A. Mahajan^{1,2}, A. Kurilshikov³, A. Vich Vila^{2,3}, U. Võsa³, M. Oosting⁴, L.A.B. Joosten⁴, M.G. Netea⁴, L. Franke³, A. Zhernakova³, J. Fu^{5,6}, C. Wimenga³, M.I. McCarthy^{1,2,7}, *Life Lines Deep*. 1) Wellcome Centre Human Genetics, University of Oxford, Oxford, UK; 2) Oxford Centre for Diabetes Endocrinology and Metabolism, Churchill Hospital, University of Oxford, Oxford, UK; 3) University of Groningen, University Medical Center Groningen, Department of Genetics, 9700 RB Groningen, the Netherlands; 4) Department of Internal Medicine and Radboud Center for Infectious Diseases, Radboud University Medical Center, 6525 HP Nijmegen, the Netherlands; 5) University of Groningen, University Medical Center Groningen; 6) Department of Pediatrics, Groningen, 9700 RB Groningen, the Netherlands; 7) Oxford NIHR Biomedical Research Centre, Oxford University Hospitals Trust, Churchill Hospital, Oxford, UK.

Microbiome-wide association studies on large population cohorts have highlighted hundreds of correlations between gut microbiome features, obesity and type 2 diabetes (T2D). However, the causal relationships between microbiome and metabolic traits remain unresolved. We conducted a bidirectional Mendelian Randomisation (MR) study to identify causal relationships between gut microbiome features and metabolic traits. We leveraged information from *Life Lines Deep* (LLD), a cohort of 952 normo-glycaemic individuals for whom genetics, gut metagenomics and faecal short chain fatty acid (SCFA) levels, to construct genetic risk scores (GRS) for microbiome features that were correlated with metabolic traits in LLD. The GRS for these microbiome features were tested for association with metabolic traits in published summary statistics from large genome-wide association studies in an MR analysis. We found that a GRS for increased abundance of gut microbiota involved in SCFA butyrate anabolism (MetaCyc acronym: PWY-5022) was associated with increased insulin response after an oral glucose test ($P=9.8 \times 10^{-5}$). Studies of butyrate supplementation, an SCFA solely produced by the gut microbiome, in humans have shown improvement in insulin response. A GRS of increased faecal SCFA propionate levels was causally related to increased body mass index (BMI, $p=0.007$) and risk of T2D ($p=0.004$). The causal relationships between increased faecal propionate increased and risk of T2D were replicated in individual level data from an independent sample in the UK Biobank. We also conducted a PheWAS analysis in the UK Biobank for the two significant microbiome feature GRS. We found that the GRS for PWY-5022 was nominally associated with increased waist-hip ratio (WHR) ($p=0.01$) and risk of obesity ($p=0.02$). The GRS for increased faecal propionate was associated with increased body fat % ($p=2 \times 10^{-4}$), increased WHR ($p=4.3 \times 10^{-3}$) and lower visceral adipose tissue ($p=0.01$). This study demonstrates how MR can be used to study causal relationships between the gut microbiome and disease traits. The data from this study demonstrate a causal role for gut microbiome produced SCFAs butyrate and propionate in energy metabolism and glucose homeostasis in man.

3271T

Estimating carrier frequency of newborn screening disorders by using the 3.5KJPN whole-genome reference panel. Y. Yamaguchi-Kabata¹, J. Yasuda¹, S. Koshiba¹, A. Uruno¹, Y. Suzuki^{1,2}, N. Fuse¹, F. Katsuoka¹, H. Kawame¹, K. Kojima¹, M. Nagasaki¹, T. Mimori¹, S. Tadaka¹, G. Tamiya¹, O. Tanabe¹, N. Yaegashi¹, S. Kure¹, K. Kinoshita¹, M. Yamamoto¹, *ToMMo Study Group*. 1) Tohoku Medical Megabank Organization, Tohoku University, Sendai, Miyagi, Japan; 2) Ageo Central General Hospital, Ageo, Japan.

Incidence rates of Mendelian diseases vary among ethnic groups, and variant types and their frequencies of causative genes also vary among human populations. To estimate frequency of potential patients in a population, it is important to clarify frequency of causative variants of genetic diseases in a general population. In this study, we 1) examine to what extent we can predict population frequency of recessive disorders from genomic data, and 2) explore better strategies of variant interpretation and classification. We used the 3.5KJPN whole genome reference panel from a general population of Japanese, which was constructed by the Tohoku Medical Megabank Organization (ToMMo). By focusing on 32 genes for 17 congenital metabolic disorders for newborn screening in Japan, we identified reported and predicted pathogenic variants through variant annotation, interpretation and multiple ways of classifications. Then the estimated carrier frequencies were compared with those from the newborn screening data in Japan based on 1,949,987 individuals. As an example of the *PAH* gene for hyperphenylalaninaemia (HPA) and phenylketonuria (PKU), the incidence rate of HPA/PKU in Japan was known to be about 1/70,000. The carrier frequency of the *PAH* variants by using the ACMG guideline of variant interpretation (2015) showed the closest estimate to that by the incidence than the other methods did. On the other hand, using only reported pathogenic variants in ClinVar resulted in a lower estimate, and including all of HGMD disease-causing (DM) variants resulted in a much higher estimate. However, there is a common missense variant, *PAH*:p.Arg53His, which has been previously reported among the patients in East Asia as heterozygous with another *PAH* variant. This type of common variant with mild effect is often discarded by conventional filtering for pathogenic variants. Our results showed that estimation of carrier frequency based on genomic data with recent guidelines of variant interpretation gave us a closer estimate to that by the observed incidence than the other methods did. However, taking other evidences and knowledge into consideration, we should be aware of that 1) the number of pathogenic variants may increase by including additional evidences, 2) common variants with mild effects also contribute to the actual frequency of the patients, 3) penetrance of each variant is not clear yet, and 4) there may exist causative indels, not detected by whole-genome sequencing.

3272F

Clinical intervention study: Assessing the effect of genomic data-guided individualized lifestyle coaching platform for body fat loss. H. Kim, J. Kang, H. Oh, W. Park, J. Kim. Samsung Genome Institute, Samsung Medical Center, Gangnam-gu, Seoul, South Korea.

The burden of obesity is increasing worldwide due to high-calorie diets and low activity level. Given the high failure rate of uniformed weight loss methods, expectations for the individual genome-guided approach are growing. However, the interaction between environmental and genetic factors on obesity is not clear. Here, we aimed to identify interactions of genetic and environmental effect on obesity and assess the effect of genome-guided individualized life-style coaching on body fat loss using genetic risk score, GRS and lifelog data including dietary habit and physical activity. Two cohorts of a total of 274 obese and non-obese adults were studied during 15 weeks in an interventional study, which include three weeks observational period and following 12 weeks interventional period. We collected genotype data before the observational period, self-reported questionnaires, anthropometric measures, blood chemistry tests and blood pressure data before and after study periods, and lifelog data using wearable device throughout both periods. The two independent cohorts were studied from June to Sep. 2017 and from Oct. 2017 to Feb 2018. To diversity lifestyle changes, the participants were engaged in different activity groups, such as control ($n = 9$), high intensity exercise ($n = 89$), moderate intensity exercise ($n = 101$), low carbohydrate diet ($n = 39$) and low fat diet ($n = 36$) groups, and were advised to change their dietary habits and amount of physical activities. We assessed GRS for previously reported models for four categories (carbohydrate intake, fat intake, total calorie intake, and exercise). We divided the participants into two groups as high and low GRS for each of these four categories. During the study periods, BMI, body fat mass, weight were significantly reduced in the participants enrolled from June (all p values < 0.001) but there were no significant reduction in the participants from Oct. ($p = 0.85$, $p = 0.91$, and $p = 0.80$ respectively). Comparing the high GRS group with the low GRS group in those who have reduced the amount of carbohydrates intake, there was a greater loss of body fat percentage in people with high GRS ($p = 0.02$). However, there were no significant differences in the other three categories (p for fat intake = 0.51, p for total calories intake = 0.77, p for exercise = 0.99). The results of this study show that high GRS for carbohydrates increases the effect of low carbohydrates diet.

3273W

FTO in the Polynesians: Association of FTO-rs9939609 with body mass index in a fixed-effect meta-analysis in people of Polynesian ancestry living in Aotearoa/New Zealand and other Pacific nations. M. Krishnan¹, A. Phipps-Green², S. Altaf³, T.J. Major⁴, M. Cadzow⁵, L. Stamp³, N. Dalbeth¹, E.M. Russell⁶, R.L. Minster¹, N.L. Hawley⁶, T. Naseri⁶, M.S. Reupena⁷, R. Deka⁸, D.E. Weeks^{4,9}, S.T. McGarvey^{10,11}, T.R. Merriman², R. Murphy¹. 1) Department of Medicine, University of Auckland, Auckland, Auckland, New Zealand; 2) Department of Biochemistry, University of Otago, Dunedin, New Zealand; 3) Department of Medicine, University of Otago, Christchurch, New Zealand; 4) Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pennsylvania, USA; 5) Department of Epidemiology (Chronic Disease), School of Public Health, Yale University, New Haven, Connecticut, USA; 6) Ministry of Health, Government of Sāmoa, Apia, Sāmoa; 7) Bureau of Statistics, Government of Sāmoa, Apia, Sāmoa; 8) Department of Environmental Health, College of Medicine, University of Cincinnati, Cincinnati, Ohio, USA; 9) Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, USA; 10) International Health Institute, Department of Epidemiology, School of Public Health, Brown University, Providence, Rhode Island, USA; 11) Department of Anthropology, Brown University, Providence, Rhode Island, USA.

The fat mass and obesity associated (FTO) locus consistently associates with obesity and increased body mass index (BMI) in most ancestral groups with rs9939609 being the most commonly described variant associated with anthropometric measures of adiposity. Yet, previous small studies have failed to replicate the association of FTO with obesity among people of Polynesian ancestry. In this study, we tested for association of rs9939609 with BMI in (1) 1,239 people of East Polynesian ancestry (indigenous NZ Māori, Cook Island Māori and Tahitian) living in Aotearoa/New Zealand and (2) 711 people of West Polynesian ancestry (Sāmoan, Tongan, Niuēan, Pukapukan, Tokelauan and Tuvaluan) living in Aotearoa/New Zealand, and in (3) 3,072 indigenous Sāmoans living in the Independent State of Sāmoa. Published statistics of three small sample sets of Polynesian (Tongan, $n = 116$), Melanesian (Solomon Islander $n = 129$) and Micronesian (Kiribati living in the Solomon Islands $n = 75$) peoples were meta-analysed with the three study groups for association of rs9939609 with BMI. The minor allele frequencies (MAF) for the East Polynesians, West Polynesians and Sāmoans were 0.25, 0.19 and 0.20 respectively. Significant evidence of association was not found ($P \geq 0.05$) between rs9939609 and BMI in the East ($\beta = 0.55 \text{ kg/m}^2$), West ($\beta = 0.51 \text{ kg/m}^2$), or Combined East and West Polynesian ancestry groups ($\beta = 0.50 \text{ kg/m}^2$) or in the Sāmoan group ($\beta = 0.35 \text{ kg/m}^2$). Meta-analysis of these groups along with the small Tongan group, did show association of the rs9939609 minor (A) allele with higher BMI in Polynesian peoples ($\beta = 0.38 \text{ kg/m}^2$, $P = 0.023$, $n = 5,404$) but not with the wider Oceanic ($\beta = 0.28 \text{ kg/m}^2$, $P = 0.064$, $n = 5,608$) samples using a Bonferroni adjusted P -value of 0.025. These data demonstrate that rs9939609 is associated with BMI in Polynesian peoples. Previous studies investigating the association of FTO with BMI among people of Polynesian ancestry may have been limited by low statistical power.

3274T

Multi-omics of BMI in Hispanics by sparse canonical correlation analysis: IRAS family study. K.D. Taylor¹, N.D. Palmer², X. Guo¹, D.W. Bowden², Y.-I.I. Chen¹, M.O. Goodarzi³, K.L.S. Lam¹, C.D. Langefeld², L.E. Wagenknecht², J.I. Rotter¹, *Insulin Resistance & Atherosclerosis Family Study*. 1) Inst. Translational Genomics, Los Angeles Biomedical Institute, Torrance, CA., USA; 2) Center for Public Health Genomics, Wake Forest School of Medicine, Winston-Salem, NC, USA; 3) Division of Endocrinology, Cedars-Sinai Medical Center, Los Angeles, CA, USA.

Introduction. Penalized matrix decomposition methods such as sparse canonical correlation analysis (SCCA) are promising for multi-omics studies that require methods appropriate for high dimensional data, for integration of different data types, and for feature-selection, or finding correlations across a multi-omics landscape. **Aim.** Perform SCCA of BMI in Hispanics with metabolomic/genomic data in the Insulin Resistance & Atherosclerosis Family Study (IRASFS). **Methods.** SCCA was performed in three steps using R packages bestNormalize and PMA: (a) *Datasets.* BMI was adjusted for age and sex. 865 metabolomic features by Metabolon were transformed and features with a Shapiro-Wilks test >0.9 were retained. GWAS SNPs (Illumina) passing stringent QC and minor allele frequency > 0.01 were retained. The final dataset consisted of 904 subjects, 644 metabolomic features, and 1,074,639 SNPs. (b) *Penalty Selection* was performed using permutation; at the penalty set used, the correlation (0.63) was greater than the highest correlation obtained with 1000 permutations (0.38). (c) SCCA was applied using the penalty set and 4 “result clusters” (RC) extracted in succession; an RC consists of a “short list” of metabolites and SNPs which together correlate with BMI. **Results.** The 4 RCs were combinations of 24 metabolites and 169 SNPs, with 26 SNPs near loci expressed in adipose tissue (GTEx), 5 SNPs near genes for mitochondrial function (MitoCarta), and 16 SNPs associated with obesity-related traits (GWAS Catalog). RC correlations across metabolites, SNPs, and BMI were 0.62, 0.58, 0.42, 0.48, respectively. RC1 correlates increased BMI, increased branched-chain and urea-cycle amino acids, and obesity loci *LRP1B*, *KIAA1217* & *HAS2* together; RC2 & RC3 correlate BMI, lysolipid increases (RC2) and decreases (RC3), and obesity loci *ATXN1*, *CRB1*, *MYT1L*, *NRXN2*, *SLC52A3*; RC4 correlates phospholipids with obesity loci *BCAT2*, *Chr1orf94*, *GBE1*, *RORA*, *TRIP11*, and *ZMIZ1*. **Discussion.** The association of branched-chain amino acids and adiposity is well-known and many known loci for the regulation of obesity, metabolism, and insulin resistance were correlated with a modest number of metabolites. In addition to confirmation of method, these correlations provide clues to metabolic consequences of SNPs leading to BMI. We conclude that SCCA is a promising tool for integration of multi-omics datasets into a single analysis and for finding major features across a large data landscape. .

3275F

Polymorphism rs822396 (-3971 A/G) of *ADIPOQ* gene as a risk factor to develop metabolic syndrome in Mexican population: Preliminary results. L. Rubio Chavez¹, R. Rosales Gómez², J. Ramos Núñez¹, D. Rubio Chavez¹, T. Garcia Cobian¹, J. Sánchez Corona³, S. Gutiérrez Rubio¹. 1) Universidad de Guadalajara, Guadalajara, Jalisco, Mexico; 2) Universidad de Guadalajara, Tonalá, Jalisco, Mexico; 3) Centro de Investigación Biomédica del Occidente, Guadalajara, México.

Introduction: The adiponectin (ApN) has a role in metabolic process and immunological effects. ApN is encoded by *ADIPOQ* gene located in 3q27. The polymorphism rs822396 in this gene has been associated with components of metabolic syndrome (MS) in Caucasian and Asiatic populations. High serum ApN levels have a protector effect to develop chronic and metabolic diseases. **Aim:** to evaluate the association of rs822396 polymorphism and serum ApN levels with metabolic syndrome development in Mexican population.

Materia and Methods: The clinical, biochemical and anthropometric data of 74 MS cases, 35 controls and 115 individuals of general population (GP) were achieved. The polymorphism detection was made by PCR-RFLP from DNA. The serum ApN levels were determinates by ELISA. **Results:** The 33% of recruited individuals have MS. The frequency of the GG genotype is 8.12%, 2.86% and 2.6% and the G allele is 30.4%, 17.1% and 21.7% in the Cases, Controls and GP respectively. The association analysis showed that dominant model (GA/GG) in carriers with MS has an OR of 2.431 (p= 0.037, CI: 1.04-5.67) regard to controls. Another hand, allelic analysis indicated that G allele carriers with MS and an OR 2.112 (p= 0.037, CI: 1.035-4.310) regard to controls. The serum ApN levels were higher (p> 0.001) in control group (9.49 ±3.8 µg/mL) regard to MS (6.03 ±2.2 µg/mL). The GP serum ApN levels (6.875 ±2.6 µg/mL) were no different compared to cases **Conclusions:** There was found an association of twice fold risk of G allele of rs822396 polymorphisms carries to develop MS. A lower serum ApN levels was found in cases with MS related with develop MS.

3276W

Transethnic meta-analysis of adiposity traits: Use of large population biobanks reveals multiple ancestry-specific association signals. Z. Fairhurst-Hunter¹, K. Lin¹, W. Gan¹, M. Arnold², A. Mahajan¹, I.Y. Millwood¹, Z. Chen¹, M.I. McCarthy¹, M.V. Holmes¹, R. Walters¹. 1) University of Oxford, Oxford, United Kingdom; 2) University of Cambridge, Cambridge, United Kingdom.

The prevalence of obesity and overweight is increasing worldwide. However, for a given body mass index (BMI), adipose tissue distribution varies considerably between ethnic groups, resulting in different relationships between measures of adiposity and disease. Association signals from previous adiposity GWAS, which predominately used European (EUR) cohorts, account for less variation in East Asians (EAS) than is explained in EUR populations. This limits the utility of EUR identified SNPs for investigating the causal influence of adiposity on disease risk in EAS. To identify novel loci useful for translational genetic studies in EAS, and to explore differences in the genetic architecture of adiposity measures between ancestries, we used data from 96K EAS individuals from China Kadoorie Biobank (CKB) and 430K EUR individuals from UK Biobank (UKB). We performed ancestry-specific GWAS for 8 adiposity measures, including BMI, waist-hip ratio (WHR) and body fat percentage. Ancestry-specific GWAS were then meta-analysed using the Han-Eskin random effects model. GWAS in UKB yielded 100s of independent loci ($p < 5 \times 10^{-8}$) for all traits assessed. Trans-ethnic meta-analysis of the GWAS data identified on average a further 15% of additional loci. While the use of two ethnically diverse biobanks increased the number of loci identified for each adiposity trait, on average 14% of lead variants displayed significant effect size heterogeneity (5% FDR) between the two ethnicities. Differences in linkage disequilibrium account for some of these instances of trans-ethnic heterogeneity, but often the divergence appears to be due to an absence of effect in one or other ethnicity. This was most evident for WHR, with 83% of heterogeneous loci showing evidence of association in only one ancestry. In addition, genetic risk scores (GRS) constructed using GWAS results from one ancestry performed poorly when applied to the other, highlighting the consistent differences in effect size between the two ethnicities. In the extreme, a GRS for WHR constructed using CKB lead SNPs, with an overall effect size in CKB of 1.00 SD [95% CI: 0.92-1.08] per weighted allele, had a quarter of the effect when applied to UKB (0.24 SD [0.21-0.28]). These findings may relate to previously characterised differences in body-fat distribution between EAS and EUR populations and may provide insights into underlying differences in genetic architecture.

3277T

Addressing the missing data issue in multi-phenotype genome-wide association studies. M.D. Anasanti¹, M. Kaakinen^{1,2}, H. Draisma¹, Z. Balkhiyarova¹, M. Jarvelin^{3,4,5,6,7}, I. Prokopenko¹. 1) Section of Genomics and Common Disease, Department of Medicine, Imperial College London, United Kingdom; 2) Division of Experimental Medicine and Toxicology, Department of Medicine, Imperial College London, United Kingdom; 3) Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, United Kingdom; 4) Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, London, UK; 5) Center for Life Course Health Research, University of Oulu, Oulu, Finland; 6) Unit of Primary Care, Oulu University Hospital, Oulu, Finland; 7) Biocenter Oulu, University of Oulu, Oulu, Finland.

Joint analysis of multiple phenotypes in genome-wide association studies (MP-GWAS) increases power for locus discovery but suffers from missingness in phenotype values, especially with high-dimensional data. This leads to inefficiency of the standardly implemented complete case (CC) analysis and loss of power. We investigated properties of missing data imputation methods within MP-GWAS and implemented a user-friendly extension to the multi-phenotype analysis software tool SCOPA. We focussed on single and multiple-imputation (SI/MI) using Bayesian approach and expectation-maximisation bootstrapping (EMB), k-nearest neighbour (kNN), left-censored imputation method (QRILC) and random forest (RF). We simulated genetic data for 5,000/50,000/500,000 individuals using Hapgen2, and highly ($r=0.64$) and moderately correlated ($r=0.33$) phenotypes (three/nine/thirty/120) for these individuals in R. We randomly selected common, low-frequency and rare variants to be significantly ($P < 5 \times 10^{-8}$) associated with the simulated phenotypes. We considered different proportions of missing data (1/5/20/50%) under three mechanisms: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). The Root Mean Squared Error (RMSE) was calculated by contrasting the regression coefficients from the analyses regressing each SNP on the linear combination of the phenotypes, after applying the selected missing data methods, to those resulting from the full data analysis. The evaluations of the RMSEs were done using one-way Analysis of Variance (ANOVA) with CC vs. full data RMSE as the reference level. The best performing methods were applied to real data from the Northern Finland Birth Cohorts (NFBC) 1966 and 1986 (4,955 and 2,687 individuals, respectively) for the imputation of anthropometric, fasting blood glucose/insulin measurements and 149 serum metabolite levels and MP-GWAS of 31 amino acids. RF and MI-EMB showed RMSE levels closest to zero and diverged the most from the reference under MCAR ($P_{RF}=7.18 \times 10^{-4}$, $P_{MI-EMB}=6.27 \times 10^{-4}$). RF also outperformed under MAR ($P=6.22 \times 10^{-4}$), whereas QRILC outperformed under MNAR ($P=7.45 \times 10^{-4}$). With the MP-GWAS of the NFBCs after imputation with RF and QRILC we observed a number of additional loci as compared to CC analysis. We propose new user-friendly high-performing solutions for phenotype imputation in high-dimensional omics data analyses and have implemented these into imputeSCOPA software tool.

3278F

Using short identical-by-descent (IBD) segments to map disease genes in unrelated case-control studies. W.-M. Chen¹, A. Manichaikul¹, J. Nguyen¹, Z. Zhu¹, C. Robertson¹, J.R.J. Inshaw², S. Onegut-Gumuscu¹, J.A. Todd², S.S. Rich¹, *Type 1 Diabetes Genetics Consortium*. 1) Center for Public Health Genomics, University of Virginia, Charlottesville, VA; 2) JDRF/Wellcome Trust Diabetes and Inflammation Laboratory, Wellcome Trust Centre for Human Genetics, University of Oxford.

While it is known that long stretches of genomes are shared identical-by-descent (IBD) between a pair of relatives, it is less well-known that unrelated individuals can also share noticeable IBD segments that are short, say < 1Mb. For example, in the 1000 Genomes Project, the total length of IBD segments between a random pair of individuals from east Asia or Europe is estimated to be >6% of the genome (*i.e.*, 5th-degree relatedness). Similarly, while inbreeding is characterized by existence of long chromosomal regions consisting of homozygotes only, an outbred individual that does not have any recent family history of inbreeding can also carry many Runs of Homozygosity (ROHs) that are short. The widely-spread short IBD segments and ROHs in human populations can be utilized for disease gene mapping in unrelated case-control studies. Built upon a computationally efficient tool KING which is mostly known for its fast and accurate inference of relatedness by estimating kinship coefficients and recently by inferring IBD segments, we developed two IBD-based gene mapping methods, IBD mapping and homozygosity mapping. The IBD mapping method compares average IBD sharing at a position in affected pairs with that in discordant pairs, with P values being assessed by a permutation test. The homozygosity mapping method tests association between the disease status and ROH status at any position in outbred individuals using contingency tables. We applied these two IBD-based gene mapping methods to three case-control cohorts where study participants were densely genotyped at 186 immune-related loci using the ImmunoChip array and type 1 diabetes (T1D) was of primary interest. The IBD mapping scan in a case-control study in European population identified loci (with $P < 0.0001$) at *MHC*, *IKZF4*, and *SH2B3*. The IBD mapping scan in a case-control study in African Americans identified loci (with $P < 0.0001$) at *C1orf106*, *IL10*, *MHC*, *CENPW*, *IKZF4* and *SH2B3*, in contrast to 3 genes only identified using the standard logistic regression method. The Homozygosity mapping scan in a large dataset identified loci at *PTPN22*, *CTLA4*, *MHC*, *BACH2*, *CENPW*, *IKZF4*, *DEXI*, *TYK2*, and *FUT2* with $LOD > 3$. The successful identification of known T1D loci without false positives provides proof of concept of the utility of our newly developed IBD-based gene mapping methods. As WGS becomes more available, genetic loci for diseases and complex traits can be mapped alternatively using IBD-based gene mapping methods. .

3279W

Machine learning in multi-omics data to assess longitudinal predictors of glycaemic trait levels. M. Kaakinen^{1,2}, L. Prolot¹, H. Draisma¹, M.D. Anasanti¹, Z. Balkhiyarova¹, M.-R. Jarvelin^{3,4}, I. Prokopenko¹. 1) Section of Genomics of Common Disease, Department of Medicine, Imperial College London, London, United Kingdom; 2) Centre for Pharmacology and Therapeutics, Department of Medicine, Imperial College London, London, United Kingdom; 3) Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, United Kingdom; 4) Center for Life Course Health Research, University of Oulu, Oulu, Finland.

Multi-omics data hold enormous potential for personalised medicine; however, analysis of high-dimensional data poses computational and statistical challenges. Type 2 diabetes (T2D) is a global health burden that will benefit from personalised risk prediction and tracking of disease progression. We aimed to identify longitudinal predictors of the levels of glycaemic traits relevant for T2D by applying machine learning (ML) approaches to multi-omics, including epigenetic and metabolomic data, from the Northern Finland Birth Cohort 1966 at 31 (T1) and 46 (T2) years of age. We predicted fasting glucose/insulin (FG/FI), glycated haemoglobin (HbA1c) and 2-hour glucose/insulin (2hGlu/2hIns) at T2 in 595 individuals using a total of 1,010 variables from T1 and T2: body-mass-index (BMI), waist-hip-ratio, sex; 10 blood plasma measurements; 453 NMR-based metabolites; 542 methylation probes established for BMI/FG/FI/HbA1c/T2D/2hGlu/2hIns and measured with Illumina Infinium HumanMethylation450 BeadChip (T1)/EPIC (T2) arrays. Metabolic and methylation data were used in their raw form (Mb-R, Mh-R) or transformed into scores (Mb-S, Mh-S) and we tested different input data combinations. We used six ML approaches: random forest (RF), boosted trees (BT) and support vector machine (SVM) with the kernels of linear/linear with L2 regularization/polynomial/radial basis function. Data were split with a nested cross validation. The outer and the inner cross validation had both 5-folds, corresponding to 20% testing and 80% training data. RF and BT showed consistent performance while SVMs struggled with higher-dimensional data. The predictions worked best for FG and FI, with the average R^2 values from the six ML models reaching values of up to 0.47 and 0.30, respectively, using Mb-S data. Both with Mb-R and Mb-S data, sex, T2 branched-chain and aromatic amino acids, HDL-cholesterol, VLDL, glycoprotein acetyls, glycerol, ketone bodies, and measurements of obesity already at T1 were amongst the most important predictors. Addition of methylation data, either raw or score, did not improve the predictions ($P > 0.3$ for model comparison); however, BMI-associated methylation probes T1_cg26361535 at *ZC3H3/T1_cg00634542* at *SLC11A1* were within the top 25 predictors of FI/FG variability when using Mb-S+Mh-R as input data. With ML we could narrow down hundreds of variables into a clinically relevant set of predictors and demonstrate the importance of longitudinal traits in prediction. .

3280T

Joint proteomic quantitative trait locus analysis sheds light on the genetic architecture of proteins involved in obesity. *H. Ruffieux^{1,2}, W.H.M. Saris³, A. Astrup⁴, M.E. Harper⁵, R. Dent⁶, A.C. Davison¹, J. Hager², A. Valsecchia².* 1) Chair of Statistics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Vaud, Switzerland; 2) Metabolic Health Unit, Nestlé Institute of Health Sciences SA, Lausanne, Switzerland; 3) Department of Human Biology, NUTRIM, School of Nutrition and Translational Research in Metabolism, Maastricht University Medical Centre, Maastricht, Netherlands; 4) Department of Nutrition, Exercise and Sports, Faculty of Science, University of Copenhagen, Copenhagen, Denmark; 5) Bioenergetics Laboratory, University of Ottawa, Ottawa, Ontario, Canada; 6) Weight Management Clinic, The Ottawa Hospital, Ottawa, Ontario, Canada.

Hundreds of genome-wide association studies have investigated the role of thousands of loci on obesity susceptibility, yet the mechanistic consequences of genetic variation on metabolic states remain elusive. This has prompted analysis of endophenotypes via quantitative trait locus (QTL) studies which assess how genetic variants affect intermediate molecular traits, such as gene or protein expression phenotypes. Owing to the number of tests that they entail, however, conventional single-SNP approaches often fail to detect weak effects, as may exist between a genetic variant and products from remote genes. We considered two obesity clinical cohorts ($n = 1644, 789$), each involving two proteomic QTL datasets with plasma samples quantified by mass spectrometry and SomaLogic assays. To bypass the multiplicity burden and increase statistical power, we performed joint analyses of these data using our molecular QTL method LOCUS (Ruffieux et al., *Biostatistics*, 2017). LOCUS implements fast variational inference for a Bayesian hierarchical regression model that leverages shared association patterns across all genetic variants and expression outcomes; we reassessed its statistical and computational efficacy on genetic variants from our cohorts and on simulated outcomes emulating the proteomic data. We took a two-stage approach, using the Optifast Canadian cohort for discovery and the DiOGenes cohort for replication. We analyzed the mass-spectrometry and SomaLogic pQTL data in two separate LOCUS runs; these data involved 130 and 1,100 proteins respectively, and 275,000 common tag variants capturing information from Illumina Exome arrays. LOCUS uncovered 18 (respectively 118) associations, of which 83% could be replicated at false discovery rate 5%. The hits significantly co-localized with many epigenomic marks and known eQTL effects. Our analyses identified 20 *trans* associations with significant evidence for functional roles. Taking advantage of the comprehensive clinical data available for both cohorts, we studied the links between our pQTL signals and measures of insulin sensitivity and total lipid levels. These analyses indicated that most proteins under genetic control act on low-grade inflammation, insulin resistance, and dyslipidemia. Our methods and results may help to better understand obesity and associated co-morbidities. In particular, we identified novel candidate biomarkers, whose roles merit further investigation, at both clinical and pre-clinical levels.

3281F

High dimensional mediation for causal gene screening. *Q. Zhang.* Statistics, University of Nebraska Lincoln, Lincoln, NE.

QTL and eQTL study the genetic regulation of the complex traits and the gene expression. The two have also been analyzed jointly, with the hope of identifying the potential causal genes of complex diseases. Such problems can be translated as selecting the gene set that mediates the genetic effect. Conventional mediation analysis typically focuses on one mediator and one predictor at a time. Even though there have been a few mediation models for multi-variate predictors and/or mediators, statistical methods for multivariate mediation with both high dimensional predictors and mediators are still rare. In this project, we developed a high dimensional mediation framework based on penalized linear mixed model for causal gene screening. We applied the proposed method to study the mechanism of insulin secretion using an f2 cross mouse data, and found that a dozen genes could block a large proportion of the genotype-phenotype causal flow, including many interesting genes with known association to pancreas function and Type 2 Diabetes.

3282W

Association of Glucose-6-phosphate dehydrogenase (G6PD) variants with non-glycemic lowering of Hemoglobin A1c (HbA1c) in East Asians.

V.J.Y. Lim¹, A. Leong^{2,3}, C. Wang^{4,5}, R. Dorajoor⁶, J.F. Chai⁷, C.K. Heng^{8,9}, R.M. van Dam^{1,8,9}, J.B. Jonas¹⁰, Y.X. Wang¹¹, J. Liu^{4,6}, D.F. Reilly¹², T.Y. Wong^{13,14}, C.Y. Cheng^{13,14,15}, X. Sim¹. 1) Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore, Singapore; 2) Massachusetts General hospital, Boston, MA, USA; 3) Harvard Medical School, Boston, MA, USA; 4) Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore; 5) Duke-NUS Medical School; 6) Department of Paediatrics, Yong Loo Lin School of Medicine, National University of Singapore; 7) Khoo Teck Puat - National University Children's Medical Institute, National University Health System, Singapore; 8) Yong Loo Lin School of Medicine, National University of Singapore, Singapore; 9) Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA, USA; 10) Department of Ophthalmology, Medical Faculty Mannheim, University Heidelberg, Germany; 11) Beijing Institute of Ophthalmology, Beijing Key Laboratory of Ophthalmology and Visual Sciences, Beijing Tongren Eye Center, Beijing Tongren Hospital, Capital Medical University; Beijing, China; 12) Genetics, Merck Sharp & Dohme Corp., Kenilworth, New Jersey, USA; 13) Ophthalmology & Visual Sciences Academic Clinical Program (Eye ACP), Duke-NUS Medical School, Singapore; 14) Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore; 15) Singapore Eye Research Institute, Singapore National Eye Centre, Singapore.

Glucose-6-phosphate dehydrogenase (G6PD) deficiency is an X-linked disorder caused by mutations in the *G6PD* gene. Prevalences of G6PD deficiency are 7.1% globally and 4.7% in Asia. A common African-American specific *G6PD* variant has been reported to be associated with non-glycemic lowering of HbA1c, a commonly used diagnostic test for diabetes (rs1050828, MAF=11%, $P=8.23 \times 10^{-135}$). HbA1c in male carriers of the *G6PD* variant was 0.81% lower than non-carriers. To identify Asian-specific *G6PD* variants associated with HbA1c, we performed imputation with the 1000 Genomes reference panel in seven Asian studies and tested the variants for association with HbA1c. We performed sex-stratified association analyses before combining by meta-analysis [9,773 East Asians (4,850 men, 4,923 women) and 1,486 South Asians (732 men, 754 women)]. No significant association was observed in South Asians. In East Asian men, 35 variants were associated with HbA1c at the *G6PD* locus ($P < 5 \times 10^{-6}$). The strongest association was a low-frequency missense mutation known as *G6PD*-Canton (rs72554665, Arg489Leu, MAF=2.2%, $P=1.25 \times 10^{-27}$, $n=2,844$, variance explained in men=6.9%). HbA1c in male carriers was 0.76% lower than non-carriers. Conditioning on this variant, 10 variants at the locus remained significant. These associations were driven by another low-frequency missense variant, *G6PD*-Kaiping (rs72554664, Arg493His, MAF=1.1%, $P=9.59 \times 10^{-11}$, $n=612$, effect size =-1.11%, $r^2=0.0$ with *G6PD*-Canton in 1000G EAS). Further conditioning on both variants, the conditional P -values at the locus was >0.1 . Imputed genotypes had good concordance with exome sequencing data (*G6PD*-Kaiping: 98.6% and *G6PD*-Canton: 100%). To determine whether genetic effects on HbA1c were mediated by glycemia, we tested whether *G6PD*-Canton and *G6PD*-Kaiping were associated with glucose measures. Both variants were not associated with fasting glucose (rs72554664, $P=0.915$, $n=548$; rs72554665, $P=0.533$, $n=1134$) or random glucose (rs72554665, $P=0.485$, $n=791$), and conditioning on these glucose measures did not attenuate their effect on HbA1c. In summary, we identified 2 *G6PD* variants in East Asian men associated with non-glycemic lowering of HbA1c. Diabetes diagnosis may be missed in carriers of these *G6PD* variants when screening Asian populations for diabetes with HbA1c, highlighting the importance of investigations in diverse ancestries to uncover ancestry-specific variants that influence the accuracy of clinically useful diagnostic tests.

3283T

Causal effects of blood lipids on breast cancer risk: A Mendelian randomization study. C. Nowak¹, J. Ärnlöv^{1,2}.

1) Division of Family Medicine and Primary Care, Department of Neurobiology, Care Sciences and Society (NVS), Karolinska Institutet, Huddinge, Sweden; 2) Dalarna University, Falun, Sweden.

Observational studies have reported associations between circulating lipids and risk of breast cancer, but whether low-density lipoprotein-cholesterol (LDL-cholesterol), high-density lipoprotein-cholesterol (HDL-cholesterol) and triglycerides causally affect breast cancer risk is unclear. We used summary genetic datasets for breast cancer and blood lipids in two-sample Mendelian randomization to assess whether lipids affect breast cancer risk. Mendelian randomization uses exposure-associated genetic variants as instruments to avoid bias from reverse causation and confounding. We constructed instruments of genome-wide significant variants for lipids and used inverse variance-weighted, Egger, weighted median and multivariable Mendelian randomization. To approximate effects of lipid-modifying drugs, we instrumentalised lipid-associated variants in the genes encoding the targets of statins, PCSK9 inhibitors, CETP inhibitors and ezetimibe. Genetically raised HDL-cholesterol was associated with higher risk of breast cancer (odds ratio, OR, per standard deviation, 1.10, 95% confidence interval, CI, 1.04, 1.16, $p = 0.001$) and ER-positive breast cancer (OR 1.10, 95% CI 1.03, 1.17, $p = 0.003$). There were no clear effects of LDL-cholesterol or triglycerides. We detected evidence that HDL-cholesterol-raising variants in *CETP* could be associated with higher risk of breast cancer (OR 1.04, 95% CI 1.01, 1.07, $p = 0.002$) and ER-positive breast cancer (OR 1.03, 95% CI 1.00, 1.07, $p = 0.025$), and that LDL-cholesterol-lowering variants in genes encoding the targets of statins, PCSK9 inhibitors and ezetimibe may have protective associations. We found evidence that genetically raised HDL-cholesterol levels may be associated with a small increase in the risk of breast cancer in women. Genetic effects may differ from pharmacologic interventions and bias from unaccounted-for pleiotropy cannot be excluded.

3284F

Genomewide association for high-density lipoprotein change levels over time in three large combined studies.

M.F. Feitosa¹, K.L. Lunetta^{2,3}, L. Wang⁴, M.K. Wojczynski⁵, C.M. Kammerer⁶, T. Perls⁷, N. Schupf⁸, K. Christensen⁹, J.M. Murabito³, M.A. Province¹. 1) Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, St Louis, MO; 2) Department of Biostatistics, Boston University School of Public Health, Boston, MA; 3) Framingham Heart Study, MA; 4) Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, PA; 5) Department of Medicine, Geriatrics Section, Boston Medical Center, Boston University School of Medicine, Boston, MA; 6) Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Department of Neurology, Columbia University Medical Center, New York, NY; 7) Department of Public Health, The Danish Aging Research Center, University of Southern Denmark, Odense, Denmark.

Elevated HDL-C levels contribute to atheroprotection, and HDL-C levels are higher among healthy elderly and exceptional aging than in younger adults. However, extreme HDL-C levels are associated with increased all-cause mortality risk. Although several genetic factors are known to contribute to levels and metabolism of HDL-C, their protective effects in cardiovascular disease (CVD) and healthy aging-longevity are complex and poorly understood. We hypothesized that a genome-wide association study (GWAS) of HDL-C longitudinal change, accounting for harmonization of HDL-C and imputed genotype in 3 family-based studies recruited for exceptional survival (Long Life Family Study, LLFS), CVD risk factors (Family Heart Study, FamHS) and community based sample (Framingham Heart Study, FHS), might identify additional loci for HDL-C. We examined fasting HDL-C change in 7738 subjects with 2-visits for LLFS (N=2002, interval=5.5–11.2 years) and FamHS (N=1895, interval=6.3–9.6 years), and 2-to-28 visits for FHS (N=3841, interval=1–36 years). All HDL measures (N=25146) from the FHS were adjusted for age, sex, and standardized (0,1). Age and sex regression coefficients from FHS were applied to adjust HDL by age-sex in LLFS (N=4004) and FamHS (N=2790). We performed imputation on 1K-Genome in each study. For genotyping harmonization, we included imputed SNPs present in all 3 studies with imputation quality ≥ 0.3 and allele frequency differences < 0.2 from 1K-Genome. We employed growth curves using linear mixed models to estimate random effects of slope of interval for each person to derive the best linear HDL-C change. We performed GWAS using a linear regression model with additive variant effects on HDL-C change accounting for family relationships. To adjust for genotype batches, changes in assays, across years and data, we corrected HDL-C for study (LLFS, FHS, FamHS) and population structure using principal components. We identified a novel association for HDL-C change with *GRID1* ($1.43E-9$), which encodes a subunit of glutamate receptor channels involved in synaptic plasticity. Suggestive novel associations ($< 1.0E-5$) for HDL-C change were with *MBOAT2* (role in lipid-metabolism pathway), *CYSLTR2* (role in cardiovascular systems), *NTNG2* (role in neurite outgrowth) and *LINC01314*. Modeling HDL-C levels over time in prospective studies, with difference in healthy-longevity and CVD risk, contributed to gene discovery and provides novel insights into mechanisms of HDL-C regulation. .

3285W

Genome-wide association study of FADS1 ω -6 activity in a homogeneous island population.

M. Daya¹, T.M. Brunetti¹, N. Rafaels¹, M. Boorgula¹, A. Shetty², S. Chavan¹, M. Campbell¹, S. Sergeant², C.R. Gignoux³, T.H. Beaty², F.H. Chilton², K.C. Barnes¹, R.A. Mathias³. 1) Department of Medicine, University of Colorado, Aurora, CO; 2) Wake Forest University Health Sciences, Wake Forest Center for Botanical Lipids, Winston-Salem, NC; 3) Department of Epidemiology, Bloomberg School of Public Health, JHU, Baltimore, MD, USA.

RATIONALE: Tangier Island, Virginia, is a small island located in Chesapeake Bay. The majority of its inhabitants trace their ancestry to three original founders who settled the island in 1722. The population has a homogeneous island-wide diet high in saturated and ω -6 fatty acids, due in part to limited access to fresh produce on the island. The unique homogeneous genetic and dietary environment of this population is well suited to detecting and interpreting the effects of genetic variants in the formation of long-chain polyunsaturated fatty acids (PUFAs), which play an important role in immunity and inflammation. In a previous candidate gene association study, we identified a cluster of SNPs in the *FADS1* gene with minor alleles that are strongly associated with a decrease in ratio of ω -6 PUFAs reflecting *FADS1* activity ($P < 1E-12$). Subsequently we performed genome-wide genotyping, and now have the opportunity to test for additional associations at the genome-wide level. **METHODS:** Samples were genotyped on Illumina's MEGA and genotype imputation was performed on the Michigan Imputation Server (using the HRC reference panel). After quality control, 220 samples with fasting serum PUFA concentrations were available for analysis. Linear mixed effect models implemented in the R Bioconductor package GENESIS were used to test for association between imputed allelic dose and the ratio of ω -6 PUFAs reflecting *FADS1* activity. Age at serum collection, sex and PC1 were included as fixed effect covariates in the statistical model, and a kinship matrix estimated from genome-wide data was included as random effect to account for relatedness between individuals. **RESULTS:** Even with the high degree of relatedness across this study, test statistic inflation was not observed ($\lambda = 0.99$ for $MAF \geq 5\%$, $\lambda = 0.96$ for $MAF < 5\%$). Two loci reached genome-wide significance ($P < 5E-8$): the *FADS1* gene on chr11q12-q13.1 (most significant $P = 1.9E-13$), as well as a region on chr1p36 including 13 SNPs (all in complete LD, $MAF \geq 7\%$, $Rsq = 0.99$, $P = 1.2E-8$). These SNPs in chr1p36 are intronic to the *ALPL* gene, which plays a role in vitamin B6 catabolism, and the significant minor alleles were associated with an increase in the *FADS1* activity ratio. **CONCLUSION:** We identified genetic variants that may increase the formation of ω -6 PUFAs through the vitamin B6 pathway. This is biochemically plausible as vitamin B6 pathway components are cofactors in fatty acid metabolism.

3286T

A powerful and adaptive gene based method to identify multiple traits associated genes using GWAS summary data. X. Guo¹, J. Zhang², Z. Zhao³, B. Guo⁴, B. Wu⁴. 1) Department of Computer Science and Engineering, UNIVERSITY OF NORTH TEXAS, DENTON, TX; 2) Department of Mathematics, University of North Texas, Denton, TX, USA; 3) Texas Academy of Mathematics & Science, University of North Texas, Denton, TX, USA; 4) School of Public Health, University of Minnesota, Minneapolis, MN, USA.

Genome-wide association studies (GWAS) have been routinely used for discovering disease susceptible genes. They have successfully identified tens of thousands of disease related variants; however, significant missing heritability remain for many traits and there remain many more variants with small effects to be discovered. Recent research has suggested that integrating multiple related traits and multiple variants can significantly boost our power to detect novel genetic variants. This has been partly motivated by the polygenic nature of most complex human traits and the widespread pleiotropy observed in many traits. For example, accumulating evidence show that pleiotropic loci affect more than one blood lipid trait and it is also the case in Alzheimer's disease, which is highly related to disrupted brain networks represented as a set of multiple traits. Despite its promise, testing for multivariate trait/locus associations is challenging. In this study, we propose a general adaptive p-value combination method (GPC) that integrate multiple correlated traits and multiple variants in a gene region or pathway to improve the power to detect disease susceptible genes in GWAS. Our method needs only publicly available GWAS summary data, avoiding the privacy and logistic issue of obtaining raw data. We conducted extensive simulation studies to check that GPC properly controls the type I errors at the genome-wide significance level. We compared the power of GPC to several popular existing tests, using real GWAS summary data for lipids and glycemic traits. GPC showed dramatically improved statistical power and identified many novel genetic variants missed by the individual trait based GWAS analysis. Many of these newly identified variants have been reported strongly related to the lipids and glycemic traits in the literature.

3287F

Variation in serum PCSK9, cardiovascular disease risk and potential unanticipated effects of PCSK9 inhibition: A GWAS and Mendelian randomization study in the HUNT study, Norway. B. Brumpton^{1,2,3}, L. Fritsche^{1,4}, J. Zheng⁵, J.B. Nielsen^{5,6}, M. Mannila^{7,8}, I. Surakka⁹, H. Rasheed¹, G.A. Vie¹, S. Graham², M. Gabrielsen¹, L.E. Laugsand⁹, P. Aukrust^{10,11,12}, L. Vatten⁹, J.K. Damås^{13,14}, J. Zwart¹⁵, F.M. van 't Hooft¹⁶, N.G. Seidah¹⁶, K. Hveem¹, C. Willer^{15,6}, G. Davey Smith², B.O. Åsvold^{1,17}, INVENT Consortium. 1) K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Trondheim, Trøndelag, Norway; 2) MRC Integrative Epidemiology Unit, University of Bristol, UK; 3) Department of Thoracic Medicine, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway; 4) Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, USA; 5) Department of Internal Medicine, University of Michigan, Ann Arbor, USA; 6) Department of Human Genetics, University of Michigan, Ann Arbor, US; 7) Cardiology Unit, Department of Medicine Solna and Center for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden; 8) Cardiovascular Medicine Unit, Department of Medicine Solna and Center for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden; 9) Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Norway; 10) Research Institute of Internal Medicine, Oslo; 11) Section of Clinical Immunology and Infectious Diseases, Oslo University Hospital Rikshospitalet, Oslo; 12) Institute of Clinical Medicine, University of Oslo, Oslo; 13) Centre of Molecular Inflammation Research, Department of Clinical and Molecular Medicine, NTNU, Norwegian University of Science and Technology; 14) Department of Infectious Diseases, St Olavs hospital, Trondheim; 15) Division of Clinical Neuroscience, Oslo University Hospital and University of Oslo; 16) Laboratory of Biochemical Neuroendocrinology, IRCM, Montreal, QC, Canada; 17) Department of Endocrinology, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway.

Proprotein convertase subtilisin/kexin type 9 (PCSK9) enzyme inhibitors reduce serum low-density lipoprotein cholesterol (LDL-C) and subsequent risk of cardiovascular disease. While some long-term trials have evaluated the safety of such inhibitors, a broad investigation of potential outcomes over the life-time, has not been conducted. We investigated genetic variants associated with serum PCSK9, and examined effects of PCSK9-lowering genetic variants on a range of potential outcomes to inform about possible unanticipated effects of long-term PCSK9 inhibition. We linked genotype and phenotype data from 69,422 participants in the population-based Nord-Trøndelag Health Study (HUNT) with hospital-recorded ICD codes. We performed a genome-wide association study of serum PCSK9 measured in 3,697 adults. In the total sample, we defined a new genetic risk score for serum PCSK9, and regressed both this and an existing score on a broad range of outcomes. For comparison, we estimated the corresponding associations for variants associated with LDL-C on *HMGCR*, the target of statins. We further performed two-sample Mendelian randomization on 48 outcomes using large biobanks. Finally, we assessed the genetic correlation between serum PCSK9 and 229 diseases or traits using LD-score regression. Three independent genetic variants were associated with serum PCSK9 at genome-wide significance: two variants in *PCSK9* (rs11591147, $P=2.9 \times 10^{-37}$; rs499883, $P=2.3 \times 10^{-16}$) and one in *PCSK2* (rs192265866, $P=2.8 \times 10^{-48}$), of which the two variants in *PCSK9* were replicated in an independent cohort. Using an existing PCSK9 genetic risk score, we demonstrated probable causal associations between lower serum PCSK9 and both lower serum LDL-C ($P=2.8 \times 10^{-43}$) and lower risk of coronary heart disease ($P=8.9 \times 10^{-64}$). Similar associations were observed using a new genetic risk score for serum PCSK9 using the two *PCSK9* variants identified in our GWAS. We did not observe any other convincing adverse or beneficial associations of the PCSK9 genetic risk scores across a range of cardiovascular, respiratory, mental, diabetes, or pain related outcomes. The lack of association between the PCSK9 genetic risk scores and a range of potential adverse outcomes is reassuring for the long-term use of PCSK9 inhibitors. However, further studies are warranted to confirm our findings, follow-up on the functional influences of the variants identified, and extend our investigations to other outcomes.

3288W

Large-scale genetic analysis identifies 17 novel loci for asthma. Y. Han^{1,2}, J.A. Hartiala^{1,2}, P. Huang^{1,2}, Q. Jia^{1,2}, N.C. Woodward^{1,2}, J. Gukasyan^{1,2}, F.D. Gilliland¹, H. Allayee^{1,2}. 1) Departments of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA; 2) Biochemistry & Molecular Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA.

To date, genome-wide association studies (GWAS), primarily in subjects of European ancestry, have identified 51 loci associated with asthma. To further understand the contribution of genetic factors to asthma, we conducted a meta-analysis of GWAS data with ~2.0 million SNPs in 62,739 asthma cases and 416,529 controls from the Trans-National Asthma Genetic Consortium (TAGC) and UK Biobank. In analyses with multi-ancestry populations, we identified 17 novel associated regions, bringing the total number of asthma susceptibility loci to 68. The risk alleles at eight of the novel loci were also associated with reduced lung function. By comparison, only two loci were associated with hay fever or eczema, indicating that the majority of the regions may be involved in adult non-allergic asthma. Bioinformatics and expression quantitative locus (eQTL) analyses identified candidate causal genes at nearly all 17 novel loci, including *CD52*, *PRDM11*, and *FADS1/2*, and provided evidence for enrichment of associated variants at all 68 asthma loci colocalizing to enhancer marks and DNase I hypersensitive sites specifically in blood and immune cells. Collectively, the results of our analyses shed additional light on the genetic architecture of asthma.

3289T

Utilizing epistasis to investigate bronchodilator response in African American youth. J. Magana¹, M. White¹, M. Contreras^{1,2,3}, O. Risse-Adams^{1,4}, P. Goddard¹, A. Zeiger^{1,5}, A. Mak¹, S. Salazar¹, J. Liberto¹, S. Oh¹, D. Hu¹, C. Eng¹, S. Huntsman¹, E. Burchard^{1,6}. 1) Department of Medicine, University of California, San Francisco, CA; 2) SF BUILD, San Francisco State University, San Francisco, CA; 3) MARC, San Francisco State University, San Francisco, CA; 4) Lowell Science Research Program, Lowell High School, San Francisco, CA; 5) Department of Biology, University of Washington, Seattle, WA; 6) Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, CA.

In the U.S. asthma is one of the leading chronic illnesses among children. Bronchodilator medications are the standard of care for asthma, yet Bronchodilator Drug Response (BDR) is a complex phenotype that is highly variable among different racial/ethnic groups. As we attempt to deepen our understanding of complex phenotypes that impact disparate health outcomes, it is important to consider interactions between environmental, psycho-social, demographic, genetic, and clinical variables that may inform BDR. We evaluated the impact of pairwise and higher-order interactions (known as epistasis) on BDR in African American youth with asthma. Our study population consisted of 688 individuals with asthma from the Study of African Americans, Asthma, Genes, & Environments (SAGE II) population with detailed measures of BDR, perceived discrimination, environmental exposures, genetic ancestry, and more. We conducted epistatic interaction analysis using ViSEN, a non-parametric entropy-based program which has been shown to be as or more powerful than standard regression-based methods in identifying interaction effects. We identified 6 significant epistatic interactions associated with BDR in the full dataset. The strongest interaction models were: Prenatal Smoke Exposure + Age (IG: 2.05%, $p=0.002$), and Sex + Experience of Discrimination + Global African Ancestry (IG: 0.81%, $p=0.015$). Having identified the interaction effects that have the greatest impact of BDR within our cohort, future studies can focus on further investigating these interaction models to characterize the nature of the relationship between these interactions and BDR. This novel methodology may contribute to the understanding of how epistasis between genetic and other health related factors impacts health disparities.

3290F

***TBC1D29* is associated with response to inhaled corticosteroids in adult asthmatics.** A.L. Wang, A.A. Dahlin, M.J. McGeachie, A.C. Wu, K.G. Tantisi-*ra*. Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA.

Up to one-third of asthmatics have poor response to inhaled corticosteroids (ICS), the primary therapy for asthma control. The influence of genetics on ICS treatment response has been described in small genomic studies. In a larger population of 3,542 adult Caucasian asthmatics on ICS therapy, we aimed to replicate and identify novel genetic loci associated with ICS treatment response. Within the 16,274 Caucasian asthmatics in the Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort, 3,542 were on inhaled corticosteroids with or without the use of long-acting bronchodilators. Subjects with respiratory co-morbidities and use of other asthma medications were excluded. In total, 1,225 (34.5%) of subjects responded to ICS treatment, as defined by the absence of asthma-related oral corticosteroid bursts, emergency department visits, and hospitalizations. Genome-wide multivariate logistic regression was performed using PLINK2. One SNP, rs216454, in the intronic region of *TBC1D29* on chromosome 17q11.2 was associated with ICS treatment response (OR 1.43, FDR-adjusted $P = 2.035 \times 10^{-4}$). Homozygotes for the major allele (G/G) had 6% less asthma exacerbations and heterozygotes had 3% less asthma exacerbations than homozygotes for the minor allele (T/T) ($P = 0.035$). In the largest ICS pharmacogenomic study, we identified a novel genetic locus associated with ICS treatment response in adult Caucasian asthmatics. *TBC1D29* has not previously been associated with ICS treatment or asthma but has been implicated in another type 2 inflammatory disease, eosinophilic esophagitis. Replication of this study is underway in 2,125 asthmatics on ICS therapy in the Partners Biobank.

3291W

Whole-blood eQTL analysis of multiple sclerosis patients from Puerto Rico. H. Bai^{1,2}, A.H. Beecham^{3,4}, K.L. Miskimen², C.P. Manrique³, W. Hulme³, J.L. McCauley^{3,4}, W.S. Bush². 1) Department of Nutrition, Case Western Reserve University, Cleveland, OH, USA; 2) Institute for Computational Biology, Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH, USA; 3) John P. Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, FL, USA; 4) Dr. John T. Macdonald Department of Human Genetics, Miller School of Medicine, University of Miami, Miami, FL, USA.

Multiple Sclerosis (MS) is an autoimmune disease that causes chronic progressive disability in patients accompanied by the formation of plaques in the central nervous system. MS has multiple established genetic factors, however, the majority of genetic studies of MS to date have been performed in European-descent populations. This is due in part to the higher prevalence of MS in European populations, specifically those at higher latitudes. However, studies of MS in Latin American countries suggest either a rising prevalence or under-diagnosis of MS in Hispanic populations, and many Latin American countries have moderate influence of European genetic admixture. In this work, we performed a small pilot eQTL analysis of six Hispanic MS cases ascertained from Puerto Rico. RNA was extracted from whole-blood and was characterized using RNA-seq. Gene expression profiles for each case were matched to available genotype data generated from the MegaEx chip and MS chip. Due to the limited sample size, we restricted this analysis to established whole-blood eQTL signals found in the GTEx project, composed of predominantly European-descent samples. We further filtered these eQTL variants to identify 1559 variants that had both available genotypes and high allele frequency within the Puerto Rican population based on the 1000 Genomes dataset. These variants (with expected minor allele frequency > 0.4) were evaluated using additive linear regression models. 731/1559 (46%) of genes with strong eQTLs from GTEx also had eQTL effects detected in our analyses of the MegaEx chip data, and 175/1559 (11.2%) were detected by analyses of the MS chip data ($p < 0.05$). Three of these variants have been previously associated to MS in European-descent samples, and in our study appear to alter expression of the GSDMB, MCM9, and PLEC genes. While this study is small in scope, we illustrate the power of eQTL analysis given the strong effects alleles have on gene expression, and the relative proportion of European-derived eQTL effects we see generalizing in the Puerto Rican population.

3292T

Developing an efficient method for identifying gene-gene interaction via random forest and least absolute shrinkage and selection operator. S. Wu¹, C. Chang¹, C. Lee¹, I. Lian², C. Fann¹. 1) Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan; 2) Department of Mathematics, National Changhua University of Education, Changhua, Taiwan.

Gene-gene interaction has been identified as a potential factor that contributes to heritability in complex traits. Multifactor Dimensionality Reduction method (MDR) is often used to test for genetic interaction¹. Another popular tool is the epistasis module of PLINK². A recent available procedure is Matrix-Epistasis³ that could adjust for covariates and detect putative epistasis with a short time. Despite the advancement, dimensionality particularly for NGS data remains a challenge. The issues encompass the problem of handling higher order interactions (> 2-way) where the number of possible marker combinations is manifold for a large number of SNPs. To circumvent, we propose a method that combine Random Forest Algorithm and Least Absolute Shrinkage and Selection Operator (LASSO) to detect gene-gene interactions. Firstly, Random Forest is used to select important variants by pre-determined criteria. Secondly, by using results from Random Forest, LASSO screens out most explanatory variants and two-way interactions. The above two procedures are repeated 1000 times; those appear most frequently are retained. Lastly, LASSO is used to select the most potential ones for a logistic regression model. The HierNet function with LASSO algorithm and randomForest in R package were implemented for our proposed method. A GAW16 dataset which consisted of 868 Rheumatoid Arthritis patients and 1194 controls was used to test feasibility. Only 84609 marginal significant SNPs were analyzed ($p < 0.1$). Aside from HLA markers, the results showed that the interaction between *GRIN2A* and *PLXDC2* were most significant ($p = 0.000529$) while the p value for *GRIN2A* (rs1448258) is 0.001301 and *PLXDC2* (rs7919778) is 0.06147 ($p = 0.06147$). By using IPA, we found that *GRIN2A*, *PLXDC2* and *PSEN1* are in a network, however, only one SNP in *PSEN1* was typed and the p -value was higher than 0.1 therefore, not included in the study. In conclusion, our proposed method provides a new way to look at gene-gene interaction that can help to uncover epistasis effects in gene mapping studies. 1. Ritchie, M.D. et al., Am J Hum Genet. 2001 2. Purcell, S. et al., Am J Hum Genet. 2007 3. Zhu, S. & Fang, G., Bioinformatics. 2018.

3293F

GWAS of six red blood cell traits identifies multiple associations with ancestry-specific lead variants: The PAGE study. C.J. Hodonsky¹, K. Nishimura², M. Graff³, R. Tao³, S. Buyske⁴, L.A. Hindorf⁵, Y. Li^{1,6}, D. Lin¹, R.J.F. Loos⁷, A.P. Reiner², K.E. North^{1,6}, C.L. Kooperberg², C.L. Avery¹. 1) University of North Carolina Gillings School of Public Health, Chapel Hill, NC; 2) Fred Hutchinson Cancer Research Center, Seattle, WA; 3) Vanderbilt University, Nashville, TN; 4) Rutgers University, Piscataway, NJ; 5) National Human Genome Research Institute, Bethesda, MD; 6) University of North Carolina Chapel Hill, Chapel Hill, NC; 7) Icahn School of Medicine at Mount Sinai, New York, NY.

Background: Red blood cell (RBC) traits—including hematocrit, hemoglobin, RBC count, mean cell hemoglobin (MCH), MCH concentration, and mean corpuscular volume—are heritable and associated with a wide array of cardiovascular- and inflammatory diseases. Recent large-scale genotype analyses suggest that genetic associations with RBC traits remain to be identified. We thus conducted a GWAS of six RBC traits in Population Architecture using Genomics and Epidemiology (PAGE II) participants genotyped on the Multiethnic Genotyping Array (MEGA). **Methods:** The Illumina MEGA platform was designed to improve common and low frequency variant discovery and fine mapping in global populations. A maximum of 29,223 PAGE participants (30% African American, 65% Hispanic/Latino, 75% female) with phenotype and MEGA genotype data were available for one or more traits. We conducted a GWAS of inverse-normal-transformed RBC traits and ~30 million genotyped and 1000 Genomes Phase 3-imputed autosomal variants in PAGE II participants. We defined a discovery association signal as a lead variant >500kb from any published lead variant. Inverse-normalized residuals were adjusted for age, sex, study, study center, self-reported race/ancestry, family structure, and ten ancestral principal components. **Results:** We identified 23 association signals that exceeded genome-wide significance ($p < 5 \times 10^{-8}$) for one or more RBC traits. Sixteen association signals had been previously reported for at least one trait; seven qualified as discoveries (proximal genes: *SLC26A*, *KHDRBS3*, *LINC00550*, *SEL1L*, *CATSPERB*, *ERCC4*, and *COX4I2*). Three discovery lead variants had a rare allele frequency (MAF < 1%). Two of these and one additional discovery lead variant were specific to African- or Amerindian 1000 Genomes populations. Replication analysis will be performed in study populations representing African, Amerindian, East Asian, and European ancestries. **Discussion:** The identification of seven discovery association signals at a stringent significance threshold emphasizes the need for ancestrally diverse study populations in genomic analyses, particularly for highly polygenic complex traits. Ancestry specificity of multiple discovery signals reinforces the benefits of a genotyping array that allows high-quality imputation outside of European-ancestry populations. Bioinformatic characterization of discovery lead variants is ongoing, including prioritization of candidate functional variants at each association signal.

3294W

Violence exposure, stress biomarkers and gender differences in buccal telomere length among African American young adults. L. Jackson^{1,2}, F. Saadatmand¹. 1) Pediatrics and Child Health, Howard University, Adelphi, DC; 2) W. Montague Cobb Research Lab, Howard University, Washington, DC.

Violence exposure has long-lasting social and biological impacts on African American (AA) health and can lead to physiological changes, including shorter telomere length (TL). While studies have investigated the relationship between TL and life stress, few focus on African American young adults and their direct nexus to violence. This study examines the effect of violence and gender on both stress biomarkers and TL in AA young adults. We examine the relationship between violence exposure, seven stress biomarkers (IgA/G/E/M, C Reactive Protein, Cortisol, Epstein Barr Virus Antigen) and TL in a cross sectional analysis of 50 buccal samples (N=50 males & 50 females) of AA 18-25 years old in Washington DC who experience differential violence exposure (physical, threat, witnessed, and sexual). Average TL was measured by qPCR. Mann-Whitney tests identified differences between males and females in exposure to violence, stress biomarkers, and the measures of TL. Correlations were calculated between TLs, biomarker levels, and violence measures. Elevated sexual violence exposure was positively correlated (RRANGE= 0.22 to 0.55) with all elevated stress biomarkers except IgE. TL in the high violence exposure group was negatively correlated to all stress biomarker levels (RRANGE= -0.09 to -0.31) except for IgG. There was no significant difference in TL among females exposed and not exposed to violence. Violence exposed females had a longer TL than men (MeanF:M=1.87 v 1.62, $p \leq .05$). Male sexual violence exposure was correlated to TL ($R=0.575$, $p \leq 0.03$). High violence levels correlate to shorter TLs and higher stress biomarker levels in AA young adults. These findings suggest that additional focus should be placed on African American young men who have experienced sexual violence.

3295T

GWAS of iron traits in Chilean children of the Santiago Longitudinal Study (SLS). G. Chittoor^{1,2}, E. Blanco³, M. Graff⁴, Y. Wang⁵, PE. Melton¹, C. Albala⁶, JL. Santos⁶, B. Angel⁶, B. Lozoff⁷, VS. Voruganti⁸, KE. North², S. Gahagan³, AE. Justice^{1,2}. 1) Biomedical and Translational Informatics Institute, Geisinger, Danville, PA, USA; 2) Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; 3) Division of Academic General Pediatrics, Child Development and Community Health at the Center for Community Health, University of California at San Diego, San Diego, CA, USA; 4) Curtin/UWA Center for Genetics Origins of Health and Disease, School of Pharmacy and Biomedical Science, Faculty of Health Science, Curtin University, Perth, Australia; 5) Department of Public Health Nutrition, Institute of Nutrition and Food Technology, University of Chile, Santiago, Chile; 6) Department of Nutrition, Diabetes and Metabolism, School of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile; 7) Pediatrics and Communicable Diseases, University of Michigan, Ann Arbor, MI, USA; 8) Department of Nutrition and UNC Nutrition Research Institute, University of North Carolina at Chapel Hill, Kannapolis, NC, USA.

Background: Genetic predisposition may contribute to iron deficiency and its negative health outcomes. Despite its high world-wide prevalence, investigations of genetic contribution to iron phenotypes are limited in pediatric and ethnically diverse populations, especially those of Hispanic/Latino descent. Estimating the genetic effects of iron biomarkers in an ancestrally diverse and admixed Chilean pediatric cohort is important to gain a comprehensive understanding of these traits and their health consequences. Thus, we aim to identify genetic risk factors for seven iron-related measures (ferritin, free erythrocyte protoporphyrin (FEP), hemoglobin (Hgb), serum iron, mean corpuscular volume (MCV), total iron binding capacity (TIBC), and transferrin saturation (SAT)) in the Santiago Longitudinal Study (SLS) of Chilean children at age 10 (N=770; 52% boys). **Methods:** The Illumina MEGA array was imputed using the 1000 Genomes Phase III AMR reference panel for all participants. We conducted linear regression, assuming an additive genetic model, adjusted for sex, population substructure, and random clinical trial arm of the parent infant study based on iron supplementation. Quality control included individual call rate >90%, SNP call rate >95%, imputation quality >0.3, effective N >10, gender mismatch, and ancestry outliers. **Results:** In total, we identified 22 significant loci ($P < 5 \times 10^{-8}$, MAF: 0.01-0.43, +/- 500kb) across the 7 traits. Of these 22 loci, 19 were novel associations, including one locus near *ARHGAP8* associated with Hgb, *HMGB1* and *MYLK* with Ferritin, and *SVIL* with MCV (MAF ≤ 0.02). Moreover, one locus associated with TIBC was *TF* ($P = 6 \times 10^{-16}$, MAF=0.08), and a variant of *TST* associated with both SAT and iron. We generalized several previous SNP-trait associations published, notably, *TF* and *TST* variant associations in these Chilean children were also observed in adult participants from Hispanic Community Health Study/Study of Latinos. **Conclusion:** We demonstrated association of both common and low-frequency variants with iron phenotypes in Chilean children, generalized some loci and reported concordant direction of effects with published findings. Importantly, we observed novel associations indicating possible age- and/or population-specific genetic effects on iron traits. Our findings support the importance of conducting GWAS studies in ancestrally diverse populations and across life stages to better understand the inherent genetic architecture of iron traits.

3296F

Characterisation and somatic mosaicism of the human glycophorin DUP4 structural variant, and association with haemoglobin levels in a malaria-endemic village. *W. Algady¹, S. Gomes², D. Carpenter¹, P. Brajer¹, A. Färnert³, I. Rooth⁴, M. Shaw⁵, F. Yang², E. Hollox¹.* 1) University of Leicester, Department of Genetics, United Kingdom; 2) Wellcome Sanger Institute, Cambridge, United Kingdom; 3) Infectious Disease Unit, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden; 4) Nyamisati Malaria Research, Rufiji, National Institute for Medical Research, Dar-es-Salaam, Tanzania; 5) School of Medicine, University of Leeds, Leeds, United Kingdom.

Human glycophorins A and B are proteins expressed on the surface of erythrocytes, and are receptors for invasion of the *Plasmodium falciparum* parasite, which causes malaria in sub-Saharan Africa. The proteins are encoded by the genes GYP A and GYP B which, together with GYPE, reside on a tandemly-duplicated repeat region on chromosome 4q31.21. Previous genome-wide analysis has shown that a structural variant within this region (DUP4), is identical to the blood group antigen Dantu NE+, and confers a clinically-important protective effect, is common in East Africans and is strongly protective against severe malaria, reducing the risk of severe malaria by up to 40%. DUP4 is a complex structural genomic variant that carries hybrid (GYP A/GYP B) fusion genes. Using fibre-FISH, we validate the structural arrangement of the glycophorin locus in the DUP4 variant, and provide evidence of somatic variation in the number of GYP A/GYP B fusion genes. Subsequently, we have developed a paralogue-specific junction fragment PCR to genotype DUP4. We demonstrate association of DUP4 variant with hemoglobin levels - a phenotype related to malaria - in 348 unrelated DNA samples from a Tanzanian village holoendemic for malaria using a family-based association test. Using the family-based association approach implemented in (QTD T), we have found a statistically significant association of the DUP4 variant with hemoglobin levels ($p=0.0054$). Our study confirms the importance of the DUP4 variant in malaria protection, and raises the intriguing possibility of heightened somatic instability and somatic mosaicism at this locus in DUP4 carriers, which might confer added protection against malaria.

3297W

Contributing to an inflammatory bowel disease meta-analysis with 8000 non-European whole genomes. *B. Avila^{1,2}, M. Rivas^{1,3}, T. Singh^{1,2}, J. Koskela^{1,2,4}, R. Maier^{1,2}, C. Stevens¹, I. Hall⁵, S. Brant^{6,7,8}, S. Kugathasan⁹, R. Newberry¹⁰, J. McCauley¹¹, M. Abreu¹², T. Matise¹³, S. Buyske¹⁴, D. McGovern¹⁵, T. Haritunians¹⁵, J. Cho¹⁶, M. Daly^{1,2,4}, NHGRI CCDG.* 1) Broad Institute of MIT and Harvard, Cambridge, MA; 2) Analytic and Translational Genetics Unit, Mass. General Hospital, Boston, MA; 3) Department of Biomedical Data Science, Stanford University, Stanford, CA; 4) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; 5) Washington University School of Medicine, St. Louis, MO; 6) Department of Medicine, Rutgers Robert Wood Johnson Medical School, New Brunswick, NJ; 7) Meyerhoff Inflammatory Bowel Disease Center, Department of Medicine, School of Medicine, Johns Hopkins University, Baltimore, Maryland, USA; 8) Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD; 9) Dept. of Pediatrics, Emory University School of Medicine, Atlanta, GA; 10) Department of Internal Medicine, Washington University School of Medicine St Louis, MO, USA; 11) John P. Hussman Institute for Human Genomics, Dr. John T. Macdonald Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL; 12) Division of Gastroenterology, Department of Medicine, University of Miami Miller School of Medicine, Miami, FL; 13) Genetics, Rutgers University, Piscataway, NJ; 14) Statistics and Biostatistics, Rutgers University, Piscataway, NJ; 15) Translational Genomics Unit. F. W. Idjaja Foundation Inflammatory Bowel and Immunobiology Research Institute, Cedars-Sinai Medical Center, Los Angeles, California, USA; 16) Icahn School of Medicine at Mount Sinai, Dr Henry D. Janowitz Division of Gastroenterology, New York, New York, USA.

As part of an ongoing IBD sequencing program, we have to date sequenced 7,915 whole genomes (3,347 IBD cases) with African American and Hispanic/Latino ancestry. While the sample size is yet relatively small for new locus discovery, we employ this data for use in several ways that aim to elucidate the largely unexplored contribution of non-European populations in rare variant studies of IBD. First, we utilize these data as a source of replication for results from an exome analysis of more than 15,000 IBD cases from four distinct European-derived populations. Here we can explore both effect and frequency distribution seen at many known IBD variants and are able to assess the utility of diverse populations to replicate new associations at different frequencies. We can further detect the existence of variants in known genes that are previously undetected because they are rare or absent in Europeans. For example, several variants in the NOD2 gene have higher effect sizes in European populations, but do not seem to contribute as much in African Americans. Nevertheless, we find NOD2 to have a similar importance in African-Americans through identification of a strongly associated missense variant not observed in Europeans. Beyond providing us with additional association and enrichment results, these genomes serve to reinforce the importance of Quality Control-related decisions in restricting and aligning whole genomes to an exome-based analysis. Preliminary results show that in both exomes and genomes, full-coverage filtering based on the GnomAD QC metrics have had the most significant impact on otherwise difficult to address noise in single-variant analysis.

3298T

Retrospective epidemiological study of Venezuelans patients with Haemoglobinopathies. Haematological characteristics, most frequent mutations, comprehensive clinical care and genetic counselling programs.

A.B. Falcon de Vargas. Clinical Genetics Unit, Hospital Vargas Caracas. Escuela Vargas. UCV. San Jose . Caracas. 1061.

A HVP-GG2020 Challenge protocol for Haemoglobinopathies in Developing Countries was used. Researchers and international publications were consulted. Major haemoglobinopathies in Venezuela are Sickle cell disease, beta thal minor, HbS, variants C, D. In 80400 haemolytic anaemia Venezuelan patients; 9% haemoglobinopathy, HbS, variants C and D. Frequency haplotypes bs gen in 272 with drepanocytosis (SCA) , heterocytotes for HbS was 50, 8%; Benin, 32,2%; CAR, 14; Senegal 2 and 2,3% Camerun. In homocytous HbS patients, 8% was Ben haplotype, 82% double heterocytote for haplotypes Ben/CAR, 8,8% haplotypes Ben, Senegal, CAR/Senegal y Ben/Camerun, one patient homocytos for haplotype CA (A Arends et al, Science and Technology, 2007, 32;8:516521). Sickle Cell Disease SCD , frequency from zero in Indians up to 5% in Mestizo and Afro-American population; 12% in the NC costal Region (Arends, 1971; Arends et al 1982). Seven polymorphic sites in b-globin gene cluster (State Aragua). Benin haplotype 0.479, Bantu haplotype 0.406; atypical Bantu A2 0.042; Senegal 0.031; atypical Bantu A7; 0.021 Saudi Arabia_Indian (0.021). (N Moreno et al, Genet. Mol. Biol. 25,1:2002). Sudanese, Bantu origins. 101.301 newborns were analysed (HPLC-CE; VBRS SC Short program ; Betathak program for family studies). 127 unrelated subjects suspicion β -thalassemia trait or clinically β -thal syndrome had DNA studies (PCR-bases reverse dot-blot method or amplification refractory mutation system (ARMS). 15 different mutations were identified , 92% of the mutant alleles studied, revealing genetic heterogeneity at the β - globin gene locus (M Bravo-Urquiola, Hemoglobin, 2012;36(3):209-18). Most frequent mutations: Codon 39 (C >T) 34.1%, IVS-I-(G >A) 11.1%, IVS-I-6(T >C=6.6%, -29 (A >G) 5.2%. Less common mutations IVS-I-5 (G >A) 3.7%; 1,393 bp deletion 3.0%; IVSII-1 (G >A) 3.0%, -86 (C >G) 2.2% , IVSII-1 (G >T) 1.5%; (C >T) codons 41/42 (-TCTT) 1.5%, IVSS-II745 (C >G) 0.7% and deletion at β - thal 0.7%. The major sources of β - thal-alleles in Venezuela are of Mediterranean and African origins. Diversity of haplotypes associated with some of the β -thal mutations explained by in situ recombination events in Venezuela.

3299F

Investigating the burden of nonsynonymous variation in PI3Kdelta pathway genes of individuals affected with idiopathic bronchiectasis.

E.M. Schmidt^{1,2}, K. Brown^{2,3}, L. He^{1,2}, S.L. Spain^{1,2}, R. McIntyre^{1,2}, I. Dunham^{2,4}, J.C. Barrett^{2,5}, S. Nejentsev^{2,6}, A.M. Condliffe^{2,7}, M. Ghousaini^{1,2}, E.M. Hessel^{2,8}, D. Michalovich^{2,8}, R.A. Floto^{2,3}. 1) Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SA, United Kingdom; 2) Open Targets, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD, United Kingdom; 3) Molecular Immunity Unit, Department of Medicine, University of Cambridge, Cambridge, CB2 0QH, United Kingdom; 4) European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD, United Kingdom; 5) Genomics plc, Oxford, OX1 1JD, United Kingdom; 6) Department of Medicine, University of Cambridge, Cambridge, CB2 0QQ, United Kingdom; 7) Department of Infection, Immunity and Cardiovascular Disease, School of Medicine, University of Sheffield, Sheffield, S10 2RX, United Kingdom; 8) Refractory Respiratory Inflammation DPU, GlaxoSmithKline plc, Stevenage, Hertfordshire, SG1 2NY, United Kingdom.

Bronchiectasis (BE) is a respiratory condition characterized by dilated bronchi, mucus hypersecretion, and chronic respiratory infection. Non-cystic fibrosis (CF) BE currently affects 300-650 in 100,000 people in the US and UK, with 35-50% of affected individuals having no obvious or easily definable disease origin (idiopathic). BE can also be caused by defined genetic disorders such as cystic fibrosis, primary ciliary dyskinesias, and immunodeficiencies (e.g. Activated PI3Kinase Delta Syndrome (APDS)). Currently, there are no specific licensed medicines for BE patients, presenting a clear need for deeper exploration into the genetic etiology of this disease. We performed whole exome sequencing of blood samples from 400 cases with idiopathic BE, and obtained whole exome sequence data generated for 4,300 healthy controls from the INTERVAL cohort (Cambridge, UK) at ~50x coverage. Variants were jointly called in the case and control cohorts followed by call set refinement, sample kinship estimation, and principal component ancestry analysis. Exome-wide association scans were carried out using both single variant (minor allele frequency (MAF) > 1%) and gene-based tests to evaluate the aggregate burden of low frequency (MAF < 5%) protein-truncating, protein-altering, and silent coding variants in a subset of 4,387 unrelated samples of recent European ancestry (373 cases and 4,014 controls). We also evaluated the burden of pathogenic variants present in curated gene sets of relevant pathways (e.g. PI3Kdelta signaling, ciliopathy) in BE cases versus controls. Of the 623,507 high quality variant sites identified in protein-coding regions, 8,505 were loss-of-function (LoF) variants (stop-gained, splice site disrupting, or frameshift; Loss-Of-Function Transcript Effect Estimator (LOFTEE) high confidence), and 155,880 were missense variants with CADD score > 21. As a first analysis, we evaluated the combined burden of LoF and missense variants in PI3Kdelta pathway genes (SKAT-O) including *AKT2* ($P=4 \times 10^{-4}$), *PIK3R1* ($P=0.02$), *FOXO1* ($P=0.02$), and *INPP5D* ($P=0.03$). We identified one affected heterozygous carrier of the APDS-causing E81K amino acid substitution in *PIK3CD*, subsequently confirmed by Sanger sequencing. Case-enrichment of pathogenic variants in genes of immune and airway-related pathways suggests potential mechanisms for disease susceptibility. Further sub-phenotyping will help identify patient subgroups that may benefit from specific therapeutic interventions.

3300W

Whole transcriptome cis-eQTL analysis of the nasal airway epithelium in childhood asthmatics reveals functional asthma risk variants. S.P. Sajuthi¹, B.A. Saefi¹, M.T. Montgomery¹, J.L. Everman¹, A.C. Mak², M. White², C. Eng², S. Oh², J. Rodriguez-Santana³, E.G. Burchard^{2,4}, M.A. Seibold^{1,5,6}. 1) Center for Genes, Environment, and Health, National Jewish Health, Denver, CO; 2) Department of Medicine, University of California San Francisco, San Francisco, CA; 3) Centro de Neumología Pediátrica, San Juan, Puerto Rico; 4) Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco, CA; 5) Department of Pediatrics, National Jewish Health, Denver, CO; 6) Division of Pulmonary Sciences and Critical Care Medicine, Department of Medicine, University of Colorado-AMC, Aurora, CO.

Genetic association studies of asthma have found most risk variants are non-coding, suggesting that these variants affect gene expression regulatory regions. Therefore, expression quantitative trait loci (eQTL) variants identified in asthma disease-involved tissues will likely be enriched for risk variants. The bronchial airway epithelium is central to asthma pathogenesis through the production of mucus, which causes airway obstruction and immune signaling leading to airway inflammation. The expense and safety concerns of performing bronchoscopies on children to collect bronchial airway epithelial brushings have impeded airway epithelial eQTL studies in children with asthma. Previously, we verified the use of minimally-invasive nasal airway epithelial brushings collected from children as an effective proxy of the bronchial airway epithelium. Therefore, we recruited a cohort of 698 children from the Genes-Environments, and Admixture in Latino Americans (GALA II) Study with nasal airway epithelial brushings to perform a cis-eQTL study. Whole transcriptome expression data was generated on nasal brushings from all subjects and paired with subject whole genome genotyping data. We found over 70% of expressed genes contained a significant cis-eQTL (eGenes=13,638). Stepwise regression analysis identified 24,737 independent eQTL variants, with 29.2%, 30.8%, and 40% of eGenes harboring 1, 2, or 3+ independent cis-eQTLs, respectively. Examining the 45 genes contained across the 18 validated asthma genomic loci identified in the largest asthma meta-analysis performed to date, we found 24 genes exhibiting a significant cis-eQTL. For example, we identified several strong independent eQTLs for both the GSDMB and PGAP3 genes at the 17q21 asthma locus. We next examined whether asthma eQTLs differed based on the airway inflammation exhibited by subjects. Specifically, we found 47% of subjects exhibited excessive airway inflammation triggered by type 2 cytokines (type 2-high disease) through gene expression clustering analyses. Meta-analysis of eQTL results for the two groups identified eGenes unique to type 2-high (n=1,547) and type 2-low (n=1,595) subjects. We are currently testing these eQTLs for association with type 2-high and -low asthma in >10,000 minority children. These results reveal the wide-spread nature of genetic regulation of airway epithelial gene expression including asthma genes, as well as the dependence of genetic effects on airway inflammation.

3301T

Novel strategies for gene-environment-wide interaction scans in family-based studies of admixed individuals. Y. Chen^{1,2}, I. Adrianto^{1,2}, M. Lanuzzi³, C. Montgomery⁴, B. Rybicki¹, A. Levin^{1,2}, J. Li^{1,2}. 1) Department of Public Health Science, Henry Ford Health System, Detroit, Michigan; 2) Center for Bioinformatics, Henry Ford Health System, Detroit, Michigan; 3) Department of Internal Medicine, State University of New York, Upstate Medical University Hospital, Syracuse, NY; 4) Arthritis and Clinical Immunology Research Program, Oklahoma Medical Research Foundation, Oklahoma City, Oklahoma.

Etiology of many complex diseases involves both environmental exposures and inherited genetic pre-disposition, and interactions between them. Genome-environment-wide interaction scans (GEWIS) provide a means to identify the interactions between genetic variation and environmental exposures that underlie disease risk. However, current GEWIS methods lack the capability to adjust for the potentially complex correlations in family based study designs carried out in admixed populations. Here, we developed novel methods based on generalized estimating equation (GEE) -- GEE-adaptive and GEE-joint to account for phenotypic correlations due to kinship while accounting for covariates, including measures of genome-wide ancestry. In two simulated family study designs of admixed individuals, both methods demonstrated higher power than traditional case-control method and controlled the family-wise error rate (FWER), an advantage over the case-only approach. Applying the proposed method to a real data analysis of insecticide exposure and risk of sarcoidosis in African American, we identified 19 potential SNPs interacting with insecticide exposure at a p-value threshold < 1*10⁻⁵.

3302F

Genome-wide scan identifies loci for white blood cell and platelet traits in a multi-ethnic study from population architecture using genetic epidemiology (PAGE). S.A. Bien¹, K.K. Nishimura¹, J. Haessler¹, C. Hodonsky², M. Graff³, R. Tao³, G. Nadkarni⁴, S. Buyske⁵, C. Haiman⁶, L. Le Marchand⁷, R.J.F. Loos⁸, T. Matise⁹, K.E. North², A. Reiner¹, U. Peters¹, C.S. Carlson¹, C.L. Avery², C. Kooperberg¹, *Population Architecture using Genetic Epidemiology (PAGE).*

1) Cancer Prevention, Fred Hutchinson Cancer Research Center, Seattle, WA; 2) Department of Epidemiology, University of North Carolina Chapel Hill, Chapel Hill, NC; 3) Vanderbilt University Medical Center, Nashville, TN; 4) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY; 5) Department of Statistics and Bio-statistics, Rutgers University, New Brunswick, NJ; 6) Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA; 7) Cancer Research Center, University of Hawaii, Honolulu, HI; 8) Department of Genetics, Rutgers University, New Brunswick, NJ.

Circulating White Blood Cell (WBC) and platelet counts are moderately heritable and vary in healthy individuals by race/ethnicity. Importantly, circulating WBC and platelet counts are used to alert clinicians to possible undiagnosed conditions, such as autoimmune diseases, immune deficiencies and blood disorders. We conducted a large, multi-ethnic GWAS in the Population Architecture using Genomics and Epidemiology (PAGE) Study to gain a deeper understanding of how ancestry-specific genetic variants may influence 8 indices of WBC and platelet counts. Transethnic meta-analysis was performed for 70,428 individuals of African American (AFR), Hispanic/Latino (H/L), Native American, or European (EUR) ancestry with WBC and platelet measurements and genotyped on either the multiethnic genotyping array (n=23,907) or another GWAS platform. These data were imputed into the 1000 Genomes Phase 3 data. Those with measurement $\pm$ 4 standard deviations were excluded. SUGEN software was used to perform linear regression for each trait stratified by race/ethnicity and adjusting for age and age² at WBC measurement, sex, body mass index, current smoking status, self-identified race/ethnicity, study center, family/household membership, and 10 ancestral principal components (PCs). Variants were then tested for association using rank-based inverse normal transformation of the residuals from these models and results were combined using fixed effects meta-analysis. We identified 100 regions independently associated at $p < 5 \times 10^{-8}$ across 8 blood cell indices. Signals were detected in 7 regions not previously associated with any of the blood cell indices analyzed (1q25.1-*RC3H1*, 1q25.1-*RABGAP1L*, 2p23.3-*HADHB*, 7q21.3-*GNG11*, 7q36.2-*INSIG1*, 12q32.2-*IGF1*, and 8q24.22-*SLA*), and 5 known regions associated blood cell trait differing from the trait previously reported, (1p34.2-*MPL*, 2q21.3-*CXCR4*, 3p22.2-*CX3CR1*, 9q34.11, and 12q24.21-*MED13L*). Additional independent associations were detected in 11 known loci, which remained genome-wide significant after conditioning on previously reported associations. As clinical practice begins to move towards precision medicine, it is critically important to perform GWAS in ancestrally diverse study populations in order to gain better understanding of genetic influences on biomarkers that should be accounted for when interpreting their circulating concentrations. Additional analyses including gene burden tests, and finemapping are ongoing.

3303W

Association between the risk of monoclonal B-cell lymphocytosis (MBL) and the chronic lymphocytic leukemia (CLL) polygenic risk score (PRS).

G. Kleinstern¹, S.J. Achenbach¹, D.P. Robinson¹, A.D. Norman¹, K.G. Chaffee¹, C.E. Lesnick¹, S.A. Parikh², T.G. Call³, J.E. Olson¹, J.R. Cerhan¹, N.E. Kay⁴, C.R. Hanson¹, T.D. Shanafelt¹, S.L. Slager¹. 1) Department of Health Sciences Research, Mayo Clinic, Rochester MN, United States; 2) Division of Hematology, Mayo Clinic, Rochester, MN, United States; 3) Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, United States; 4) Department of Medicine, Division of Hematology, Stanford University, Stanford, CA, United States.

Background MBL is a precursor state to CLL, with a prevalence of ~5-8% in the general population and 15-18% in CLL families. To date, 41 single nucleotide polymorphisms (SNPs) have been found to be associated with CLL. We previously showed that a PRS of the weighted average of the number of risk alleles of these SNPs is associated with CLL risk using cases and controls from the International Lymphoma Epidemiology (InterLymph) Consortium. We validated this score in an independent sample of CLL cases and controls from the Genetic Epidemiology of CLL (GEC) Consortium, a cohort of families with ≥ 2 members with CLL. In the CLL families we also reported an association between the PRS and MBL risk. Here we evaluate the PRS in an independent sample ascertained from the Mayo Clinic Biobank. **Methods** The Mayo Clinic Biobank is a large-scale bio-repository of adult patients assembled to provide a wide array of health-related research studies. Biobank participants were screened for MBL using a highly-sensitive, 8-color flow-cytometry assay on stored cryopreserved peripheral blood mononucleotide cells. Low-count MBL was defined as those individuals who had $< 85\%$ clonal B-cells out of total B-cells. Individuals without MBL ("controls") were frequency matched to MBLs based on age and sex. Participants were genotyped using the Illumina OmniExpress array. Standard genotyping quality-control metrics were applied. We computed the PRS as previously published and categorized it by quintiles based on values previously used with InterLymph controls. The middle quintile served as the reference category. We used logistic regression to estimate odds ratios (ORs) and 95% confidence intervals (CI), adjusting for age at consent, sex, and socioeconomic status. **Results** We identified 276 MBLs and 844 controls; 97% of MBLs had low-count MBL. Family history of leukemia/lymphoma was 11.3% among MBLs and 8.8% among controls. The continuous PRS had a 1.75-fold increased risk for MBL (CI: 1.5-2.1, $P = 2.65 \times 10^{-12}$). Compared to the middle quintile, the highest quintile had 2.2-fold increased risk for MBL (CI: 1.5-3.3, $P = 1.46 \times 10^{-4}$) and the lowest quintile had 0.6-fold decreased risk (CI: 0.4-0.96, $P = 0.03$). **Conclusion** In this independent sample, we validated our previous findings that the CLL PRS is associated with MBL risk. These results may help identify individuals at higher risk of MBL (especially low-count MBL) beyond the risk associated with age and family history of CLL.

3304T

Whole genome sequences association with E-selectin levels reveals

loss-of-function variant in African Americans. L.M. Polfus¹, L.M. Raffield², R.P. Tracy^{3,4}, L.A. Lange⁵, G. Lettre⁶, J.G. Wilson⁷, J.C. Bis^{8,9}, R.P. Bowler¹⁰, A. Correa¹, R.J.F. Loos^{11,12,13}, B. Li^{14,15}, A.P. Reiner^{16,17}, P.W.L. Auer¹⁸ on behalf of the TOPMed Inflammation Working Group. 1) Center for Genetic Epidemiology, University of Southern California, Los Angeles, CA; 2) Department of Genetics, University of North Carolina, Chapel Hill, NC, USA; 3) Department of Pathology and Laboratory Medicine, Robert Larner, M.D. College of Medicine, University of Vermont, Burlington, VT, 05405, USA; 4) Department of Biochemistry, Robert Larner, M.D. College of Medicine, University of Vermont, Burlington, VT, 05405, USA; 5) Department of Medicine, University of Colorado Denver, Anschutz Medical Campus, Aurora, CO, 80045, USA; 6) Department of Medicine, Université de Montréal, Montréal, QC H3T 1J4, Canada; Montreal Heart Institute, Montréal, QC H1T 1C8, Canada; 7) Department of Pediatrics and Medicine, University of Mississippi Medical Center, Jackson, MS, USA; 8) Cardiovascular Health Research Unit, University of Washington, 1730 Minor Avenue, Suite 1360, Seattle, Washington 98101, USA; 9) Department of Medicine, University of Washington, 1959 Northeast Pacific Street, Box 356420, Seattle, Washington 98195-6420, USA; 10) Division of Pulmonary, Critical Care and Sleep Medicine, Department of Medicine, National Jewish Health, Denver, Colorado; and Department of Radiology, USA; 11) Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 12) Genetics of Obesity and Related Metabolic Traits Program, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 13) Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 14) Vanderbilt University, Genetics Institute, Nashville, TN, USA; 15) Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA; 16) Department of Epidemiology, University of Washington, Seattle, WA, USA; 17) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA; 18) Zilber School of Public Health, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin, USA.

E-selectin is highly expressed on endothelial cells when activated by cytokines or chemokines, and is specifically involved in the tether and rolling phases of the leukocyte-endothelial adhesion during various inflammatory and immune responses. Previous genome-wide association studies (GWAS) have identified the *ABO* locus associated with E-Selectin levels in European and Asian individuals. Using whole-genome sequencing (WGS) data in 2,458 African Americans (AAs) from the Jackson Heart Study (JHS) in the Trans-Omics for Precision Medicine (TOPMed) Program, we examined genome-wide associations with soluble E-Selectin levels. We performed discovery analyses for 30,016,534 single nucleotide variants and small indels with minor allele frequency (MAF) ≥ 0.001 . Three loci surpassed a genome-wide significance threshold: 9q24 (85 variants, sentinel: *ABO* rs532436, $P=2 \times 10^{-83}$), 1q24 (70 variants, sentinel *SELL-SELE* rs139825795, $P=7 \times 10^{-48}$), and 19p13 (31 variants, sentinel: *FUT6* rs17855739, $P=7 \times 10^{-24}$). Because the *FUT6* rs17855739 (p.Glu274Lys) loss-of-function variant is considerably more common in populations of African ancestry (MAF=0.28-0.29) compared to non-African ancestry populations (MAF=0.03-0.06), we performed targeted replication and phenotypic association studies in another African-American population. We replicated the association of *FUT6* (rs17855739) with higher soluble E-selectin in an independent population of 748 AAs from the Women's Health Initiative ($P=4 \times 10^{-5}$) and identified a pleiotropic association with vitamin B12 levels. We also conducted electronic medical record (EMR)-based phenome-wide association scan of these loci in over 9,000 African Americans from the BioVU and BioMe EMR-based biobanks. Despite the broad role of both selectins and fucosyltransferases in various inflammatory, immune and cancer-related processes, we were unable to identify any additional disease associations of the *FUT6* rs17855739 variant in an EMR-based phenome-wide association scan. The African *FUT6* rs17855739 is associated with plasma alpha(1,3)-fucosyltransferase deficiency, and should be further investigated for association with chronic disease outcomes in African Americans, particularly given the potential importance of E-selectin-interactions as a therapeutic target for various chronic inflammatory conditions.

3305F

Understanding the genetic determinants of peanut-specific IgG4 in the learning early about peanut allergy (LEAP) study.

A.H. Winters¹, M. Boor-gula², H.T. Bahnson^{3,7}, S. Chavaan², G. Du Toit^{4,5}, I. Ruczinski¹, P.H. Sayre^{3,6}, K.C. Barnes², G.T. Nepom⁷, G. Lack^{4,5}, R.A. Mathias¹, the Immune Tolerance Network Learning Early About Peanut Allergy study team. 1) Johns Hopkins University, Baltimore, MD; 2) University of Colorado, Anschutz, CO; 3) Immune Tolerance Network, San Francisco, CA; 4) Department of Paediatric Allergy, Division of Asthma, Allergy and Lung Biology, King's College London, UK; 5) Guy's and St. Thomas' NHS Foundation Trust, London, UK; 6) Department of Medicine, University of California, San Francisco, CA; 7) Benaroya Research Institute, Seattle, WA.

The LEAP study showed a protective effect of early oral peanut exposure for infants at high risk of developing peanut allergy, with an 81% reduction in peanut allergy incidence in the peanut consumption group compared to the peanut avoidance group. We have now performed whole genome sequencing (WGS) to understand the genetic determinants of clinical outcomes and biomarker correlates of peanut allergy in the LEAP Study. It has been shown that peanut-specific IgG4 increased more in the peanut consumption group as compared to the peanut avoidance group and showed the largest difference between the two groups amongst the biomarkers measured in LEAP. In this report, we leverage array data from the Illumina Omni 2.5 bead chip, generated as part of our WGS quality control pipeline, to begin to understand the genetic determinants of peanut-specific IgG4. Single SNP tests for association for peanut-specific IgG4 at 60 months of age (the conclusion of the LEAP trial) were performed on 267 participants in the peanut consumption group. The peak association identified was with the variants rs75590598 and rs11681497 ($p=5.79 \times 10^{-7}$ and 9.33×10^{-7} , respectively). These are located on chromosome 2 in the *HDLBP*, *SEPT2*, and *FARP2* gene cluster. Both the SNPs are common intronic variants with minor allele frequencies $>10\%$. While rs75590598 maps to *HDLBP* and rs11681497 to *FARP2*, both are reported to be expression quantitative trait loci (eQTLs) for the *SEPT2* gene in a number of tissues in GTEx. A second signal was seen on chromosome 6 for rs17612852 ($p=5.80 \times 10^{-7}$) which maps to the region between *HLA-DQA1* and *HLA-DQB1*. rs17612852 is an eQTL for *HLA-DQB1* in multiple tissues in GTEx. *SEPT2* has previously been shown to regulate airway epithelial barrier function, and deficiencies in the epithelial barrier are known to be associated with allergy. Additionally, previous work has shown association with variants in *HLA-DQB1* and increased risk of peanut allergy independent of asthma, implicating this locus in peanut allergy-specific risk and not just general atopic risk. We are currently analyzing all sequence-identified variants from the WGS to follow up on these results from the GWAS array and to identify additional rare and novel variants associated with peanut-specific IgG4. We are also extending these analyses to other biomarkers associated with peanut allergy as measured in the LEAP subjects.

3306W

Genetic modifiers of IgE response in AD. C. Polymeropoulos, S. Smieszek, M. Polymeropoulos. Vanda Pharmaceuticals, DC.

Immunoglobulin E (IgE) mediated allergy to environmental allergens plays a central role in the pathophysiology of asthma, atopic dermatitis, and food allergy. Atopy and plasma IgE concentration are genetically complex traits, and the specific genetic risk factors that lead to IgE dysregulation and clinical atopy are an area of active investigation. To ascertain the genetic risk factors which lead to IgE dysregulation we conducted a genome-wide association analysis using 117 whole genome sequencing atopic dermatitis samples. Using linear regression, we have directly tested the association between 14,322,979 SNPs and log transformed IgE response, correcting for covariates including sex and principal components. We detected 24 loci at $P < 9.5 \times 10^{-8}$. Interestingly there was a strong signal within 6q23 region containing immune response genes such as Interleukin 20 Receptor Subunit Alpha (*IL20RA*) and Interleukin 20 Receptor Subunit Alpha (*IL22RA2*). The *IL20RA* gene encodes the IL-20 receptor α subunit, which can form a heterodimeric receptor with either IL-20RB to bind IL-19, IL-20 and IL-24, or with IL-10RB to bind IL-26. Evidence suggests that this family of cytokines have a pro-inflammatory effect and involvement in skin inflammation. This gene and interleukin 20 receptor beta are highly expressed in the skin, and are upregulated in psoriasis. Top loci were enriched in cytokine receptor activity and interleukin-20 binding (GO:0042018, GO:0042020). Haploreg v4.1 was used to investigate the regulatory potential, showing that rs276571 demonstrates a number of lines of evidence to support a function in disease causality, including mapping to an enhancer in B-lymphoblastoid cell lines, primary and T-regulatory cells. It also maps to a region of open chromatin, characterised by DNase hypersensitivity, shows evidence of binding of STAT3. Furthermore the locus has been shown to be a significant (0.000004) eQTL for *IL20RA* in GTEx. The risk allele is effectively correlated with higher expression of *IL20RA*. We can explain a high proportion of variance (GCTA) in IgE within the AD samples (p val 8.882×10^{-16} , SE 0.09). Moreover top hits include some other interesting associations such as a missense variant rs2280090 within a disintegrin and metalloprotease 33 gene *ADAM33*. Interestingly, our analysis led to functional and actionable results of relevance to the greater understanding of AD and IgE responses that could aid in therapeutics with genetic support.

3307T

A genome-wide association study of intracellular phenotypes reveals novel haematological biology. P. Akbari¹, T. Jiang¹, D. Vuckovic², D.J. Roberts^{2,3}, E. Di Angelantonio¹, J. Danesh¹, W.H. Ouwehand⁴, A.S. Butterworth¹, N. Soranzo², W.J. Astle^{1,3}. 1) Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Strangeways Research Laboratory; 2) Department of Human Genetics, The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge; 3) Department of Haematology, Churchill Hospital, Headington, Oxford; 4) Department of Haematology, University of Cambridge, Cambridge Biomedical Campus, Long Road, Cambridge.

An adult human synthesises approximately 1012 blood cells per day. These cells have crucial functions in respiration, immunity and wound healing, and have been implicated in the aetiology of many common diseases, ranging from asthma to stroke. However, our understanding of the regulation of the production, clearance and activation of blood cells is limited. Up to now, population genetic studies of quantitative cellular blood phenotypes have focused on cell count/volume (CBC) traits, which are largely insensitive to intracellular functional mechanisms. We used the Sysmex XN advanced haematology analyser to assay ~80 novel cellular traits in ~40,000 blood donors participating in the INTERVAL study. These traits include measures of nucleic acid-content, membrane lipid content, granularity, neutrophil activation and lymphocyte reactivity, previously unstudied by human geneticists. We conducted a GWAS of ~30 million imputed variants, identifying 929 loci associated with at least one phenotype. Comparisons with a recent GWAS of CBC traits measured in ~170,000 participants (Astle et al, Cell, 2016) showed that just 54% of the intracellular trait associated loci had been detected by the previous study despite its 4 fold larger sample, demonstrating that the new phenotypes detect a novel spectrum of biological variation. Functionally relevant ontology terms such as 'defense response to bacterium' and 'T cell differentiation' were respectively enriched in genes associated with neutrophil and lymphocyte traits. Neutrophil granularity was significantly associated ($p=4.2 \times 10^{-313}$) with the azurophilic granule residing antimicrobial defensin locus (DEFA4, DEFA6, DEFA8P &c) and Mendelian randomisation implicated defensins in the aetiology of coronary artery disease. Platelet granularity was associated with a common variant linked to NBEAL2 ($p=4.1 \times 10^{-12}$), rare mutations in which cause gray platelet syndrome, a disorder with bleeding symptoms caused by the dysfunction of alpha-granules. We identified associations with a number of alleles implicated in rare recessive disorders, such as those in SPTA1 ($p=5.4 \times 10^{-47}$) known to cause hereditary pyropoikilocytosis, a form of hemolytic anemia characterized by an abnormal sensitivity of red blood cells to heat. Our analysis, which is the first large scale investigation of flow cytometry derived intracellular phenotypes, reveals novel haematological biology with implications for understanding the aetiology of rare and complex diseases.

3308F

Optimizing chronic hepatitis C treatment in a Canadian cohort using precision medicine based methods. J.J. Lin¹, C.M. Loucks², J.N. Trueman², E.M. Yoshida³, J. Ford³, S.S. Lee⁴, E. Tam⁵, A. Ramji⁵, B.I. Drogemoller^{2,6}, G.E.B. Wright^{1,2}, C.J. Ross⁶, B.C. Carleton^{1,2,6}. 1) Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada; 2) Department of Pediatrics, University of British Columbia, Vancouver, BC, Canada; 3) Division of Gastroenterology, University of British Columbia, Vancouver, BC, Canada; 4) Division of Gastroenterology and Hepatology, University of Calgary Liver Unit, Calgary, AB; 5) LAIR Centre, Vancouver, BC, Canada; 6) Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver, BC, Canada.

Background: Chronic Hepatitis C virus (HCV) infection is a major cause of cirrhosis and liver disease, making it a leading indicator for liver transplantation. In the era of direct acting antiviral therapy for HCV with impressive cure rates over 95%, ribavirin (RBV) is still used to improve outcomes for difficult to treat patients and shorten therapy length. However, RBV can cause serious anemia in up to 50% of patients, requiring dose reductions to prevent further harm. This can compromise treatment success, resulting in longer treatment periods and higher costs. We aim to validate previously identified genomic variants, and discover novel variants associated with RBV-induced anemia in Canadian patients and develop a pharmacogenomic prediction model identifying patients most likely to develop anemia due to RBV-containing treatments. **Methods:** We have recruited patients treated with RBV-containing regimens from five sites in three cities across Canada. A genome-wide association study will then be conducted to (i) validate the role that previously associated variants (e.g. rs1127354 and rs7270101 in *ITPA*) play in current RBV treatment regimens and (ii) uncover novel genomic predictors of RBV-induced anemia. **Results:** 191 patients treated with RBV-containing regimens were recruited. Of these, 50 cases were identified and defined as those with serious anemia that warranted a change in treatment in the form of dose reductions in RBV or erythropoietin supplementation. These patients had a median hemoglobin decline of 38 g/L during treatment. The case cohort is comprised of a higher proportion of viral genotype 1 infected patients and lower proportion of viral genotype 3. We are currently assessing recruited patients for ~700,000 variants across the genome for genomic analyses which will be completed by September 2018. **Conclusions:** Downstream genomic analyses will need to take into account the differences in HCV genotype distribution in the studied cohort. Knowing in whom serious anemia to RBV is likely to occur, based on precision medicine, will allow for tailoring of therapy to individual patients to increase the probability of treatment success and minimize the likelihood of developing serious anemia. This will be particularly interesting in our Canadian cohort, as it is composed of patients of various ancestries that may have unique predispositions to anemia. Together, this will help decrease both economic and psychological burdens on affected families.

3309W

Identifying transcriptional and epigenetic mechanisms associated with genetic risk for inflammatory bowel disease using causal mediation analysis. K.T. Nguyen^{1,2}, Y. Park^{1,2}, L. Hou^{1,2}, C. Boix^{1,2}, L. He^{1,2}, M. Kellis^{1,2}. 1) Massachusetts Institute of Technology, Cambridge, MA; 2) Broad Institute of MIT and Harvard, Cambridge, MA.

Identifying the mechanisms by which genetic variants influence risk of complex disease is a fundamental problem in human genetics. Summary-based transcriptome-wide association studies (sTWAS) have helped address this problem, identifying genes associated with human disease by aggregating cis-regulatory effects on gene expression learned in expression quantitative trait loci (QTL) studies. Our contribution to this problem, Causal Multivariate Mediation within Extended Linkage disequilibrium (CaMMEL) is a Bayesian inference framework that jointly models multiple mediated and unmediated effects of GWAS risk variants. We showed that this method addresses limitations of other methods of detecting causal mediation by allowing for horizontal pleiotropy and other unmediated effects, and jointly modeling mediators in order to select a sparse set of mediators in LD regions. In addition to expression QTLs, the availability of QTL data for other types of molecular phenotypes has been rapidly increasing, allowing us to investigate additional mechanisms of disease etiology. Recently, the BLUEPRINT consortium generated a rich dataset of QTLs for a wide variety of molecular phenotypes, including gene expression (eQTL), histone modifications H3K27ac and H3K4me1 (hQTL), splicing (sQTL), and DNA methylation (meQTL), generated in purified samples of monocytes, T-cells, and neutrophils. We applied CaMMEL to this data to investigate the mediation effect of these molecular phenotypes on inflammatory bowel disease (IBD), a complex autoimmune disease. This has allowed us to identify regulatory regions whose activity is predicted to mediate IBD risk. These regions are enriched in motifs of transcription factors such as LXR and KLF4, which have both been experimentally linked to protection against induced colitis in mice. We also determined that those regions are associated with GO pathways such as negative regulation of developmental growth and regulation of T cell mediated cytotoxicity. This work allows us to gain specific insights into the mechanisms of action of risk variants and to identify therapeutic targets for IBD. More generally, we will also be able to evaluate the contribution of the various types of molecular QTL and the importance of cell type-specific information to disease etiology.

3310T

Whole-genome sequencing identifies *CRISPLD2* as a candidate gene for lung function in children with airway hyperresponsive asthma. P. Kachroo¹, J. Hecker^{2,3}, B. Chawes^{1,4}, C. Lange^{1,3}, S.T. Weiss¹, J. Lasky-Su¹. 1) Channing Division of Network Medicine, Department of Medicine, Brigham & Women's Hospital and Harvard Medical School, Boston, MA; 2) Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA; 3) Department of Genomic Mathematics, University of Bonn, Bonn, GE; 4) COP-SAC, Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital, University of Copenhagen, Copenhagen, DK.

Background: Resequencing of common asthma candidate genes discovered by genome-wide association studies has shown that uncommon and rare variants contribute to asthma susceptibility. This implies the need for high-coverage sequencing to unravel the genetic basis of asthma. **Objective:** To identify regions of genetic variation contributing to airway hyperresponsive asthma affection status and lung function utilizing whole-genome sequencing (WGS) data from Costa Rican children with asthma. **Methods:** WGS data were generated on 1,053 individuals from trios and extended pedigrees participating in the family-based 'Genetic Epidemiology of Asthma in Costa Rica' cohort. Asthma affection status was defined by asthma plus bronchial hyper-responsiveness to methacholine, and lung function (FEV1, FEV1/FVC-ratio) was measured at baseline and post-bronchodilator. Analyses were done with family-based association test (FBAT) methods for both single-variant and region-based tests. P-values were combined for meta-analysis in two other independent cohorts that defined asthma similarly: CAMP and Hartford-Puerto Rico using metaP package in R. **Results:** After data cleaning, 25.76 million autosomal SNPs remained for single-variant FBAT, showing a genome-wide significant association with baseline FEV1/FVC-ratio for a SNP in *CRISPLD2* (rs12051168, $p=3.6 \times 10^{-8}$). This SNP also had nominally significant associations for asthma affection status ($p=0.025$), baseline FEV1 ($p=0.044$) and post-bronchodilator FEV1/FVC-ratio ($p=2.6 \times 10^{-3}$). Additionally, rs12051168 could be replicated across two independent cohorts (Puerto Rico and CAMP) for baseline FEV1 ($p=0.048$). Combined P value for rs12051168 was significant across all cohorts for baseline FEV1 ($p=0.002$) and FEV1/FVC ratio ($p=1.47 \times 10^{-4}$) and post-bronchodilator FEV1 ($p=0.006$) and FEV/FVC ratio ($p=0.016$). A total of 327,761 regions with at least 6 SNPs were constructed for the region-based FBAT analysis, which identified several regions with association p-value below 10^{-5} that supports the single-variant findings. **Conclusions:** This WGS analysis of asthma defined utilizing airway hyperresponsiveness identified *CRISPLD2* as a lung function candidate gene. It has previously been shown that *CRISPLD2* regulates anti-inflammatory effects of glucocorticoids in the airway smooth muscle cells, whereas these WGS findings suggest that *CRISPLD2* may play a larger role in asthma pathogenesis.

3311F

Identifying modifying genes to explain the variation in severity of fragile X-associated primary ovarian insufficiency. C.E. Trevino, E.G. Allen, P. Jin, H.R. Johnston, D.J. Cutler, M.E. Zwick, S.T. Warren, S.L. Sherman. Emory University, Atlanta, GA.

Primary ovarian insufficiency (POI) is a condition characterized by symptoms of early menopause and infertility that affects 1% of the general population, making it a leading cause of infertility in women. Patients with this condition also display elevated follicle stimulating hormone, low anti-Müllerian hormone levels and early estrogen deficiency. Women who carry the *FMR1* premutation, a CGG repeat expansion in the range of 55-200 repeats in the 5'UTR of the X-linked *FMR1* gene, are at a significantly increased risk for POI: about 20% of carriers are diagnosed with fragile X-associated POI (FXPOI). Although the risk for FXPOI is significant, not all carriers experience the disorder, and among those who do, the range of severity is broad. We hypothesize that there are other genetic variants, in addition to premutation repeat number, that modify the age of onset and severity of FXPOI. To test this hypothesis, we conducted a whole genome sequencing (WGS), case/control study among women who carry the premutation. For this study, cases are defined as women with the premutation who experience secondary amenorrhea before the age of 35 and controls are women with the premutation who experience natural menopause after the age of 50. For the first 65 cases and 55 controls, variants were called using a well-established pipeline of WGS analyses and candidate genes were prioritized using burden testing of rare variants, expression in ovaries, and presence of a fly ortholog. The top candidate genes (*DCP2*, *TIA1*, *SUMO1*, *STAT1*, and *KRR1*) are currently being screened using the *Drosophila* model as a high-throughput, whole organism functional screen. This intermediate screen points to the next set of experiments to identify a set of genes that influence ovarian dysfunction in a cost effective way. The overall goal of this project is to elucidate novel molecular mechanisms for FXPOI to facilitate early risk assessment and diagnosis, allowing more informed choices about family planning and fertility options. In addition, the identified genes may also play a role in idiopathic POI and lead to a new understanding of possible treatment targets.

3312W

Kidney Genome Atlas: A whole-genome landscape of more than 2,000 kidney disease patients.

W. Zhang¹, T. Soare¹, A. Tebbe¹, L. Walsh¹, M. Kretzler², G.N. Nadkarni³, R.A. Gbadegesin⁴, J.L. Wenke⁵, P.H. Mundel¹, T.T. Tibbitts¹. 1) Goldfinch Bio, Cambridge, MA; 2) U.Michigan, Ann Arbor, MI; 3) Ichan School of Medicine, New York, NY; 4) Duke University Medical Center, Durham, NC; 5) Nashville Biosciences, Nashville, TN.

Chronic kidney disease (CKD) is a heterogeneous disease affecting more than 30 million people in the US, which is about 10% of the US population. Despite the urgent need for targeted therapeutics, the understanding of the mechanistic basis of CKD including the genetic variants that potentially drive it has lagged other diseases (e.g. cancer) for decades. Here we report the generation and initial analysis of the Kidney Genome Atlas (KGA), the world's largest whole-genome landscape of individuals with molecularly defined kidney diseases. Focusing initially on focal segmental glomerulosclerosis (FSGS), related disorders, and diabetic kidney disease (DKD), KGA contains whole genome sequences (>30X coverage) from more than 2,000 patients and 2,500 matched healthy controls. Each genome is linked to longitudinal clinical records and for a subset of 500 patients, the atlas also includes matched transcriptomic data from microdissected glomerular and tubulointerstitial samples. The KGA has enabled the discovery of genetic variants associated with kidney disease and the integration of genomic and transcriptomic data to identify kidney disease-specific expression quantitative trait loci (eQTLs). The atlas also provides the foundation for establishing the relationships between genetic variants, histologic diagnoses, and quantitative clinical phenotypes of kidney function and disease progression. Computational integration of these datasets will enable the prioritization of candidate variants with putative disease modulating effect. Investigating and validating biological pathways derived from these analyses can also be used to stratify patients into subtypes most likely to respond to specific targeted therapies. The KGA is a valuable resource fueling our understanding of the molecular mechanisms of kidney diseases at the whole genome scale.

3313T

Genome-wide meta-analysis of susceptibility to idiopathic pulmonary fibrosis.

R.J. Allen¹, B. Guillen-Guio², J.M. Oldham³, S.F. Ma⁴, T.E. Fingerlin^{5,6}, M. Ng⁷, R. Braybrooke^{8,9}, I. Sayers^{8,10}, D. Furniss⁷, K.T. Zondervan¹¹, A.P. Morris^{11,12}, I.P. Hall^{8,10}, M.D. Tobin^{1,13}, C. Flores^{2,14,15}, D.A. Schwartz^{16,17}, I. Noth⁴, R.G. Jenkins^{8,10}, L.V. Wain^{1,13}, UK ILD Consortium. 1) Department of Health Sciences, University of Leicester, Leicester, UK; 2) Hospital Universitario N.S. de Candelaria, Universidad de La Laguna, Santa Cruz de Tenerife, Spain; 3) Department of Internal Medicine, University of California Davis, Davis, USA; 4) Division of Pulmonary & Critical Care Medicine, University of Virginia, Virginia, USA; 5) Center for Genes, Environment and Health, National Jewish Health, Denver, USA; 6) Department of Biostatistics and Informatics, University of Colorado, Denver, USA; 7) Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK; 8) NIHR, Nottingham BRC, Nottingham, UK; 9) Division of Epidemiology and Public Health, University of Nottingham, Nottingham, UK; 10) Division of Respiratory Medicine, University of Nottingham, Nottingham, UK; 11) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 12) Department of Biostatistics, University of Liverpool, Liverpool, UK; 13) NIHR, Leicester Respiratory BRC, Leicester, UK; 14) CIBER de Enfermedades Respiratorias, Instituto de Salud Carlos III, Spain; 15) ITER, Santa Cruz de Tenerife, Spain; 16) Department of Immunology, University of Colorado, Denver, USA; 17) Department of Medicine, University of Colorado, Denver, USA.

Idiopathic pulmonary fibrosis (IPF) is a rare complex lung disease of unknown cause characterised by scarring of the lung. There are few effective treatments and half of patients die within three years of diagnosis. Previous genome-wide association studies (GWAS) have identified 17 genetic signals of association with IPF susceptibility, implicating genes involved in host defence, cell-cell adhesion, regulation of growth factors and telomere maintenance. The strongest genetic association is in the promoter of *MUC5B* where individuals have around 5 times the risk of disease for each copy of the risk allele. Using polygenic risk scores we were able to show IPF is highly polygenic with possibly hundreds of variants associated with the disease. Therefore, we performed a genome-wide meta-analysis on three IPF studies to identify additional genetic variants associated with susceptibility to IPF. In total 2,769 IPF cases and 8,591 controls were included in the meta-analysis. Each study was imputed to the HRC reference panel and GWAS were performed adjusting for 10 principal components to account for population structure. There were 10,768,131 variants measured in at least two of the studies and included in the final meta-analysis. Variants were deemed to show an association with IPF if they reached genome-wide significance (meta-analysis $p < 5 \times 10^{-8}$), were nominally significant ($p < 0.05$) in each study, and had consistent direction of effects across studies. Association with gene expression was investigated using the NESDA-NTR (blood) and GTEx (multiple tissues) eQTL databases. Three common novel association signals met the significance criteria and all had supporting evidence from directly genotyped variants in all three studies. Additional replication is in progress. The top variants of these three signals were found in introns of *KIF15* (rs78238620, GWAS $p = 8.85 \times 10^{-10}$, OR=1.57 (95% CI [1.36, 1.82]), eQTL for *KIF15* in brain tissue), *MAD1L1* (rs12699415, GWAS $p = 2.95 \times 10^{-13}$, OR=0.78 (95% CI [0.73, 0.83]), eQTL for *MAD1L1* in heart tissue) and *DEPTOR* (rs6469878, GWAS $p = 5.20 \times 10^{-10}$, OR=0.81 (95% CI [0.76, 0.87]), eQTL for *DEPTOR* in multiple tissues including lung). In addition, low frequency variants in *RTEL1* (a telomere regulating gene previously implicated in familial pulmonary fibrosis), and rare variants in *HECTD2* and at 5p14.1 were identified and require further investigation. This is the largest GWAS of IPF performed to date and provides new insight into disease risk.

3314F

Gene prioritization in kidney function associated loci through systematic colocalization with gene expression. K.B. Sieber, CKDGen Consortium. Target Sciences, GlaxoSmithKline, Collegeville, PA.

Genome-wide association studies (GWAS) have successfully identified genetic loci associated with complex traits; however, identification of effector genes is an ongoing challenge. One promising method to prioritize effector genes is to perform colocalization of GWAS and expression quantitative trait loci (eQTL) data to identify shared genetic variants that implicate effector genes. We applied GWAS-eQTL colocalization analysis to GWAS meta-analyses of estimated glomerular filtration rate (eGFR, n=560K) and urinary albumin-to-creatinine ratio (UACR, n=540K) based on European ancestry individuals within the CKDGen Consortium. For eQTL data, we used all GTEx v6p tissues and the first publicly available kidney eQTLs (NephQTLs) with glomerular and tubular data derived from nephrotic syndrome patients. In addition, we introduce a novel method to calculate the direction of effect across the eQTL and GWAS results using a posterior probability weighted average of the effect ratios. Colocalization testing identified genetic evidence for 367 genes with >80% posterior probability of colocalization across 314 GWAS loci. Across the 314 GWAS loci, we identified at least one colocalized gene in 55% (140/253) of the eGFR loci and 52% (32/61) of the UACR loci. Of the GWAS loci with at least one colocalized gene, 48% (67/140) of the eGFR loci and 38% (12/32) of UACR loci had a single effector gene. Focusing on the positive kidney tissue eQTL colocalizations, UACR had six genes colocalizing in the kidney tissues, of which two were unique to the NephQTL dataset. In contrast, fourteen of the twenty eGFR genes that colocalized within the kidney tissue were unique to the NephQTLs. For example, an eGFR locus colocalizes with the *KNG1* eQTL exclusively in the NephQTL tubular data. Along with the systematic colocalization, we evaluated the effect direction to assess consistency across tissues. For example, *MTTL10* colocalized with an eGFR locus in 34 tissues, all of which support that higher expression of *MTTL10* is associated with higher eGFR. In contrast, *SH3YL1* colocalized with an eGFR locus in 24 tissues; however, only 13 of the 24 positive colocalization tissue eQTLs support that higher *SH3YL1* expression associates with higher eGFR. Together, our findings support the utility of using a large data source such as GTEx for colocalization testing, but also emphasize the need for additional tissue-specific open access eQTLs datasets such as the NephQTL data.

3315W

Using human genetics to identify a novel biomarker for idiopathic pulmonary fibrosis. A. Cerani^{1,2}, S. Ross^{1,2}, D. Assayag³, D.A. Schwartz³, M. Lathrop⁴, P. Wolters⁴, J.B. Richards^{1,2,5,6}. 1) Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, QC, Canada; 2) Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, QC, Canada; 3) Department of Medicine, University of Colorado, Denver, CO; 4) Department of Medicine, University of California, San Francisco, San Francisco, CA; 5) Department of Human Genetics, McGill University, Montreal, QC, Canada; 6) Department of Medicine, McGill University, Montreal, Canada; 7) McGill University and Genome Quebec Innovation Centre, Montreal, QC, Canada.

Introduction: Idiopathic pulmonary fibrosis (IPF) is a lethal disease with limited effective treatments. There is, therefore, an urgent need to find and validate biomarkers for diagnosis, prognosis, as to serve as potential drug targets. We undertook combined genetic epidemiology and metabolomics studies to identify a metabolic biomarker for idiopathic pulmonary fibrosis. **Methods and Results:** To identify potential metabolites associated with idiopathic pulmonary fibrosis, we combined GWAS for IPF and blood metabolites and identified a shared genetic determinant at isovaleryl dehydrogenase (*IVD*). The product of this gene is an enzyme whose inhibition is known to increase its substrate, the metabolite isovalerylcarnitine (IVC). Through two-sample Mendelian randomization (MR), we observed that genetically decreased IVC levels were strongly associated with increased risk of IPF ($p = 9.8 \times 10^{-4}$). We confirmed this result by using another blood metabolomics GWAS in a second MR study ($p = 5.5 \times 10^{-7}$). We next found that the non-coding alleles from the two metabolite GWAS that were associated to decreased IVC blood levels increased *IVD* expression ($p = 2.10 \times 10^{-20}$). We therefore tested whether low IVC levels were associated with IPF risk in a study of 382 cases and 217 control. We found that decreased IVC blood concentration was associated with a 26% increase in IPF risk (OR: 1.26 per 50 nM decrease; 95% CI: 1.00 -1.60, $p = 0.048$). **Conclusion:** IPF presents important unmet clinical needs. Building on strong orthogonal evidence from genetics and metabolomics data, we found evidence that low IVC levels were associated with higher risk of IPF. IVC represents a clinically relevant biomarker and its enzyme, *IVD*, could represent an entirely novel IPF drug target.

3316T

Towards precision medicine management of vitamin D inadequacy. K.E. Hatchell, C.D. Engelman. Department of Population Health Sciences, University of Wisconsin-Madison, MADISON, WI.

Vitamin D inadequacy affects more than 50% of adults in the United States and is associated with numerous adverse health outcomes. Vitamin D concentrations are influenced by genetic factors which may determine how much vitamin D intake is required to reach optimal vitamin D blood concentration (25(OH)D). Therefore, knowledge of the genetic factors of 25(OH)D could be useful for the prevention of vitamin D associated morbidity and mortality. To date, only a small portion of the genetic factors of 25(OH)D is understood. Understanding the complete set of genetic factors that contribute to 25(OH)D could allow for personalized treatment of vitamin D inadequacy. Participants (European ancestry $n=9,569$, African ancestry $n=2,761$) come from dbGaP studies: ARIC, MESA and WHI. In an ancestry specific manner, using PRSice and published GWAS summary statistics, polygenic risk scores (PRSs) were created to capture a more complete set of genetic factors. PRSs were iteratively calculated, using p-value thresholds from 5×10^{-5} to 0.5, incrementing the threshold by 5×10^{-5} each iteration. SNPs were clumped ($r^2 < 0.5$) for independence in the PRS. The relationship between 25(OH)D and PRS was tested, selecting the PRS with the highest coefficient of determination (R^2) as the most optimal. Additionally, using ancestry-specific estimates from GCTA, heritability of the PRS was quantified and compared to SNP heritability and heritability of published GWAS findings. The most optimal PRS for African ancestry was derived using a p-value cut-off of 0.013, includes 32,269 SNPs and has an R^2 of 4.4%; for European ancestry, the PRS was derived using a p-value cut-off of 3.5×10^{-4} , includes 341 SNPs and has an R^2 of 1.4%. Results show that the PRS is associated with 25(OH)D in training and validation cohorts. For example, in the validation cohort, those with lowest genetic risk had 25(OH)D concentrations 2.7 ng/mL higher than those of median risk ($p < 0.0001$) in European ancestry, and concentrations 1 ng/mL higher (NS) in African ancestry. Additionally, where sample size allowed for analysis (European ancestry), the PRS explains more of the heritability of 25(OH)D than do published GWAS findings (3.7% vs 1.5%). However, the PRS only explains 17.1% of the total SNP heritability (21.6%). Given that much of the heritability remains unexplained, further genetic exploration of 25(OH)D concentrations is needed to leverage genetic information for personalized vitamin D dosing.

3317F

A weighted genetic risk score of 279 signals discovered in individuals of European descent is associated with impaired lung function and chronic obstructive pulmonary disease in multiple ancestries. A.L. Guyatt¹, N.

Shrine¹, R. Boxall^{2,3}, P. Sakornsakolpat^{4,5}, K. Lin², D. Prokopenko⁴, M.M. Parker⁶, C.A. Melbourne¹, K.A. Fawcett¹, A.M. Erzurumluoglu¹, V.E. Jackson^{1,6,7}, L. Li⁸, D.P. Strachan⁹, A.P. Morris^{10,11}, Z. Chen², E.K. Silverman^{4,12}, R.G. Walters², B.D. Hobbs^{4,12}, M.H. Cho^{4,12}, I.P. Hall¹³, M.D. Tobin^{1,14}, L.V. Wain^{1,14}, SpiroMeta Consortium. 1) Department of Health Sciences, University of Leicester, Leicestershire, United Kingdom; 2) Nuffield Department of Population Health, University of Oxford, United Kingdom; 3) Medical Research Council Population Health Research Unit, University of Oxford, United Kingdom; 4) Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, Massachusetts, United States of America; 5) Department of Medicine, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand; 6) Population Health and Immunity Division, The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, Australia; 7) Department of Medical Biology, University of Melbourne, Parkville, Victoria, Australia; 8) Department of Epidemiology & Biostatistics, Peking University Health Science Centre, Peking University, Beijing, China; 9) Population Health Research Institute, St George's, University of London, London, United Kingdom; 10) Department of Biostatistics, University of Liverpool, Liverpool, United Kingdom; 11) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 12) Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, Massachusetts, United States of America; 13) Division of Respiratory Medicine and National Institute for Health Research – Nottingham Biomedical Research Centre, University of Nottingham, United Kingdom; 14) National Institute for Health Research Biomedical Research Centre, Glenfield Hospital, Leicester, United Kingdom.

Introduction: Chronic obstructive pulmonary disease (COPD) is a condition with high global morbidity. It is characterised by reduced lung function, quantified by the ratio of the forced expiratory volume in 1 second to forced vital capacity (FEV₁/FVC). We have shown that constructing genetic risk scores for reduced FEV₁/FVC (as an endophenotype for COPD) is a well-powered method of predicting COPD in people of European descent. However to date, scores defined in European-ancestry individuals have not shown strong associations with lung function in other ancestries. Here, we construct a new, *weighted* genetic risk score of FEV₁/FVC from 279 signals (139 novel), and test its association with FEV₁/FVC and COPD in multiple ancestries from the UK and China Kadoorie Biobanks, and the COPDgene study. **Methods:** The European-derived weighted genetic risk score comprised 140 known lung function signals, and 139 novel signals discovered in over 400,000 individuals from UK Biobank and the SpiroMeta consortium. Unbiased weights were derived from the largest cohort of UK Biobank or SpiroMeta not included in signal discovery. The score was tested for association with FEV₁/FVC in UK Biobank participants of South Asian, Chinese and African ancestry, and the China Kadoorie Biobank. Associations with COPD (defined as FEV₁/FVC < 0.7, FEV₁ < 80% predicted) were tested in China Kadoorie Biobank, African American participants of COPDgene, and UK Biobank ancestries with >100 COPD cases. Associations were adjusted for age, age², sex, height, smoking status, and population structure. **Results:** The weighted risk score was associated with FEV₁/FVC in all ancestries, with a 1-SD increase in score corresponding to a decrease in ~0.1-0.2 SDs in FEV₁/FVC. In non-European ancestry participants, effect sizes were largest in 6,358 UK Biobank individuals of South Asian descent ($\beta = -0.18$ [95%CI -0.21, -0.16] $p = 3.7 \times 10^{-47}$), and smallest in 4,225 UK Biobank individuals of African descent ($\beta = -0.09$ [95%CI -0.12, -0.06] $p = 2.1 \times 10^{-9}$), and 72,796 individuals within China Kadoorie Biobank ($\beta = -0.08$ [95%CI -0.09, -0.07] $p = 5.1 \times 10^{-99}$). The score predicted COPD in all ancestries tested. **Conclusion:** For the first time, we show that a weighted genetic risk score constructed from signals discovered in European individuals is predictive of lung function and COPD across global ancestries. This work provides evidence of shared genetic architecture of COPD across ancestries.

3318W

Defining phenotypes with non-ignorable missingness on the basis of heritability: Application to lung disease in cystic fibrosis. S.O. Kim^{1,4}, B. Xiao², N. Panjwani¹, G. He³, J. Gong¹, E. Corfield², L.J. Strug^{1,2,4}. 1) Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada; 2) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, ON, Canada; 3) Department of Psychiatry, The Hospital for Sick Children, Toronto, ON, Canada; 4) Biostatistics Division, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada.

Heritability is the proportion of phenotypic variance attributable to genetic variation. Although heritability provides little information on the genetic architecture, it offers a meaningful comparison of traits within a population and can help define phenotypes to improve gene-mapping studies. Quantitative measurements are often used as an indirect measure of a trait, such as forced expiratory volume in 1 sec (FEV1) as a measure of lung disease in cystic fibrosis (CF). Since FEV1 changes with factors including age, sex and height, it needs to be standardized to population norms (CF-specific reference equations). This standardization reduces the phenotypic variation due to environmental factors and results in a more heritable trait. However, the progressive mortality due to lung disease in CF results in cohort and sex-specific non-ignorable missingness and left truncation; this mortality limits the utility of standardized FEV1 phenotypes in genetic studies. Here, we propose several comprehensive, alternative model-based solutions. To facilitate modelling choices of phenotypes, we use heritability comparisons. We propose the use of non-parametric quantile regression with pattern-mixture modeling (PMM) for a variety of imputation choices that directly incorporates both mortality and CF-specific reference equations into a single phenotype calculation. We used SNP-based heritability by Yang *et al.* (2010) to discriminate between the imputation choices including complete, neighboring and available case missing values in PMM. To estimate the heritability, we included unrelated individuals aged between 6 and 40 years from the Canadian CF Gene Modifier Study (n=1,233). We used a parametric bootstrapping approach by Schweiger *et al.* (2016) to construct 95% confidence intervals (CI) and to test the difference in heritability between phenotypes. The highest estimated heritability across phenotypic modeling choices was $h^2=0.446$ (CI: 0 - 0.632). It was more than 2-fold greater than the heritability estimate ignoring the missingness ($h^2=0.207$; CI: 0 - 0.367) and an improvement of 11% over previously defined phenotypes used by the International CF Gene Modifier Consortium (Taylor *et al.*, 2011) ($h^2=0.401$; CI: 0 - 0.575). Tests of the differences, however, did not reach statistical significance. Phenotypes defined using quantile regression and PMM would be useful for other progressive diseases with mortality, and heritability comparisons facilitate modeling choices.

3319T

Copy number polymorphism on Y chromosome and adult height: The gr/gr deletion does not contain the Y-specific growth gene. A. Hattori^{1,2}, K. Saito¹, M. Fukami¹. 1) Department of Molecular Endocrinology, National Research Institute for Child Health and Development, Tokyo, Japan; 2) Department of Advanced Pediatric Medicine, Tohoku University School of Medicine, Tokyo, Japan.

In the 1990s, Ogata *et al.* proposed a hypothesis that the Y-specific growth gene, which is termed GCY (growth control, Y chromosome influenced), accounts for about two-thirds of the sexual difference in human adult height. Thus far, the genomic position of GCY has been investigated by case-based phenotype-genotype correlation studies. However, it is difficult to set an appropriate adult height cut-off value because GCY abnormality does not always cause an adult height below -2.0 standard deviations. Here, we performed a group-based statistical analysis focusing on the 1.6 Mb gr/gr deletion, which is a common polymorphism in Japanese populations. A total of 315 healthy Japanese men were enrolled in this study. We performed PCR amplification analysis of two sequence-tagged site markers (sY1291 and sY1191) to examine the presence of the gr/gr region. The participants were classified into two groups with and without gr/gr deletion. Then, we compared the mean adult height of the two groups. The statistical difference was examined by student's *t*-test. Of the 315 men, 127 and 188 were classified into the deletion-positive and -negative groups. There was no significant difference in the mean adult height between the two groups (170.3 cm vs. 170.1 cm, $P = 0.833$). The present study statistically demonstrated for the first time that the gr/gr region does not encompass growth control genes. The group-based study focusing on a common polymorphism enabled effective statistical analysis. We may further narrow candidate regions for GCY by applying this method to rarer recurrent variants.

3320F

Identification of CpG sites as mediators for genetic influence on osteoporosis by Mendelian randomization. *F.T. Yu, C. Qiu, C. Xu, H.W. Deng, H. Shen.* Center for Bioinformatics and Genomics, Department of Global Biostatistics and Data Science, School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA, USA.

Osteoporosis is one of the most serious public health concern that affect more than 200 million people worldwide. It can be characterized by low bone mineral density (BMD), which is primarily caused by excessive bone resorption by osteoclasts. Peripheral blood monocytes (PBMs) serve as an important source of precursors of osteoclasts and produce cytokines affecting the differentiation, activation, and apoptosis of osteoclast. Due to the biological relevance of PBM with bone metabolism and disorders, it has been determined as a working cell model for genomic and epigenomic studies of osteoporosis and other-bone related disease. Recent studies have assessed the cell-type-specific correlation between diverse genetic, epigenetic and transcriptomic factors of PBMs and evaluated the impact of these factors on immune disease. However, the higher-resolution epigenetic profile of PBMs measured by Illumina Human Methylation 450K array and the mediation effects of DNA methylation on the pathway of genetic factors to disease has not been examined. In this study, we performed a comprehensive analysis of DNA methylation data measured from PBMs of 105 Caucasian women with extreme BMD phenotype. By testing the association between each pair of SNP and CpG, we identified a large amount of methylation quantitative loci (mQTLs). For the mQTLs that show significance in GWAS, we perform Mendelian Randomization (MR) analysis and reverse MR to further evaluate the mediation role of the corresponding CpGs in the causal pathway of SNP and BMD. A novel SNP (rs12811331) has been identified as a mediator of the genetic effect on BMD. We will also check the colocalization between the mQTL signals and the GWAS signals of BMD trait, which would allow us to detect more precisely the common causal variant of DNA methylation and BMD trait. This work is supported by grants from the NIH (R01AR059781, P20GM109036) and Edward G. Schlieder Endowment as well as the Drs. W. C. Tsai and P. T. Kung Professorship in Biostatistics from Tulane University.

3321W

A genome-wide association study of bone mineral content and bone mineral density of multiple skeletal sites in Hispanic children. *R. Hou¹, S.A Cole², K. Haack³, S. Laston³, N.R Mehta⁴, R. Shypailo⁴, A.G Comuzzie⁵, N.F Butte⁴, V.S. Voruganti¹.* 1) Department of Nutrition and Nutrition Research Institute, University of North Carolina at Chapel Hill, Kannapolis, NC; 2) Texas Biomedical Research Institute, San Antonio, TX; 3) South Texas Diabetes and Obesity Institute, University of Texas Rio Grande Valley, Brownsville, TX; 4) Pediatrics and USDA/ARS Children Nutrition Research Center, Baylor College of Medicine, Houston, TX; 5) The Obesity Society, Silver Spring, MD.

Background: Peak bone mass attained in early life is a significant predictor of bone mass in later life and is known to be highly heritable. **Aim:** To identify genetic variants affecting bone mineral content (BMC) and bone mineral density (BMD) of multiple skeletal sites in Hispanic Children from the Viva La Familia Study. **Methods:** We performed a genome-wide association study of bone outcomes with 1.1 million single nucleotide polymorphisms (SNPs) genotyped using the Illumina Human Omni-Quad v1.0 BeadChip marker panel in 815 Hispanic children. BMC and BMD were determined by dual-energy X-ray absorptiometry with a Hologic Delphi-A whole-body scanner (Version 11.2). Measured genotype analysis was performed between SNPs and BMC and BMD of multiple skeletal sites after accounting for kinships. Empirical thresholds for genome-wide significant and suggestive associations were defined as $p < 1 \times 10^{-7}$, and $p < 1 \times 10^{-6}$, respectively. **Results:** The heritabilities for BMC and BMD across different skeletal sites ranged from 42-76% ($p < 3.1 \times 10^{-9}$). Evidence for statistical significant association ($p < 1 \times 10^{-7}$) was observed for several SNPs with BMC, including BMC of right arm with rs762920 in *PVALB* ($p = 4.6 \times 10^{-8}$), BMC of right leg with rs4871750 in *LOC105375751* ($p = 7.3 \times 10^{-8}$) and with rs142813632 in *PVT1* ($p = 7.6 \times 10^{-8}$) and association with BMD pelvis was observed for rs7000615 in *PTK2B* ($p = 7.4 \times 10^{-8}$). Suggestive evidence ($p < 1 \times 10^{-6}$) for association was found for BMC and BMD on different skeletal sites with genes relevant to bone metabolism (*USP12*, *RSU1*, *DGKB*, *ROBO2*, *IGF2R* and *PRLR*) and novel genes with unknown relationship with bone metabolism (*LRIF1*, *METTL10*, *STMN4*, *DSCAM*, *VSIG10*, *LOC105378516* and *ZDHHC1*). **Conclusion:** These findings shed light on the genetic structure underlying BMC and BMD variation and osteoporosis susceptibility. .

3322T

Identification of putative causal genes regulating hip bone mineral density using tree-based methods. X. Meng¹, L. Tan¹, H. Xiao^{1,2,3}, H. Deng³.

1) Hunan Normal University, Changsha, China; 2) Tulane University, New Orleans, USA; 3) Central South University, Changsha, China.

Objective: Osteoporosis is a complex disease characterized by low bone mineral density (BMD) and an increased susceptibility to fractures. Several studies have identified dozens of genes associated with BMD from gene expression profiles using the linear regression model. However, this approach cannot distinguish those genes whose expression levels have a non-linear correlation with BMD or determine whether these genes causally regulate BMD. **Method:** We analyzed gene expression profiles from bone biopsies in 84 postmenopausal women using a tree-based method known as Gene Network Inference with Ensemble of trees (GENIE3). The inferred regulation network was used to identify putative genes which may regulate hip BMD. Enrichment analysis was performed to identify transcription factors whose binding sites are significantly enriched in femoral neck BMD (FN-BMD) genome-wide association studies (GWAS) signals. **Results:** In the gene regulatory network we successfully identified six genes which were previously identified using the standard linear regression model, and another 14 novel predictive genes. Among 161 transcription factors, FN-BMD GWAS signals were significantly enriched within JUN binding sites (fold enrichment = 5.45, p value = 1.51×10^{-4}). Three predictive genes (*PLIN5*, *CLIP4* and *LINC00312*) were significantly enriched in JUN regulation network (fold enrichment = 4.44, p value = 0.045). **Conclusion:** We identified three putative causal genes and one transcription factor that may be involved in the regulation of FN-BMD. The transcription factor JUN interacts with NFAT and is crucial for RANKL-regulated osteoclast differentiation. Our results suggest that while the traditional linear regression model is useful to identify some trait associated genes, the tree-based method may identify more putative causal genes. These novel insights into the transcription factor regulation network are helpful to understand the biological mechanisms of these predictive genes regulating BMD.

3323F

HLA-DQB1*03:01 as a biomarker for genetic susceptibility to bullous pemphigoid induced by DPP-4 inhibitors. T. Ozeki¹, H. Ujii², K. Muramatsu², T. Mushiroda¹, H. Miyoshi³, H. Iwata², T. Ito², K. Izumi², W. Nishie², H. Shimizu². 1) Laboratory for Pharmacogenomics, RIKEN Center for Integrative Medical Sciences, Yokohama, Kanagawa, Japan; 2) Department of Dermatology, Hokkaido University Graduate School of Medicine, Sapporo, Japan; 3) Department of Rheumatology, Endocrinology and Nephrology, Faculty of Medicine, Graduate School of Medicine, Hokkaido University, Sapporo, Japan.

Dipeptidyl peptidase-4 inhibitor (DPP-4i) has been widely used to treat type 2 diabetes. DPP-4i has been known as a safe drug; however, an increased risk of bullous pemphigoid (BP), which is an autoimmune subepidermal blistering disease, during DPP-4i exposure has been reported in diabetic patients administered with DPP-4i. Although the exact mechanism behind the association of DPP-4i exposure and BP has not been elucidated, several studies have reported an association between human leukocyte antigens (HLA) and drug-induced reactions. Thus, we examined HLA genotypes in Japanese patients with BP who had been taking DPP-4i for type 2 diabetes for at least 3 months before the BP onset (DPP-4i-BP). It was reported that DPP-4i-BP tends to show a "noninflammatory" phenotype with few erythematous lesions, in sharp contrast to conventional BP unrelated to DPP-4i intake. We encountered 30 patients with DPP-4i-BP in the last 3 years and found that the majority of cases (21/30) showed the noninflammatory phenotype. We compared the 6 HLA alleles in noninflammatory DPP-4i-BP with those in Japanese general population controls or in DPP-4i-tolerant patients with type 2 diabetes and found that 86% (18/21) of noninflammatory DPP-4i-BP patients, 18% (156/873) of general population controls and 31% (19/61) of the tolerant patients carry *HLA-DQB1*03:01*. The frequency was significantly higher in DPP-4i-BP patients than in general population controls ($P = 5.86 \times 10^{-11}$; OR = 27.6, 95% CI = 8.0-94.8) and in the tolerant patients ($P = 2.13 \times 10^{-4}$; OR = 13.3, 95% CI = 3.5-50.5). This work was supported by JSPS Core-to-Core Program, A. Advanced Research Networks.

3324W

ABCA7 and ancestry-aware association study for Alzheimer disease in Puerto Ricans. K. Celis¹, F. Rajabli¹, B. Feliciano-Astacio², K.L. Hamilton-Nelson¹, L.D. Adams¹, V.C. Rodriguez¹, A. Rodriguez Betancourt², H. Acosta³, A. Chinea⁴, T. Thornton⁵, H.N. Cukier⁶, G.S. Bird⁶, J.M. Vance^{1,7}, M.L. Cuccaro¹, G.W. Beecham¹, M.A. Pericak-Vance^{1,7}. 1) John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL, USA; 2) Universidad Central del Caribe, Caguas, PR, USA; 3) Clinica de la Memoria, San Juan, PR, USA; 4) Universidad Central del Caribe, Bayamon, PR, USA; 5) University of Washington, Department of Biostatistics, WA, USA; 6) Department of Biology, North Carolina A & T State University, Greensboro, NC, USA; 7) Dr. John T. Macdonald Department of Human Genetics, University of Miami, Miami, FL, USA.

Background: Caribbean Hispanic (CHI) populations have a higher prevalence and incidence of late-onset Alzheimer disease (LOAD) than non-Hispanic white (NHW) individuals. However, CHI populations are understudied compared to NHW. Moreover, the ancestral heterogeneity of CHI makes studies of this population critical in discovery ancestry-specific genetic factors in AD. To identify novel genetic variation influencing AD, to test generalization of the NHW findings and to better understanding AD in the CHI we performed genome-wide association analysis in CHI Puerto Rican (PR) population. **Methods:** Through PR Alzheimer Disease Initiative we have ascertained a family-based dataset consisting of 412 individuals of PR ancestry. The dataset includes 231 AD cases and 181 cognitively intact individuals matched for age and sex. Genome-wide genotyping was performed on two platforms: MEGA and GSA. Single variant association analysis was performed by using the GWAf geepack package in R. Logistic model was adjusted for principal components, age, and sex. We assessed the ancestral background of the candidate genes ("local" ancestry) by using RFMix. Variants were annotated with Ensemble VEP annotation. A gene-wise adjusted *p*-value was calculated to correct for multiple markers within the gene. **Results:** Our results showed that 15 GWAS candidate AD loci significantly associated with AD in PR population, (e.g. ABCA7, BIN1). The top SNP among the candidate genes was observed in the exonic region of the ABCA7 gene (rs3752240, *p*=2e-4). By stratifying the dataset according to the local ancestral backgrounds, we observed that the European and African genetic ancestry contributed risk in ABCA7 gene. We also found that APOE ε4 risk allele was associated with the LOAD (OR=1.96, *p*=6e-4), but the effect size was much less than in NHW. By using the local ancestry we found that African ancestral background mitigate the effect of ε4 allele. Genome-wide the most significant association was observed in the SYNDIG1 gene (rs2892216, *p*=9.1e06). SYNDIG1 gene plays a critical role during synapse development, and associated with a mental and behavioral disorder. **Conclusions:** Our findings support the hypothesis that there may be an interaction between the genetic factors and the ancestral backgrounds located around the genomic region of interest. Results on PRs demonstrated the risk in ABCA7 was contributed by AF and EU ancestry. Further, we will apply genome-wide ancestry-aware approach.

3325T

Development of a polygenic risk score for Alzheimer's disease using a Bayesian hierarchical model. S.P. Dickson¹, S.B. Hendrix¹, B.L. Brown², P.G. Ridge², S.B. Booth², R.R. Fortna³, J.S.K. Kauwe², NIAGADS. 1) Pentara, 2261 E. 3300 S. Millcreek, UT; 2) Brigham Young University, Provo, UT; 3) Affirmativ Group 3614 Meridian Street, Suite 100 Bellingham, WA.

Our knowledge of the genes associated with Alzheimer's disease has grown steadily over the past three decades. While the effect of the APOE gene still remains the single largest genetic risk factor, dozens of other genes have been implicated over the years. Summary results from published studies based on large, well-curated datasets were combined with data from multiple genetic studies to incorporate the most important genetic risk factors identified to date with age, sex and their interaction into a single polygenic risk score. This combined information represents the best information available about the genetic risk of Alzheimer's disease. We explored a range of techniques for combining past results, including additive and multiplicative risk models, regularization techniques like LASSO and elastic nets, as well as Bayesian models, and use cross validation to select the best performing method. A Bayesian hierarchical model is reported and used to demonstrate the calculation of the unconditional lifetime risk of developing Alzheimer's disease, as well as the lifetime risk of developing Alzheimer's disease conditional on the current absence of disease.

3326F

Shared causal paths underlying Alzheimer's disease and type 2 diabetes.

Z. Hu¹, R. Jiao⁴, Y. Zhu², J. Zhao³, D. Bennett², L. Jin¹, M. Xiong¹. 1) School of Life Science, Fudan University, Shanghai, China; 2) Rush Alzheimer's Disease Center, Rush University Medical Center; 3) Department of Epidemiology, University of Florida; 4) University of Texas School of Public Health.

Although Alzheimer's disease (AD) is a central nervous system disease and type 2 diabetes MELLITUS (T2DM) is a metabolic disorder, an increasing number of epidemiological and genetic epidemiological studies show clear link between AD and T2DM. AD with great economic, political and social consequences is a progressive, irreversible degenerative disease of the brain and is the most common cause of dementia due to the gradual accumulation of accumulation of amyloid-beta (A β) and twisting of tau protein. AD is also involved in inflammation and oxidative stress and exhibits memory loss and cognitive dysfunction. The current approaches to identifying several shared pathophysiology processes between AD and T2DM have several limitations. Firstly, the pathophysiological links between AD and T2DM have been assessed mostly through literature review. Secondly, the most previous works have focused on identifying biological pathways underlying AD and T2DM. Few attempts to discover the role of dysregulated SNPs, gene expressions and methylations have been carried out. Thirdly, the conventional evidences for linking AD and T2DM purely depend on the statistical association. It is time to develop a new generation of genetic analysis for shifting the current paradigm of genetic analysis from shallow association analysis to deep causal inference and from genetic analysis alone to integrated causal genomic, epigenomic, and phenotypic data analysis for unraveling the mechanistic link between AD and T2DM. We develop novel causal inference methods for genetic studies of AD and T2DM and unified frameworks for systematic causal analysis of integrated genomic, epigenomic, and clinical phenotype data and to infer multilevel omics causal networks for the discovery of common paths from genetic variants to AD and T2DM via methylations, gene expressions and multiple phenotypes.

3327W

Constrained instruments and their application to Mendelian randomization with pleiotropy.

L. Jiang^{1,2}, K. Oualkacha³, D. Didelez⁴, A. Ciampi^{1,2}, P. Rosa^{5,6}, A. Benedet⁶, S. Mathotaarachchi⁶, B. Richards⁷, C. Greenwood^{1,2}. 1) Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Quebec, Canada; 2) Lady Davis Institute, Jewish General Hospital, Montreal, QC, Canada; 3) University of Quebec at Montreal, Montreal, Quebec, Canada; 4) Leibniz Institute for Prevention Research and Epidemiology, University of Bremen, Germany; 5) Department of Neurology and Neurosurgery, McGill University, Montreal, Quebec, Canada; 6) Translational Neuroimaging Laboratory, McGill University Research Centre for Studies in Aging, Douglas Hospital, McGill University, Montreal, Quebec, Canada; 7) Department of Medicine, McGill University, Montreal, Quebec, Canada.

Introduction: In Mendelian randomization (MR), genetic variants are used to construct instrumental variables, which enable inference about the causal effect of a phenotype of interest on a response or disease outcome. Conventional estimation methods depend on a set of prior assumptions regarding the validity of instrumental variables, such as the genetic variants only influence the response through the phenotype of interest. In practice, these assumptions are violated when pleiotropy occurs, in which genetic variants influence the response through multiple phenotypes and render conventional MR analysis invalid. **Methods:** We present a set of intuitive methods (Constrained Instrumental Variable methods [CIV]) to construct valid instrumental variables and perform adjusted causal effect estimation when pleiotropy exists, given that pleiotropic phenotypes have been measured. One of the methods in this set, CIV_smooth, implement an automatic and valid selection of genetic variants when building the instrumental variables. Despite the fact that the variable selection problem is NP-hard, we implement an approximate solution that is computationally efficient. **Results:** The performance of the proposed methods, compared to competitor methods, is demonstrated on data simulated under different types of pleiotropy. We also analyzed data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) Mueller et al. (2005) to disentangle causal relationships of several biomarkers with AD progression. A list of valid genetic instrument for each of these biomarkers of interest are identified, and adjusted estimation of their causal effect on AD progression are implemented. CIV_smooth found only one significant causal effect, that of Amyloid beta 1-42 on AD progression; this indicates that the previously known association of this biomarker with AD progression may be causal. However, due to the limitation of the ADNI study design, additional analysis are still needed to validate our results.

3328T

Searching for the causal effects of genetic variants for Alzheimer's disease in UK Biobank. R. Korologou-Linden¹, L.A.C. Millard^{1,2}, W. Spiller¹, G. Davey-Smith¹, L.D. Howe¹, E.L. Anderson¹, E. Stergiakouli^{1,3}. 1) MRC Integrative Epidemiology Unit (IEU), Bristol Medical School (Population Health Sciences), University of Bristol, Bristol, UK; 2) Intelligent Systems Laboratory, Department of Computer Science, University of Bristol, Bristol, UK; 3) School of Oral and Dental Sciences, University of Bristol, UK.

Introduction: Observational studies for Alzheimer's disease have reported conflicting evidence for most potential modifiable risk factors. This may be due to bias from confounding and reverse causation that can be problematic in observational studies. Furthermore, the hypothesis-driven nature of previous work biases research on Alzheimer's disease to hypotheses with a prior. Alzheimer's disease is highly heritable with estimates as high as 79%. Genetic studies have identified 21 single nucleotide polymorphisms associated with late-onset Alzheimer's disease, all exerting low to modest effects (except for those in the apolipoprotein E gene). We perform a phenome-wide association analysis to investigate the causal effects of a polygenic risk score (PRS) for Alzheimer's disease on a large and diverse array of phenotypes, without a prior and without the biases present in most observational studies. **Methods:** We used a sample of 334,968 participants from UK Biobank, and generated a PRS calculated as the weighted sum of the risk-increasing alleles identified in a genome-wide association study of Alzheimer's disease. We performed a phenome-wide association analysis of this PRS on 21,868 traits, using the PHENome Scan Analysis tool (PHESANT). We also examined whether findings varied by age categories (39 to 53, 53 to 62 and 62 to 72 years). **Results:** We identified associations of the PRS for Alzheimer's disease with family history of dementia, own diagnosis of dementia, death from dementia, and brain related-measures (e.g. lower mean intra-cellular volume in parahippocampal part of cingulum, lower cognitive function). Other factors associated with the PRS included lifestyle (e.g. more days of moderate/vigorous activity), biological measures (e.g. lower monocyte percentage), higher risk of diseases (e.g. atherosclerotic disease, delirium), and physical measures (e.g. lower body mass index). The associations observed for the entire sample appear to be driven by the older age groups. For the youngest sample with mean age 47 years (39-53), we only identified associations with biological measures (e.g. reticulocyte percentage), hypercholesterolaemia and family history of dementia. **Conclusion:** The novel factors identified by our phenome-scan may either lie on the causal pathway from genetic predisposition to Alzheimer's disease or may be a consequence of a higher genetic risk for Alzheimer's disease. These novel associations should be replicated in independent cohorts.

3329F

Genome-wide association analysis identifies novel Alzheimer disease candidate risk loci for African-Americans. B.W. Kunkle¹, K.L. Hamilton-Nelson¹, M. Jean-Francois¹, M. Schmidt¹, E.R. Martin¹, L.A. Farrar², J.L. Haines³, G.S. Byrd⁴, G.D. Schellenberg⁵, R. Mayeux⁶, M.A. Pericak-Vance¹, C. Reitz², Alzheimer's Disease Genetics Consortium. 1) University of Miami, Miami, FL, USA; 2) Columbia University, New York, NY, USA; 3) Boston University, Boston, MA, USA; 4) Case Western Reserve University, Cleveland, OH, USA; 5) North Carolina A&T State University, Greensboro, NC, USA; 6) University of Pennsylvania, Philadelphia, PA, USA.

Background: The largest Alzheimer disease (AD) genome-wide association study (GWAS) to date in African-Americans (AA), conducted by the Alzheimer's Disease Genetics Consortium (ADGC) in 2013, identified in addition to *APOE*, two susceptibility loci for Late-onset Alzheimer disease (LOAD), *ABCA7* and an intergenic locus at 5q15. This analysis used the 1000 Genomes (1000G) 2011 panel for imputation, a panel based on haplotypes of 1,094 individuals (246 with African ancestry). Following up these analyses, we conducted a GWAS using two reference panels: the African Genome Resource (AGR) panel, constructed from 4,956 individuals (~2,650 with African ancestry), and the latest 1000G panel, used for indel imputation. We increase our analysis sample by 37 percent (2,845 cases and 5,271 controls). **Methods:** All datasets were imputed to the AGR (~93 million variants) and 1000G panels (~85 million variants) using the Sanger imputation server. Single-variant association analysis was conducted adjusting for age, sex, and principal components (PCs). Additional models adjusted for *APOE*. Individual datasets were analyzed applying logistic regression for case-control datasets and general estimating equations for family-based datasets. Within-study results were meta-analyzed using METAL. **Results:** Following quality control, approximately 35.8 million variants were analyzed. In addition to the previously reported AA risk loci (*APOE*, *ABCA7*, and 5q15), we identified several novel candidate loci ($P < 5 \times 10^{-8}$) for LOAD in AA. These include signals centered at 15q26.1 ($P = 7.89 \times 10^{-7}$), 14q23.3 ($P = 9.75 \times 10^{-7}$), 15q13.3 ($P = 1.11 \times 10^{-6}$), 4p15.32 ($P = 1.72 \times 10^{-6}$), and 6q26 ($P = 2.69 \times 10^{-6}$). Of the 18 other known non-Hispanic white GWAS loci besides *APOE* and *ABCA7*, only *BIN1* showed suggestive association in AA ($P = 0.005$). Gene-based results for these loci are pending however. **Conclusions:** We identified several novel candidate loci for LOAD in AA. These loci contain genes with functions relevant to Alzheimer disease and neurodegeneration including: *CHD2* (highly expressed in microglia) and *RGMA* (synapse formation/accumulated in amyloid plaques) at 15q26.1, *FMN1* (highly expressed in microglia) at 15q13.3, *LDB2* (associated with coronary artery disease) at 4p15.32, and the Parkinson's disease risk gene *PARK2* in 6q26. While overlap between AD risk genes exists between AA and NHW, there are also ethnic specific loci for AA in AD.

3330W

Genome-wide rare variant imputation in 60,178 subjects identifies novel rare variant candidate loci in Late-Onset Alzheimer's Disease (LOAD):

The International Genomics of Alzheimer's Project (IGAP). A.C. Naj^{1,2}, J. Sha¹, G. Leonenko³, Y. Chen⁴, V. Chouraki⁵, M. Leber⁶, R. Sims⁷, B. Grenier-Boley⁸, J.C. Bis⁹, Y. Zhao², A. Kuzma², B. Kunkle⁹, H. Comic¹⁰, S.J. van der Lee¹¹, C. Bellenguez⁵, A. Frizzati¹², P. Holmans⁷, J. Dupuis¹³, L.-S. Wang², L.A. Farrer^{13,14}, C. van Duijn¹⁰, J.L. Haines¹⁵, A.L. DeStefano¹³, M.A. Pericak-Vance^{9,16}, A. Ramirez⁶, P. Amouyel⁶, J.-C. Lambert⁶, S. Seshadri¹⁷, J. Williams⁷, G.D. Schellenberg², *International Genomics of Alzheimer's Project (IGAP).* 1) Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 2) Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 3) Division of Psychological Medicine and Clinical Neurosciences, School of Medicine, Cardiff University, Wales, UK; 4) Graduate School of Arts and Sciences, Boston University, Boston, MA, USA; 5) Epidemiology and Public Health Unit, University of Lille, Lille, France; 6) Department of Psychiatry & Psychotherapy, University of Cologne, Cologne, Germany; 7) Dementia Research Institute, School of Medicine, Cardiff University, Wales, UK; 8) Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA; 9) John P. Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, FL, USA; 10) Department of Epidemiology, Erasmus Medical Center, Rotterdam, the Netherlands; 11) Alzheimer Center, VU University Medical Center, Amsterdam, the Netherlands; 12) Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Wales, UK; 13) Department of Biostatistics, School of Public Health, Boston University, Boston, MA, USA; 14) Department of Neurology, School of Medicine, Boston University, Boston, MA, USA; 15) Department of Population & Quantitative Health Sciences, School of Medicine, Case Western Reserve University, Cleveland, OH, USA; 16) Dr. John T. Macdonald Foundation, Department of Human Genetics, Miller School of Medicine, University of Miami, Miami, FL, USA; 17) Alzheimer's Institute, UT Health San Antonio, San Antonio, TX, USA.

IGAP meta-analyses of genome-wide association studies (GWAS) have previously identified 19 LOAD susceptibility loci in addition to *APOE*, however the majority of these were common (minor allele frequency (MAF)>0.05); only recently have rare variant associations with LOAD such as *TREM2* been observed in these meta-analyses. The Haplotype Reference Consortium (HRC) released a dense reference panel (~65K haplotypes/~39.2M SNPs) allowing imputation of rare variants (MAF>0.0008) for discovery association testing. IGAP consortia imputed 42 GWAS datasets to HRC to identify novel rare variant associations. We imputed 22,847 cases and 36,889 controls to the HRC r1.1 reference panel using Minimac3 on the University of Michigan Imputation Server. Logistic regression on individual variants with MAF>0.01 was performed in PLINKv1.9 (generalized linear mixed model in R for family-based variants) using imputed genotype probabilities, while variants with MAF≤0.01 were analyzed using score-based tests with the program SNPTEST; both analyses adjusted for age, sex, and population substructure. Fixed-effects meta-analysis across datasets was performed in METAL. Gene-based association was also performed using SKAT-O. Preliminary analyses of ~39.2M genotyped or imputed SNVs identified single variant associations of $P<5\times10^{-8}$ in 10 known IGAP LOAD candidate loci (*APOE*, *BIN1*, *MS4A* region, *PICALM*, *CR1*, *CD2AP*, *CLU*, *SPI1*, *SORL1*, and *ABCA7*), including some in more recent IGAP analyses (*LILRA5*, *ADAMTS1*) and multiple suggestive associations ($P<10^{-4}$) at each of an additional 14 loci (including known IGAP loci), rs13155750 in *MEF2C* (OR(95% CI): 1.13(1.08,1.20); $P=5.09\times10^{-6}$) and rs755951 in *PTK2B* (OR 1.11(1.06,1.16); $P=5.61\times10^{-6}$). Novel associations include signals at *LILRA5* (19:54821819; OR 1.14(1.08,1.20); $P=4.84\times10^{-7}$), involved in innate immunity pathways; and at *ADAMTS1* (rs2830500; OR 1.11(1.06,1.17); $P=5.17\times10^{-6}$), which is downregulated in Down Syndrome with AD/neurodegeneration. Rare variant and gene-based analyses demonstrated significant *APOE* $\epsilon 2$ and *TREM2* associations, and a novel association in *STK32C* ($P=6.96\times10^{-7}$) previously associated with cognitive decline, while gene-based testing identified strong but marginal association for *SORL1* ($P=5.55\times10^{-6}$). Several novel candidate loci for LOAD have been identified using high-quality imputation of rare and low-frequency variants in the ADGC, reinforcing the utility of high-density imputation panels in whole genome studies.

3331T

A new powerful co-localization test. W. Pan, Y. Deng. University of Minnesota Twin Cities, Minneapolis, MN.

Recently some integrative methods like transcriptome-wide association study (TWAS) have been proposed to detect genetic association between a gene's expression (or another molecular trait) and a GWAS trait, which may suggest, however does not directly determine, co-localization of GWAS causal variants and eQTL (or other molecular trait) causal variants, in spite of their causal inference frameworks of instrumental variable analysis or Mendelian randomization due to likely violations of their imposed assumptions. Nevertheless, co-localization testing can move statistical inference closer to establishing causal relationships: for example, if it can be established that a causal variant for a GWAS trait and that for an eQTL are the same, then it may suggest a regulatory role of the causal SNP on the GWAS trait through gene expression. Accordingly, there have been extensive efforts in developing and applying various co-localization testing approaches. However, they all have some severe limitations as briefly discussed below. First, the null hypothesis of the popular *coloc* and *HEIDI* specifies a co-localization, which will not be rejected if the sample size is too small and/or the significance level is stringent, clearly a downside. Second, many (e.g. *coloc*, *HEIDI* and *JLIM*) assume the presence of no more than one causal SNP in a locus, contradicting the recently discovered widespread allelic heterogeneity; when the assumption is violated, they may have substantial power loss. Third, although *eCAVIAR* does not impose such an assumption, it requires specifying an upper bound on the number of causal SNPs in a locus, typically at 6, to be computationally feasible. Fourth, some (e.g. *JLIM*) require the availability of individual-level genotypic and gene expression data. Fifth, at least with their current implementations, none of the existing methods can handle more than two traits with dependent samples with only genome-wide association summary statistics for all the traits, which is important to identify pleiotropy and/or for multi-trait analysis. We develop a novel and powerful test overcoming the above shortcomings of the existing methods. We demonstrate its power advantages over the existing methods using both simulated and real data, including a large-scale Alzheimer's disease GWAS and a gene expression datasets, and a large-scale blood lipids GWAS summary association dataset.

3332F

Genome-wide association of plasma tau levels in the Framingham study.

C. Sarnowski¹, C.L. Satizabal^{2,3}, V. Chouraki⁴, M.P. Pase^{5,6}, C. DeCarli⁷, L.A. Cupples^{1,2}, R.S. Vasan^{2,8}, A.S. Beiser^{1,2,3}, A.L. DeStefano^{1,2,3}, S. Seshadri^{2,3,9}. 1) Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA; 2) Boston University and the NHLBI Framingham Heart Study, Boston, MA, USA; 3) Department of Neurology, Boston University School of Medicine, Boston, MA, USA; 4) Lille University, Inserm, CHU Lille, Institut Pasteur de Lille, U1167 - RID-AGE - Risk factors and molecular determinants of aging-related diseases; Lille, France; Labex Distalz; 5) Centre for Human Psychopharmacology, Swinburne University of Technology, Hawthorn, Australia; 6) Melbourne Dementia Research Centre, The Florey Institute for Neuroscience and Mental Health and the University of Melbourne, Australia; 7) Department of Neurology and Center for Neuroscience, University of California at Davis, CA, USA; 8) Preventive Medicine & Epidemiology, Boston University School of Medicine, Boston, MA, USA; 9) Glenn Biggs Institute for Alzheimer's & Neurodegenerative Diseases, University of Texas Health Sciences Center, San Antonio, TX, USA.

Background: Amyloid beta (A β) and tau levels are two biomarkers of Alzheimer Disease (AD). Few genome-wide association studies (GWAS) were conducted on A β or tau levels and these studies were limited by their sample sizes and their inability to explore less common genetic variants. **Methods:** We performed a GWAS of plasma tau levels measured using an ultrasensitive assay in 6,018 Framingham Heart Study participants. We used mixed-effect linear regression models to examine the association between tau levels and imputed genetic variants from the Haplotype Reference Consortium, while accounting for familial relatedness and adjusting for age, sex and principal components. We conducted bivariate analyses to explore potential pleiotropy between tau and three measures of A β levels (A β ₄₂ or A β ₄₀ dosages and A β ₄₂/A β ₄₀) in plasma. Finally, we performed pathway analyses based on results from single-variant association analyses. **Findings:** We detected, at the genome-wide level ($P < 5 \times 10^{-8}$), a new locus on 1p36 (rs34683021, MAF=0.26, $P = 2.8 \times 10^{-9}$) in the intron of *PANK4*, a gene overexpressed in the brain that belongs to a family of proteins associated with neurodegeneration. We also confirmed the association of the 17q21 locus and identified two distinct signals (*MAPT*, rs242557, MAF=0.36, $P = 9.9 \times 10^{-83}$ and *ARL17A*, rs4630592, MAF=0.35, $P = 1.3 \times 10^{-50}$). We found associations ($P \leq 0.05$) in loci associated with tau (CSF p-tau and t-tau and plasma t-tau), AD, and AD-related traits (AD age of onset, accelerated cognitive decline, total ventricular volume, hippocampal atrophy, whole-brain volume and small vessel stroke). Bivariate analyses of tau and A β levels uncovered pleiotropic effects for variants located in the AD gene *C10orf11*, the neurological disorder gene *GOSR2*, the tau and AD gene *GLIS1* and the neuromuscular disease gene *PKD1L2*. Finally, we found enrichment for pathways linked to synaptic function and immunity. **Interpretation:** Our GWAS of plasma tau levels reveals new variants, genes and pathways associated with tau levels and identifies potential pleiotropic effects with A β in brain-related genes. Studying relevant AD biomarkers in genetic analyses can help to uncover new genes and pathways that may contribute to AD-related processes and be promising targets for developing new drugs and preventive medicines for AD.

3333W

African American whole exome sequencing suggests risk coding variants in IDH1 gene.

P. Whitehead¹, F. Rajabli¹, K. Hamilton¹, J.M. Vance¹, E.R. Martin¹, G.D. Schellenberg², L. Wang³, L.A. Farrer³, J.L. Haines⁴, R. Mayeux⁵, G.S. Byrd⁶, G.W. Beecham¹, M.A. Pericak-Vance¹, Alzheimer Disease Genetics Consortium. 1) John P. Hussman Inst for Human Genomics, University of Miami, Miami, FL; 2) Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 3) Departments of Medicine (Biomedical Genetics), Neurology, Ophthalmology, Epidemiology, and Biostatistics, Boston University Schools of Medicine and Public Health, Boston, MA; 4) Inst for Computational Biology, Case Western Reserve University, Cleveland, OH; 5) Department of Neurology, College of Physicians and Surgeons, Columbia University, New York, NY; 6) Department of Biology, North Carolina A & T State University, Greensboro, NC.

African Americans (AA) have a higher prevalence and incidence of AD than non-Hispanic white (NHW) individuals. However, AA populations are understudied, and genomics successes in AD have primarily arisen from studying NHW participants. To identify novel genetic variation influencing AD, and to better understand AD in the AA population we performed whole-exome sequencing (WES) in a set of AA individuals. Our study is the largest to date WES sample for genetic studies of AD in AA individuals. Exome was captured using the Agilent's SureSelect Human All Exon V6 kit. WES was performed on the Illumina's HiSeq 3000 System at average sequencing depth 101. General alignment was performed using BWA algorithm. Genotyping calling was done with GATK and ATLAS calling pipelines. Sequencing was performed on 2,998 unrelated individuals (1,680 cognitively intact controls and 1,318 AD cases). Samples were matched for sex (65.6% vs 73.7%, case vs control) and age (74.8(8.1) vs 74.7(8.7), case vs control). Single variant association and gene-based analysis were performed by using SeqMeta score test analysis and SKAT-O respectively by adjusting for age, sex, PC1 and PC2. Variants were annotated with Ensemble VEP annotation, and were given a high priority if they have a functional impact by a high CADD score. WES confirmed the association of APOE with LOAD (rs741780, p -value=1.2e-55; gene-based test, p -value=1.1e-16). The IDH1 synonymous variant (rs552214533, p -value=5.4e-6) on chromosome 2 did not fully reach genome-wide significance, but was observed as one of the top significant variants. We observed nominal association at 8 variants in AKAP9 gene (rs190545859, p -value=2.6e-3). Similarly, 15 variants in ABCA7 were nominally significant (rs73505232, p -value=6.8e-5). Our results confirm that variants in ABCA7 and AKAP9 genes are the risk factors in AA, and suggest the presence of AD risk variants in IDH1 gene. The IDH1 gene is expressed in brain tissue and related to the immune system pathways, an important pathway in AD. Identification of population-specific variation that influences disease is important to understand the genetic basis of health disparities.

3334T

Correcting for confounding from batch effects and genotype imputation with whole genome sequence data: Application to the ADSP family sample. E.M. Wijsman^{1,2}, T.R.C. Day¹, T.A. Thornton², E.E. Blue¹, *Alzheimer's Disease Sequencing Project*. 1) Division of Medical Genetics/Department of Medicine, University of Washington, Seattle, WA; 2) Department of Biostatistics, University of Washington, Seattle, WA.

Disease-gene identification often employs whole genome sequence (WGS) data. Evaluation of WGS variants may be carried out with association methods used in GWAS studies, including adjustment for related subjects, if necessary. Costs and time required to generate WGS data introduce practical complications. High cost induces incentives to limit the number of subjects with newly-generated WGS. Limiting sample size may include sequencing cases while using previously sequenced shared controls, or sequencing only some subjects, with genotype imputation into the rest. The time needed to generate WGS data in large samples also leads to temporal and/or spatial batch-effects. We show in analysis of the Alzheimer's Disease Sequencing Project (ADSP) family sample that these issues lead to confounding, requiring correction to avoid spurious results in association analyses. We carried out association analysis of Alzheimer's disease (AD) with genotype dose from WGS data in 110 ADSP families with GWAS SNP data from the initial discovery sample, using a linear mixed model in GENESIS, principal components as covariates, and adjustment for relatedness via a kinship matrix. WGS was initially generated for an ~8:1 AD case-control ratio. A second extension round of WGS, providing 854 subjects with WGS in total, emphasized controls in the same families, but still yielded an overall case-control ratio of ~2:1. WGS data and AD status was therefore confounded. Only p-values obtained from the initial WGS conformed to the expected null distribution. Analysis of the joint WGS from both rounds of sequencing lead to a large genomic inflation factor, $\lambda=1.25$, for $-\log_{10}p$. Family-based genotype imputation led to a severe ($\lambda > 3$) inflation factor, which was worse in the presence of sequencing batch effects. Inflation in λ was reasonably controlled by adding several covariates: 1) sequence batch, 2) GWAS status, 3) a measure of imputation quality (imputation deviance) that is a function of family size and proximity to subjects with observed data, and 4) functions of family-based WGS fractions from each batch. Our results provide approaches to adjust for issues in WGS data that are likely to be widespread. They also show that care is needed to capture multiple sources of confounding, both qualitative and quantitative, to avoid spurious conclusions to be drawn from the WGS data.

3335F

Genetically-informed association analysis identifies risk factors for late-onset Alzheimer's disease. D. Yan¹, B. Hu², Y. Wang¹, Q. Lu^{2,3}. 1) University of Wisconsin-Madison, Madison, WI, USA; 2) Department of Statistics, University of Wisconsin-Madison, Madison, WI, USA; 3) Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI, USA.

Genome-wide association studies (GWAS) have provided insights into the genetic basis of numerous complex traits. Increasingly accessible GWAS summary statistics and biobank data have made it possible to further dissect the complex relationship between traits. Here, we introduce GAACT (Genetically-informed Association Analysis for Complex Traits), an efficient method to identify associations between a disease of interest and thousands of genetically-imputed traits using GWAS summary statistics. Compared to traditional association analysis, GAACT does not require traits to be measured on the same individuals and effectively reduces reverse causation from associations. We demonstrated the effectiveness of GAACT through extensive simulations and application to more than 2,000 traits from the UK Biobank (N~337,000) and summary statistics of 30 complex diseases. Our results confirmed many classic causal risk factors including smoking behavior for lung cancer ($p=3.9e-55$). Additionally, we performed detailed analyses to identify risk factors for late-onset Alzheimer's disease (LOAD) using UK biobank traits and GWAS summary statistics from the International Genomics of Alzheimer's Project (IGAP; N=54,162). We further replicated our findings in an independent summary dataset (N=7,050) and results from two stages were meta-analyzed. We identified a total of 58 LOAD-associated traits with Bonferroni-corrected significance and consistent effect directions in two independent analyses. Unsurprisingly, family history of dementia ($p=1.3e-71$ and $2.1e-34$ for illnesses of mother and father, respectively) and cognitive impairment ($p=6.4e-07$) were strongly associated with higher LOAD risk. In addition, various intelligence and education traits, e.g. fluid intelligence score ($p=1.1e-16$) and age completed full time education ($p=6.4e-08$), showed consistent associations with lower risk of LOAD. In contrast, not having any education specified in the survey were associated with a higher risk ($p=3.4e-09$). Further, a few traits with previously mixed results in the literature showed strong and consistent associations with LOAD in our analysis, including high cholesterol ($p=1.8e-14$) and leukemia ($p=4.2e-09$). Finally, we demonstrated that multivariate conditional analysis can be performed in the GAACT framework and suggested a number of traits with independent effects on LOAD risk. GAACT has the potential to be widely used in future studies to screen risk factors for complex diseases.

3336W

A genetic epidemiological study for rare spinocerebellar degeneration based on whole exome sequencing analysis in the Japanese population.

Y. Hama¹, H. Date¹, S. Watanabe¹, H. Ishiura², J. Mitsui², J. Yoshimura³, K. Doi³, S. Morishira³, S. Tsuji², M. Murata¹, H. Mizusawa¹, Y. Takahashi¹. 1) Department of neurology, national center hospital, national center of neurology and psychiatry, Japan; 2) Department of neurology, graduate school of medicine, the university of Tokyo, Japan; 3) Department of computational biology, graduate school of frontier sciences, the university of Tokyo, Chiba, Japan.

Background Hereditary spinocerebellar degeneration (hSCD) is genetically heterogeneous group of disorders with more than 50 causative genes. Common disease types of hSCD include triplet repeat expansion diseases and SCA31 in the Japanese population, which collectively account for ~80% of hSCD. However, genetic epidemiology of the remaining SCD, which is considered to be aggregation of rare disease types, has not been fully delineated. Purpose The purpose of our study was to elucidate the genetic epidemiology of rare SCD in the Japanese population. Methods Seventy subjects diagnosed as SCD were enrolled in this study with informed consent, in whom mutations in triplet repeat expansion diseases, EAOH/AOA1, AOA2, AR-SACS and ataxia telangiectasia were excluded. Whole exome sequencing analysis was conducted and nonsynonymous variants were filtered with the allele frequency <0.002. The present analyses compared age, sex, family history, and clinical findings between patients with mutation and without mutation. This study was approved by the Institutional Review Board of National Center of Neurology and Psychiatry. Results Twenty-seven (39%) subjects harbored likely pathogenic mutations in 16 genes including *AARS2*, *B4GALNT1*, *CACNA1A* (3 subjects), *FGF14*, *GCH1*, *ITPR1*, *KIF1A* (2 subjects), *LYST*, *NPC1*, *PLA2G6* (2 subjects), *SCYL1*, *SETX*, *SPTBN2* (2 subjects), *SYNE1* (5 subjects), *TGM6* (2 subjects), *TTBK2*, and *ZFYVE26*. Nine (33%) genes were causative for diseases not classified as SCD including *AARS*, *B4GALNT1*, *GCH1*, *ITPR1*, *KIF1A*, *LYST*, *NPC1*, *PLA2G6*, and *ZFYVE26*. Pathogenic mutations were more likely to be identified in female patients (female vs. male, 29% vs. 16%, $p=0.046$) or patients with mental retardation ($p=0.048$). There was a tendency towards enrichment of pathogenic mutations in patients with peripheral neuropathy ($p=0.074$). There was no significant difference in age and presence of family history, pure cerebellar symptom, pyramidal sign, parkinsonism, autonomic disturbance, and cognitive dysfunction between the two groups. Conclusion This study demonstrated the advantage of exome-first approach for rare hSCDs, particularly for female patients or the disease subtypes complicated with mental retardation or peripheral neuropathy. Broad categories of genes should be taken into consideration in molecular diagnosis of this entity.

3337T

A third linear association between Olduvai (DUF1220) copy number and the severity of the classic symptoms of inherited autism. J.M. Davis¹, I. Heft¹, S.W. Scherer^{2,3}, J.M. Sikela¹.

1) University of Colorado Denver, Anschutz Medical Campus, Department of Biochemistry and Molecular Genetics, Aurora, CO; 2) McLaughlin Centre and Department of Molecular Genetics, University of Toronto; 3) The Centre for Applied Genomics and Program in Genetics and Genome Biology, Hospital for Sick Children.

Autism is a highly heritable condition which typically involves social deficits, communication impairments and repetitive behavior. We previously have reported, and subsequently confirmed, that the copy number of sequences encoding the Olduvai protein domain (formerly DUF1220) shows a linear association with the severity of the primary symptoms of autism (social deficits and communication impairment). Here, using another independent population ($n = 215$) and a new higher resolution measurement method that uses whole genome sequence data, we show, for the third time, that as the copy number of the CON1 Olduvai subtype increases autism symptoms become incrementally worse. Specifically, in male multiplex children a linear association between CON1 copy number and Social Diagnostic Score (SDS) ($b = 0.24$, $se = 0.12$, $p = 0.041$) and Communication Diagnostic Score (CDS) ($b = 0.23$, $se = 0.08$, $p = 0.007$) was confirmed. Combining data from the three studies ($n = 524$) identified linear associations of CON1 dosage with SDS ($b = 0.18$, $se = 0.05$, $p = 0.0008$) and CDS ($b = 0.13$, $se = 0.04$, $p = 0.0009$), further supporting a role for Olduvai sequences in the classic symptoms of autism. In addition, the enhanced discriminatory power of our sequence read depth method allowed us to implicate Olduvai subtypes specifically found in two of the 23 Olduvai-encoding genes (*NBPF1* and *NBPF14*) as primary contributors to this association. The variability of a latent genotype constructed using the copy number of *NBPF1* CON1 and *NBPF14* HLS1 subtypes explained 6.1% of the variability of a latent social phenotype ($R^2 = 6.1\%$, $p = 0.002$). We also note that our associations with these core autism phenotypes are consistently found in multiplex but not simplex families, and thus may be particularly important in inherited forms of autism. Finally, Olduvai sequences are hyper-amplified specifically in humans and have been implicated in human brain evolution and cognitive function. Such findings, taken together with our autism findings, lend further support to the proposal that Olduvai domains may constitute a genomic trade-off in which the same sequences, depending on how they vary, can produce either evolutionary benefit or neurological disease.

3338F

Analysis of APOE and known depression SNPs in patients with multiple sclerosis. S.W. Brugger¹, J.C. Denny², M.F. Davis¹. 1) Microbiology and Molecular Biology, Brigham Young University, Provo, UT; 2) Departments of Biomedical Informatics and Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, USA.

Multiple Sclerosis (MS) is a debilitating disease characterized by the demyelination of the axons of the central nervous system. Although symptomatic manifestations of MS are unique to each patient, previous studies have demonstrated that up to 50% of MS patients will experience clinically significant depression at some point in their life. This correspondence warranted an investigation into a potential genetic link between MS and depression. Certain APOE variants have been reported to be linked to depressive affect in MS patients, which this study aimed to replicate. We also examined the relationship between MS and a collection of SNPs recently identified as being significantly associated with depression in the general population. This study was performed by dividing an EHR population of MS patients into three groups: MS patients with billing codes for depression at any point in their medical record (n=270), at least one year before their MS diagnosis (n=31), and at least one year after their MS diagnosis (n=154). Controls were defined as patients with MS but no billing codes for depression (n=968). Genome-wide association analyses did not identify any significant SNPs within each of these categories. Logistic regression analyses with both APOE and other depressive SNPs failed to reach statistical significance for association between MS and depression, although moderate significance was discovered with one known depression SNP on chromosome 1 (rs1432639, $p = 0.058$). We were unable to replicate the results linking known APOE SNPs to depression in MS patients, nor were we able to identify any associations between reported depression SNPs in the general population and MS in our patients.

3339W

Transcriptomic analysis in multiple system atrophy (MSA). C. Bleuh^{1,2}, I.S. Piras^{1,2}, J.S. Talboom^{1,2}, M.A. Naymik^{1,2}, I. Schrauwen³, M.D. De Both^{1,2}, M.J. Huettelman^{1,2}. 1) Translational Genomics Research Institute, Phoenix, AZ; 2) Arizona Alzheimer's Consortium, Phoenix, AZ; 3) Baylor College of Medicine, Houston, TX.

Multiple System Atrophy (MSA) is a rare neurodegenerative disorder characterized by ataxia, autonomic dysfunction, parkinsonism and aggregation of α -synuclein in oligodendrocytes. Here we present the RNA-sequencing profiling results of post-mortem brains (cerebellum white matter) from MSA patients ($n = 67$) and healthy controls (HC; $n = 66$), also stratifying for MSA subtype (MSA-P and MSA-C). This is the largest known RNA profiling study conducted on MSA post-mortem brains. Samples are from Banner Sun Health (Sun City, AZ), New South Wales (Sydney, AU) (Cohort1; C1) and Queen Square (London, UK) (Cohort 2; C2) brain banks. RNA was prepared using Illumina's Truseq kit (C1) and TempO-Seq protocol (BioSpyder) (C2) and sequenced using Illumina HiSeq 2500 and NextSeq. Differentially expressed genes (DEGs) were detected using DESeq2 including age, sex, and PMI as covariates. We conducted Gene Ontology (GO) and protein-protein interaction (PPI) network analysis on DEGs lists. We detected 33 DEGs in C1 (mostly upregulated), and 3 genes in C2 associated with MSA. In the MSA-P group, we detected 20 DEGs in C1 (mostly upregulated) and 3 DEGs in C2. No overlap between the cohorts were detected. In MSA-C, we detected 519 DEGs in C1 and 304 DEGs in C2 with a slight prevalence of upregulated genes (55.3% in C1 and 52.3% in C2). 30 genes were detected in both cohorts and not detected in MSA-P. GO analysis highlighted the enrichment of genes associated with "Negative regulation of phosphate metabolic processes" (C1) and "Astrocytes and glial cell projection" (C2). We generated a PPI network using the 30 candidate genes, detecting a sub-network including 9 DEGs, with 5 directly connected with APP. In conclusion, we show the presence of 30 candidate genes associated with MSA-C, the subtype characterized by olivopontocerebellar ataxia. The PPI network analysis showed a potential direct interaction of five of them (IL1RL1, TTLL7, GAS2, PCYT1B, and MTURN) with APP. It is known that β -amyloid may drive α -synuclein pathology through different mechanisms (e.g impairment of protein clearance, enhancing inflammation and phosphorylation). At this time, β -amyloid pathology has been detected only in one case of MSA-C. However, it has been suggested that β -amyloid plaques may contribute to cognitive impairment related to MSA and decreased levels of Ab42 were observed in MSA CSF, similarly to AD. Finally, GO processes detected could be related to the α -synuclein phosphorylation.

3340T

Genome-wide association studies of brain structure and function in the ~20,000 UK Biobank participants. K.J. Sharp¹, L.T. Elliott¹, F. Alfaro-Almagro², S. Shi¹, K. Miller², G. Douaud², J. Marchini^{1,3}, S. Smith². 1) Dept. of Statistics, University of Oxford, Oxford, Oxford, United Kingdom; 2) FMRIB, Wellcome Centre for Integrative Neuroimaging, University of Oxford, Oxford, UK; 3) The Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK.

Genome-wide association studies of brain imaging derived phenotypes (IDPs) have great potential to shed light on the genetic basis of many neurodegenerative and psychological traits. UK Biobank is in the process of brain imaging 100,000 participants. We recently analyzed the first tranche of ~8,500 subjects for over 3,000 IDPs and identified 148 clusters of SNP-imaging associations with lead SNPs that replicate at $p < 0.05$ in an additional set of 3,456 samples (consistent with a post-hoc winner's curse corrected power analysis). Notable significant and interpretable associations include: iron transport and storage genes, related to changes in T2* in subcortical regions; extracellular matrix and the epidermal growth factor genes, associated with white matter micro-structure and lesion volume; and genes regulating mid-line axon guidance development associated with pontine crossing tract organization. We will report an expanded set of results from an analysis of ~20,000 participants and highlight the role of computationally efficient methods for tests using multi-trait mixed models. These complement individual IDP tests by utilising estimates of genetic correlation to boost power, while needing to correct for a much smaller number of tests. In our multi-trait analysis of 23 groups of IDPs for ~8,500 subjects with up to 243 IDPs per group, we found 170 associations that survived a correction for 23 scans with $-\log_{10}$ p-value > 8.86 . 138 of these replicated at $p < 0.05$ including 25 loci that showed no genome-wide significance in the flanking region for any individual IDP in the corresponding group. Notable examples were rs62073157 ($P=4.07E-11$) associated with diffusion MRI measures of mode of anisotropy (dMRI MO) and rs12928124 ($P=1.31E-10$) associated with FreeSurfer measures of brain volume. rs62073157 is an eQTL of *MAPT* which has been associated with several neurodegenerative disorders, while rs12928124 is located in *IL-34* which has been shown to protect against oligomeric amyloid- β mediated synaptic dysfunction and neuronal damage in Alzheimer's disease. We also report results from a new Sparse Bayesian Multitrait Testing method (SBAT) which can identify the most probable IDPs within a group with greater precision than a full model test. By analysing a group of 15 subcortical volumes, we demonstrate markedly greater power to identify known associations than a recently published approach based on summary statistics from single phenotype GWAS of the same data.

3341F

Replication of the *SCFD1* associated locus in ALS using linear mixed models. A. Nicolas¹, M. Nalls^{2,3}, B. Traynor¹, International ALS Genomic Consortium. 1) Neuromuscular Diseases Research Section, Laboratory of Neurogenetics, National Institute on Aging, NIH, Porter Neuroscience Research Center, Bethesda, MD 20892, USA; 2) Molecular Genetics Section, Laboratory of Neurogenetics, National Institute on Aging, NIH, Porter Neuroscience Research Center, Bethesda, MD 20892, USA; 3) Data Tecnica International, Glen Echo, MD, USA.

ALS is a progressive and fatal neurodegenerative disease causing muscle weakness and atrophy and leading to death within 2 to 5 years of symptom onset, due to respiratory failure. This disease kills approximately 7,000 Americans annually and no effective treatments currently exist. A genome-wide association study published in 2016 (Van rheenen *et al.*, 2016) nominated five novel loci as being associated with ALS, based on linear mixed modeling: rs616147 in *MOBP*; rs7813314 in *LOC101927815*; rs10139154 in *SCFD1*; rs35714695 in *SARM1*, and rs75087725 (*C21orf2*). We have previously replicated the association of rs75087725 (*C21orf2*) and ALS (Nicolas *et al.*, 2018). Here, we attempted to replicate the remaining four loci in an independent dataset of 8,229 ALS cases and 36,329 controls. To do this, we performed standard quality control of genotyping data using PLINK software package (version 1.9). Haplotypes were estimated with SHAPEIT (version 2.r790). Variants were imputed using Minimac3 software (version 1.0.11) and using the reference panel of 1000 Genomes Project dataset (phase 3, version 5a, release 2013-05-02). Post-imputation quality control was performed to filter out variants with a $R^2 < 0.3$ and a $MAF < 0.01$. Joint association analysis was run using the MLMA-LOCO linear mixed models algorithm as implemented in GCTA and corrected for 5 principal components, age and gender. Within our data, the *SCFD1* locus only is replicated with a $p=2.89 \times 10^{-5}$, in the Van rheenen's study, $p=4.95 \times 10^{-5}$ and in the combined dataset, $p=5.342 \times 10^{-11}$ ($OR=1.014$, 95% $CI=1.010-1.018$). We did not find definitive evidence of association for *MOBP*, *LOC101927815* or *SARM1*. Another GWAS of 1,234 ALS cases and 2,850 controls did not replicate any of the 5 loci (Benyamin *et al.*, 2017), in a Chinese population. *SCFD1*, for Sec1 Family Domain Containing 1, is involved in the Golgi-to-ER retrograde transport and interacts with syntaxin, for the docking of synaptic vesicles.

3342W

A factor analysis of early stressful life events identifies evidence for G-E interaction in multiple sclerosis. *M.K. Horton¹, X. Shao¹, H. Quach¹, D. Quach¹, K.H. Bellesis², T. Chinn², C. Schaefer², L.F. Barcellos^{1,3}.* 1) Genetic Epidemiology and Genomics Laboratory, Division of Epidemiology, School of Public Health, University of California, Berkeley, Berkeley, CA; 2) Kaiser Permanente Division of Research, Oakland, CA; 3) Computational Biology Graduate Group, University of California, Berkeley, Berkeley, California.

Causes of multiple sclerosis (MS [MIM 126200]), a chronic inflammatory and autoimmune disease, remain largely unexplained, but evidence suggests gene-environment interactions (G-E) account for part of its etiologic complexity. Childhood stress is implicated in pathogenesis of adult disease, and because stress is mechanistically related to inflammation, it may play a role in MS etiology. Prior studies reported an association between childhood stress and MS; however, none have tested whether SNPs associated with MS or stress interact with childhood stress to modify MS risk. We hypothesize that individuals who experienced childhood stressful life events and carry MS- or stress-related SNPs will have an increased risk of MS compared to individuals with one or neither exposure. We analyzed 1,500 MS cases and 12,000 controls from Kaiser Permanente Northern California. Ten stressful life events during ages 0-10 and 11-20 years were retrospectively measured using a modified version of the Life Event Record. Factor analysis was used to combine events into fewer latent factors and derive factor scores. Genotyping was performed using Illumina microarrays on saliva or blood samples. We selected *a priori* established MS risk variants, including those in the major histocompatibility complex (MHC) and 200 non-MHC variants, and SNPs within 30 candidate stress-related genes involved in hypothalamic-pituitary-adrenal axis regulation, such as *CYP21A1*, *CRHR1*, and *HSD11B2*. To test for G-E interaction and risk of MS, we used logistic regression with interaction-terms for latent factor scores and SNPs. We adjusted for genetic ancestry, gender, year of birth, and mother's educational attainment. Factor analysis revealed four latent factors of childhood stressful life events: "crime/abuse", "left home", "death/ illness" of immediate family member, and "divorce/remarry" of parents. Results identified 18 significant G-E interactions ($p < 0.05$). The top interaction was rs1177228X"left home" ($p < 0.01$). Stratified results by genotype showed that as the latent factor "left home" increased, the odds of MS increased among rs1177228 AA and AG genotypes but not GG genotypes [$OR_{AA} = 1.42$ (95% CI: 1.20, 1.71); $OR_{AG} = 1.40$ (95% CI: 1.02, 2.10); $OR_{GG} = 1.04$ (95% CI: 0.94, 1.14)]. Although further studies with larger sample sizes are needed to increase power, our results suggest there is evidence for interaction between MS- and stress-related SNPs and childhood stressful life events in MS.

3343T

Analysis of predicted loss of function variants in MSSNG identifies potentially causative variants in autism spectrum disorder. *B. Przychodzen¹, S. Smieszek².* 1) Cleveland Clinic, Cleveland, OH; 2) Case Western Reserve University.

The complex genetic architecture of ASD encompasses profound clinical heterogeneity, which poses huge challenges in understanding its pathophysiology. We conducted a large scale analysis of the loss of function variants in the MSSNG whole genome sequencing data encompassing 7231 individuals to elucidate variants potentially relevant to ASD. We focused on rare LOF (MAF < 0.01, frameshift/stopgain/splicing), inherited variants within this data. We then calculated the cumulative frequency among patients. As a control group we processed GnomAD database to extract similar variants among normal population. We then focused on statistically relevant genes that is single gene level over-represented, group/family of genes over-represented. Altogether we used X predicted LOF variants We identified interesting variants including KCNU1 and FBN2, amongst others, a missense variant rs168885577 (6.18E-08), eQTL, that has been shown to be implicated in schizophrenia. Enrichment analysis led us to calcium ion binding (GO:0005509). In addition we have conducted TDT meta-analysis of the MSSNG data. Less than 3% of protein coding genetic variants are predicted to result in loss of protein function through the introduction of stop codon for example, frameshift, disruption of splice site however such predicted variants provide insights regarding the direction of effect. We focus on such as such can lead us to actionable targets providing insights about pathophysiology of ASD.

3344F

Genome-wide association analysis of excessive daytime sleepiness

in the UK Biobank identifies 42 loci. R. Saxena^{1,2,3,4}, H. Wang^{1,4}, S. Gill¹, S. Jones⁵, J. Lane^{2,3,4}, H. Dashti^{2,4}, B. Cade^{1,4}, B. Leger², Y. Song², K. Patel¹, D. Ray⁶, D. Lawlor^{7,8}, X. Zhu⁹, M. Rutter^{6,10}, M. Weedon⁵, J. Walker^{2,3}, S. Redline^{1,11}.
 1) Division of Sleep and Circadian Disorders, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; 2) Center for Genomic Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA; 3) Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA; 4) Broad Institute of MIT and Harvard, Cambridge, MA, USA; 5) Genetics of Complex Traits, University of Exeter Medical School, Exeter, UK; 6) Division of Endocrinology, Diabetes & Gastroenterology, School of Medical Sciences, Faculty of Biology, Medicine and Health, University of Manchester, UK; 7) MRC Integrative Epidemiology Unit at the University of Bristol, UK; 8) Population Health Sciences, Bristol Medical School, University of Bristol, UK; 9) Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH, USA; 10) Manchester Diabetes Centre, Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK; 11) Division of Pulmonary, Critical Care, and Sleep Medicine, Beth Israel Deaconess Medical Center, Boston, MA 02215, USA.

Excessive daytime sleepiness (EDS) affects 10-20% of the population and is associated with substantial functional deficits and a wide variety of adverse health outcomes. EDS results from chronic insufficient sleep duration or poor sleep quality related to suboptimal sleep behaviors or underlying primary sleep disorders. However, there is significant individual variation in susceptibility to EDS as well as evidence for heritability of this trait. To improve our understanding of the genetic basis of EDS and related sleep disorders, we performed a genome-wide association analysis (GWAS) of self-reported EDS using 452,071 individuals of European ancestry in the UK Biobank. We identified 42 significant loci ($P < 5 \times 10^{-8}$) associated with EDS, many of which also associate with other sleep traits, including self-reported short sleep duration, insomnia, chronotype, and accelerometer-derived sleep efficiency and daytime inactivity. Novel EDS loci include genes previously associated with primary sleep disorders, psychiatric and metabolic traits (e.g. *BTBD7*, *CACNA1C*, *GABRA2*, *HCRTR2*, *HTR7*, *KSR2*, *PCLC1*, and *RAI1*) as well as other loci. Identified signals were enriched for genes expressed in multiple brain tissues and central nervous pathways. Experimental analysis further showed knock out of *CG8916* (an ortholog of *GABRA2*) results in sleep alteration in *Drosophila melanogaster*. Mendelian randomization analyses were consistent with a causal association between BMI with EDS. Shared genetic basis were suggested between EDS and coronary artery disease, psychiatric diseases, cognitive traits, and reproductive aging. This largest GWAS for EDS identified multiple genetic loci and phenotypic associations that offer biological clues into pathways that contribute to EDS and chronic health disorders. .

3345W

Pathway signal detection analyses with an application to amyotrophic lateral sclerosis.

MQ. Zhang^{1,2,3}, A.S. Allen^{1,2,3}. 1) Department of Biostatistics and Bioinformatics, Duke University, Durham, NC; 2) Center for Genomic and Computational Biology, Duke University, Durham, NC; 3) Center for Statistical Genetics and Genomics, Duke University, Durham, NC.

Gene-set-based analyses are often used to identify pathways where pathogenic variants may lead to disease. These analyses combine information across the genes in a pathway and attempt to identify any evidence of non-null effects within. For example, Higher Criticism (HC) statistics can be used to detect the weak-sparse signal within a pathway by looking for the largest scaled departure of ordered p-values from expectation. Here we investigate HC-like statistics that are based on integrating the scaled departure of ordered p-values from expectation instead of taking the max. We can show that this approach improves signal detection of certain alternatives over the traditional HC test. We illustrate our approach with a pathway analysis using data from a whole-exome sequencing data of amyotrophic lateral sclerosis (ALS) and show that our approach prioritizes ALS related pathways.

3346T

The influence of puberty on risk of multiple sclerosis: A Mendelian

randomization study. A. Harroud^{1,2}, J.A. Morris^{2,3}, V. Forgetta^{2,3}, R. Mitchell⁴, G.D. Smith⁴, S. Sawcer⁵, J.B. Richards^{2,3,6,7,8}. 1) Department of Neurology and Neurosurgery, McGill University, Montreal, Quebec, Canada; 2) Centre for Clinical Epidemiology, Department of Epidemiology, Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, Quebec, Canada; 3) Department of Human Genetics, McGill University, Montreal, Quebec, Canada; 4) MRC Integrative Epidemiology Unit, School of Social and Community Medicine, University of Bristol, Bristol, United Kingdom; 5) Department of Clinical Neurosciences, Cambridge Biomedical Campus, University of Cambridge, Cambridge, United Kingdom; 6) Department of Medicine, McGill University Montreal, Quebec, Canada; 7) Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Quebec, Canada; 8) Department of Twin Research and Genetic Epidemiology, King's College London, London, United Kingdom.

Several observational studies have reported an increased risk of multiple sclerosis (MS) with earlier age at puberty, although others found no relationship. In addition, these studies are susceptible to bias from confounding, further limiting causal inference. Therefore, we performed a Mendelian randomization (MR) study to investigate the potential for a causal effect of age at puberty on MS susceptibility. We also explored how body mass index (BMI) might influence this association, as weight status relates to both puberty timing and risk of MS. We used 375 genome-wide significant variants associated with age at menarche in a genome wide association study (GWAS) involving 329,245 women from the ReproGen consortium. The genetic architecture of puberty timing across both sexes is highly correlated ($r_g=0.75$, $p=1.2 \times 10^{-79}$), allowing these variants to provide reliable insight into puberty timing in males as well. Of the 375 variants, 372 were ascertained either directly ($n=345$) or through proxies ($n=27$, median $r^2=1.00$) in the latest International MS Genetics Consortium GWAS, including 14,802 MS cases and 26,703 controls. The effect of each variant on MS was weighted by its association with age at puberty. Estimates were combined using an inverse-variance weighted meta-analysis. To test whether puberty effects on MS were dependent on BMI, we performed multivariable MR controlling separately for the effects of adult and childhood BMI. For this, we used summary statistics from the ENGAGE (87,048 individuals) and EGG (35,668 individuals) consortia, respectively. The inverse-variance weighted MR revealed that a 1-year increase in age at puberty decreased odds of MS by 8% (OR=0.92, 95% CI 0.86-0.99, $p=0.03$). However, multivariable MR analysis showed that after accounting for effects on adult BMI, the association of age at puberty with MS susceptibility attenuated significantly (OR=0.96, 95% CI 0.89-1.02, $p=0.36$). Similar results were obtained when adjusting for childhood BMI (OR=0.97, 95% CI 0.89-1.05, $p=0.40$). Sensitivity analyses, including multivariable MR-Egger regression, provided no evidence of directional pleiotropy or reverse causation. This study found support for an inverse association between age at puberty and risk of MS with a magnitude comparable to that reported in observation studies. This effect is largely mediated by the strong association between age at puberty and obesity, and a large causal effect of puberty timing independent of BMI is unlikely.

3347F

Regional collapsing of rare variation implicates specific regions in known and novel ALS genes.

S. Gelfman¹, S. Dugger¹, C. De Araujo Martins Mo¹, Z. Ren¹, C. Wolock¹, N.A. Shneider¹, H. Phatnan², E.T. Cirulli³, B.N. Lasseigne⁴, T. Harris⁵, T. Maniatis¹, G.A. Rouleau⁶, JrR.H. Brown⁷, A.D. Gitler⁸, R.M. Myers⁴, S. Petrovski⁹, A. Allen¹⁰, M.B. Harms¹, D.B. Goldstein¹. 1) Columbia University Medical Center, New York, NY, 10032 USA; 2) New York Genome Center, New York, NY, 10013 USA; 3) Human Longevity INC, San Diego, CA 92121 USA; 4) HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806 USA; 5) SV Health Investors, Boston, MA 02108 USA; 6) McGill University, Montreal, H3A 2B4 Canada; 7) University of Massachusetts Medical School, Worcester, MA 01655 USA; 8) Stanford University School of Medicine, Stanford, CA 94305 USA; 9) Austin Health and Royal Melbourne Hospital, University of Melbourne, Melbourne, Australia; 10) Duke University, Durham, NC, USA.

Large-scale sequencing efforts in amyotrophic lateral sclerosis (ALS) have implicated novel genes using gene-based collapsing methods. However, pathogenic mutations may be concentrated in specific genic regions. To address this, we developed two collapsing strategies, one focuses rare variation collapsing on homology-based protein domains as the unit for collapsing and another gene-level approach that, unlike standard methods, leverages existing evidence of purifying selection against missense variation on said domains. The application of these two collapsing methods to 3,093 ALS cases and 8,186 controls of European ancestry, and also 3,239 cases and 11,808 controls of diversified populations, pinpoints risk regions of ALS genes including *SOD1*, *NEK1*, *TARDBP* and *FUS*. We identify *LGALS1* as a novel ALS gene that is also associated with early age-at-onset and implicate *PKP4* as a novel candidate gene. This work establishes a paradigm for identifying genic regions contributing to ALS occurrence in both known and novel candidate genes.

3348W

Associations between hematologic laboratory values and multiple sclerosis subtypes. *J.M. Miller¹, J.C. Denny², M.F. Davis¹.* 1) Department of Microbiology and Molecular Biology, Brigham Young University, Provo, Utah, USA; 2) Departments of Biomedical Informatics and Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, USA.

Multiple Sclerosis (MS) is a complex disease of the central nervous system in which the myelin sheaths of the nerve cells in the brain and spinal cord are damaged. Because expression of the disease varies widely between patients, four subtypes of MS have been defined based on patterns of its progression. Two of these subtypes are the focus of this study: Primary Progressive Multiple Sclerosis (PPMS), and Relapsing Remitting Multiple Sclerosis (RRMS). PPMS is the gradual progression of the disease with no relapses or remissions, while RRMS is characterized by unpredictable attacks (with potential permanent deficits) followed by periods of disease quiescence. Patients are classified into a subtype after monitoring of the disease over time. There are no known risk factors that vary significantly between PPMS and RRMS. In this study, we evaluated common laboratory tests to see if there are differences between the two populations without monitoring disease activity. We identified 5,416 people from the Vanderbilt Synthetic Derivative matching a case definition of MS, and then identified MS subtypes using previously developed natural language processing algorithms. We then took laboratory tests comprising the Complete Blood Count (CBC) from 2,440 patients with either PPMS (n=352) or RRMS (n=2,088) and compared them to see if there was a correlation between each specific lab value and MS subtype. Logistic regression was performed on the highest laboratory value for each patient, using patient's age at the date of the laboratory test as a covariate. This analysis showed that hemoglobin, PCV, platelet count, red blood cell count, and white blood cell count (.024, .0207, .0186, .00484, and .0102, respectively) were significantly different between the two subtypes of MS. We will look for genetic associations with the significant CBC and Comprehensive Metabolic Panel (CMP) lab values.

3349T

Optimal and data-adaptive p-value combination tests with applications in omics data analysis. *Z. Wu¹, H. Zhang².* 1) Worcester Polytechnic Institute, Worcester, MA; 2) Merck & Co., Inc., Kenilworth, NJ.

P-value combination is an important statistical approach for information-aggregated decision making. It is foundational to a lot of applications such as meta-analysis, data integration, signal detection, and others. We propose two generic statistic families for combining p-values: gGOF, a general family of goodness-of-fit type statistics, and tFisher, a family of Fisher type p-value combination with a general weighting-and-truncation scheme. The two families unify many optimal statistics over a wide spectrum of signal patterns, including both sparse and dense signals. We provide efficient solutions for analytical calculations of p-value and statistical power, as well as studies of asymptotic efficiencies, while emphasizing the conditions of realistic data analysis: small or moderate group size, Gaussian and non-Gaussian distributions, correlated data, generalized linear model based alternative hypotheses, etc. Based on these two families of statistics, omnibus tests are also designed for adapting the family-retained advantages to unknown signal patterns. Applications of these methods are illustrated in integrative analyses of large-scale omics data.

3350F

Bivariate causal discovery and its applications to gene expression and imaging data analysis. R. Jiao¹, N. Lin¹, Z. Hu², D.A. Bennett³, L. Jin², M. Xiong^{1,2}. 1) The University of Texas School of Public Health, Houston, TX; 2) School of Life Sciences, Fudan University; 3) Rush University.

Genetic data analysis has experienced significant progress on statistical association, but it remains elusive to understand the etiology and mechanism of complex phenotypes. As a major analytical platform, association analysis may hamper the theoretic development of genomic science and its application. Thus, many researchers suggest making the transition from association to causation. To explore bivariate causal discovery, we introduced independence of cause and mechanism (ICM) as a basic principle, using additive noise model (ANM) as a major tool for bivariate causal discovery. Large-scale simulations were performed to evaluate the feasibility of the ANM for bivariate causal discovery. To further evaluate its performance, ANM was applied to the gene regulatory network construction of Wnt signaling pathway with RNA-Seq of 145 genes measured in 447 tissue samples. Around fifty most significant causal relations were selected to test detection accuracy. For comparison, the score-based linear structural equation model (SEM), causal additive model (CAM), PC algorithm and random network were also included in analysis. Results showed that 38% of ANM-detected paths agreed with the KEGG, while SEM only reached 20% accuracy, CAM 16%, PC 21.57% and random network 25.41%. The ANM outperformed the other three causal algorithms. Also, ANM was applied to the Alzheimer's Disease Neuroimaging Initiative (ADNI) data with 91 individuals with Diffusion Tensor Imaging measured at four time points: baseline, 6 months, 12 months and 24 months. We identified causal relationships between working memory and activities of the temporal R (hippocampus) at 24 months with $P = 0.00014$ (image working memory), but found no association ($P = 0.59$). Corresponding images showed that the decreased neural activities in the temporal R region deteriorated AD patients' working memory. This result provided evidence that causation may be identified in the absence of association signals. We also observed only association between cholesterol and imaging signal variation at 12 and 24 months in the Occipital_Mid brain region ($P = 0.0003$ at 12m, $P = 0.00004$ at 24m), but no causation ($P = 0.68$ at 12m, $P = 0.19$ at 24m), which was consistent to the fact of no significant imaging signal variation in the occipital lobe region. It infers that association may not provide enough information on mechanism of complex phenotypes.

3351W

Polygenic risk score analysis of antidepressant response in major depressive disorder in the STAR*D study. W. Guo¹, A. Rush², L. Hou³, I. Henter⁴, M. Ritter¹, S. Kang¹, N. Akula³, F. McMahon³, Y. Shugart¹. 1) Statistical Genomics and Data Analysis Core, National Institute of Mental Health, National Institutes of Health, Bethesda, MD; 2) Duke – National University of Singapore, Singapore; 3) Human Genetics Branch, National Institute of Mental Health, National Institutes of Health, Bethesda, MD; 4) Section on PET Neuroimaging Sciences, National Institute of Mental Health, National Institutes of Health, Bethesda, MD.

Major depressive disorder (MDD) is one of the most common mental disorders in the United States, affecting about 6.7% of the population in a given year. The Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study provided the largest cohort and was designed to assess effectiveness of treatments in patients with nonpsychotic major depressive disorder. In the STAR*D data, four levels of the study were designed to assess the effectiveness of depression medication in patients with major depressive disorder. Antidepressant response has been shown as a polygenic trait and involves a large number of SNPs with small effect sizes, and approximately 42% of phenotypic variance in antidepressant response was explained by common SNPs across the human genome. The STAR*D Study aimed to assess the effectiveness of depression treatments in 4,041 outpatients with major depression, recruited in 18 primary care and 23 psychiatric clinical sites across the United States. Of the initial 4,041 participants, 1,165 were extracted because they either were not qualified to have "at least moderate" depression or they chose not to participate. Then as entering level 1, 2,876 participants were treated with the antidepressant citalopram, a selective serotonin reuptake inhibitor (SSRI) from 2 to 14 weeks. In this study, QIDS-SR₁₆ was used as the primary measure to report citalopram's efficacy. First, the most stringent quality control (QC) was carried out for the STAR*D data at both SNP-level and sample-level before imputation. The STAR*D samples were genotyped on Affy 5.0 platform. Second, the imputation of genotypes was conducted using the IMPUTE2 software program using the 1000 Genomes Phase 3 dataset as reference. Furthermore, as a comparison, we re-genotyped a subset of STAR*D-Omni1S samples by Illumina Omni1S BeadChip to inspect the imputation concordance within this replication samples to evaluate the quality of imputation. To detect single-nucleotide polymorphisms (SNPs) associated with antidepressant response, we re-analyzed the antidepressant efficacy in MDD patients of STAR*D data using well-imputed dosage data. Finally, a polygenic risk score analysis was carried on the antidepressant response at all four levels in Major Depressive Disorder in the STAR*D samples. .

3352T

Selecting best polygenic risk scoring method based on the features of a study. A.A. Shabalin¹, J.S. Anderson¹, A.R. Docherty^{1,2}. 1) Psychiatry, University of Utah, Salt Lake City, UT; 2) Virginia Institute for Psychiatric & Behavioral Genetics, Virginia Commonwealth University, Richmond, VA.

Polygenic risk scoring (PRS) is a new approach for phenotype prediction designed for highly polygenic outcomes. The predictive power is achieved by integrating millions of effects across the entire genome into a single metric. Due to the success of PRS paradigm, multiple PRS calculation methods have been proposed, each with individual strengths and weaknesses and a distinct set of parameters (or priors for Bayesian methods). The choice of the method and parameters can strongly affect the performance of calculated PRS and there is no single best method / set of parameters. All of the following features affect polygenic risk score calculation: 1) polygenicity of the phenotype, 2) heritability of the phenotype, 3) reliability/validity of the phenotype measurement, 4) whether the outcome is binary or continuous, 5) prevalence of binary outcomes / distribution of continuous outcomes, 6) sample size of the training data and the test cohort, 7) number of SNPs tested in original GWAS and test cohort. These features can be used to select a PRS method and its parameters. Here we present empirically derived guidelines for selection of PRS calculation methods and for the choice of the parameters based on the aforementioned features. The assessment is conducted using both real and simulated data with a variety of clinical phenotypes and psychiatric diagnoses.

3353F

Prioritizing risk genes for neurodevelopmental disorders using pathway information. T. Nguyen¹, A. Dobbyn^{1,2}, J. Bryois³, A. Charney¹, L. Huckins¹, W. Wang¹, D. Ruderfer⁴, X. Xu⁵, J. Buxbaum⁶, D. Pinto^{5,6,7}, X. He⁸, P. Sullivan⁹, E. Stahl¹. 1) Division of Psychiatric Genomics, Department of Genetics and Genomic Sciences, Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 2) Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 3) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 4) Division of Genetic Medicine, Departments of Medicine, Psychiatry and Biomedical Informatics, Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA; 5) Seaver Autism Center, Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 6) The Mindich Child Health & Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 7) Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 8) Department of Human Genetics, University of Chicago, Chicago, IL, USA; 9) Departments of Genetics and Psychiatry, University of North Carolina, Chapel Hill, North Carolina, USA.

Gene-set enrichment analysis has contributed to the understanding of the genetic architectures of complex diseases. In current de novo and rare case-control sequencing studies, either gene set analyses or individual gene analyses are conducted, but typically results of gene set enrichment analysis are not used to better identify individual risk genes. Here, we have developed a Bayesian method-based pipeline, termed gTADA, which identifies enrichment in gene sets or gene-expression levels through their direct incorporation into models of de novo (DN) and rare inherited/case-control (rCC) gene burden data. With this pipeline, arbitrary significance thresholds can be circumvented. Our method can leverage external gene-level information to prioritize risk genes. Applying gTADA to available whole exome sequence data sets including developmental disorder (DD), intellectual disability (ID), autism spectrum disorder (ASD), epilepsy (EPI), and congenital heart disease (CHD) phenotypes, gTADA is able to replicate previously reported gene-set enrichments and identify novel risk genes. For EPI, gTADA prioritized multiple significant genes of which 25 are novel significant genes (posterior probabilities > 0.95) in a meta-analysis of multiple types of EPI (356 trios, 5704 cases and controls). Several of our novel EPI risk genes replicate in a recent whole genome sequencing study using independent samples. Top prioritized EPI genes have high protein-protein interaction network connectivity, and their expression during human brain development correlates with subnetwork functional annotations. Finally, we also saw more enriched drug-target gene sets for EPI than for other disorders.

3354W

Methods for estimating covariate effects on 5hmC methylation: A proposed method and a comparison. C.M.T. Greenwood^{1,2}, K. Greenlaw¹, J. Gross³, A. Canty⁴, A. Ciampi¹, G. Turecki^{2,3}. 1) Lady Davis Institute for Medical Research, Montreal, Quebec, Canada; 2) McGill University, Montreal, QC, Canada; 3) Douglas Mental Health University Institute, Montreal, QC, Canada; 4) McMaster University, Hamilton, ON, Canada.

Introduction: 5hydroxymethylcytosine (5hmC) methylation levels in DNA can be estimated by applying both an oxidative treatment and a bisulfite treatment to DNA samples, and then comparing the resulting signal distributions. This possibly transitory methylation state is more commonly detected in brain tissues than in other tissues, and may be associated with psychiatric and behavioural phenotypes. Although methods have been developed to estimate 5hmC methylation levels, little attention has been paid to estimating the effects of covariates on 5hmC methylation. This could be important, for example to assess whether 5hmC methylation, at particular CpG sites, is associated with a disease state. Hence, we propose 2 new approaches and compare performance in a simulation study. **Methods:** A set of 38 postmortem brain samples were assayed with the Illumina 450K platform, measuring methylated and unmethylated signals with and without additional oxidative treatment. We carefully selected a few CpG sites showing signature signal patterns across the samples. These signatures include high overall methylation with and without strong apparent 5hmC methylation, and low overall methylation with and without apparent 5hmC methylation. From these raw data distributions, we designed a rich simulation study to detect covariate effects on 5hmC methylation. We simulated presence or absence of covariate effects on either 5hmC methylation, 5methylcytosine (5mC) methylation, both, or neither. We then compared the type 1 error and power for detection of covariate influences on 5hmC methylation across a published method (Houseman et al. 2016), and two proposed approaches built on generalized linear models with contrasts, with and without random effects. **Results:** Both our approaches control type 1 error and show adequate power for detection of 5hmC methylation. The previously published approach yielded strongly inflated type 1 error for detection of covariate effects on 5hmC methylation when covariate effects on 5mC methylation were present. In contrast, one of our approaches, using random effects, controls the type 1 error well in this situation. **Conclusion:** Care must be taken to choose an appropriate analytic method when looking for factors influencing 5hmC methylation levels. Our proposed approach with random effects takes the signal intensity into account, as well as the inherent pairing between signals before and after oxidative treatment, and performs well.

3355T

Environmental factors are often heritable. Does this bias polygenic gene-by-environment interaction analyses? W.J. Peyrot¹, M.C. Keller², W. van Rheenen³, N.R. Wray², B.W.J.H. Penninx¹. 1) VU medical center, Noord-Holland, Netherlands; 2) University of Colorado, Boulder, US; 3) University of Queensland, Brisbane, Australia.

Background In the era of genome-wide genotype data, research in gene-by-environment interaction (GxE) has shifted its focus from candidate genes to polygenic methods, such as analyses of polygenic risk scores (PRS) and the genetic relatedness matrix (GRM). This progress acknowledges the highly polygenic nature of psychiatric disorders, but also introduces new challenges. In particular, most environmental factors tested in GxE analyses have a genetic component, such as smoking, cannabis use and even exposure to childhood trauma (CT). We hypothesize that commonly used polygenic GxE analyses may be biased when analyzing heritable environmental factors. **Methods** Polygenic data were simulated using population parameters representative of Major Depressive Disorder (MDD), i.e., heritability of 35% and lifetime risk 15%. In addition, we simulated CT for the same individuals, assuming a cumulative CT risk of 20%, and an odds ratio for MDD of 2.5 for exposed individuals (CT=1) compared to unexposed individuals (CT=0). We simulated 5 scenarios: (1) no interaction & CT not heritable (CT-h2I=0), (2) different direction of genetic effect on MDD in CT=1 compared to CT=0, CT-h2I=0, (3) larger genetic effects on MDD in CT=1 compared to CT=0 & CT-h2I=0, (4) no interaction & CT-h2I=50% with no genetic correlation with MDD (other than via the causal effect of CT on MDD), and (5) no interaction & CT-h2I=50% with genetic correlation with MDD. On these simulated data, we performed PRS analyses (logistic regression of MDD on PRSxCT) and GRM-based analyses (cross-product Haseman-Elston regression to estimate (1) the genetic correlation between MDD in CT=1 and MDD in CT=0, and (2) the difference in magnitude of the liability-scale heritability of MDD in CT=1 compared to CT=0). **Results** The PRS- and GRM-based analyses provided non-biased results for scenarios 1-4. For scenario 5, a slight bias was introduced when simulating a relatively strong genetic correlation between MDD and CT of 0.53 (MDD-h2I of 0.31 in CT=1 and 0.33 in CT=0). **Discussion** Contrary to our hypothesis, these simulations suggest that polygenic GxE analyses generally provide reliable results. However, GxE results should be interpreted with caution when a heritable environmental factor is studied with a strong genetic correlation with the disorder. A limitation of our work is that we cannot rule out potential bias from other genetic architectures or polygenic methods than those studied here.

3356F

Genome-wide association study of amphetamine related stimulant dependence risk variants. J.W. Cox^{1,2}, R. Sherva³, L. Wetherill⁶, T. Foroud⁶, H. Edenberg⁷, H. Kranzler³, J. Gelernter^{4,5}, L. Farrer^{1,2,8,9,10,11}. 1) Section of Biomedical Genetics, Dept of Medicine, Boston University School of Medicine, Boston, MA; 2) Program in Genetics and Genomics, Dept of Medicine, Boston University School of Medicine, Boston, MA; 3) Dept of Psychiatry, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; 4) Dept of Psychiatry, Genetics, and Neuroscience, Yale Univ. School of Medicine; West Haven, CT; 5) VA CT Healthcare Center, West Haven, CT USA; 6) Dept of Medical And Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN; 7) Dept of Biochemistry and Mol. Biology, Indiana University School of Medicine, Indianapolis, IN; 8) Dept of Neurology, Boston University School of Medicine, Boston, MA; 9) Dept of Ophthalmology, Boston University School of Medicine, Boston, MA; 10) Dept of Biostatistic, Boston University School of Medicine, Boston, MA; 11) Dept of Epidemiology, Boston University School of Medicine, Boston, MA.

Amphetamine dependence is an understudied disorder of public health significance. Studies have consistently found it to have a genetic basis, but to date there has not been an adequately-powered genome-wide association study (GWAS) to identify genetic risk factors for stimulant dependence. We performed a GWAS for DSM-IV stimulant dependence in the Yale-Penn study sample, composed of small families and unrelated African Americans (AA, 149 cases, 5243 controls) and European Americans (EA, 368 cases, 3951 controls) ascertained for dependence on no/other substances. Dependence diagnosis was regressed on imputed SNP dosage adjusted for sex, age and the first 5 ancestry PCs in a logistic regression solved with generalized estimating equations to adjust for relatedness. All included subjects reported lifetime exposure to stimulants. Association tests were performed within each array with AA, EA separately. Summary statistics were analyzed via inverse variance meta-analysis. We attempted to replicate SNPs with $p < 1 \times 10^{-6}$ in 2 independent datasets of AAs (N=445) and EAs (N=2333) groups. In the discovery cohort, 10 genomic regions passed a $p < 1 \times 10^{-6}$ threshold for replication. The most significant association in discovery AAs was rs55670624 ($p = 7.1 \times 10^{-8}$, OR=0.46) located 91 kb upstream of *KYNU*, intergenic between *KYNU* and *LRP1B*. We had a genome-wide significant (GWS) finding for rs73000123 after combining with replication AAs (meta $p = 4.66 \times 10^{-8}$, OR=0.3), located 44 kb upstream of rs55670624 ($r^2 = 0.178$, $D' = 0.955$). An intergenic SNP (rs76036414) between *ISL1* and *PELO* showed marginal association in discovery AAs ($p = 3.91 \times 10^{-7}$, OR=0.38) with the same (non-significant) effect direction in replication AAs (OR= 0.74) and EAs (OR=0.95). To date, this is the first GWAS for non-cocaine stimulant dependence to identify GWS results. *KYNU* encodes kynureninase, an enzyme that has been implicated in cannabis dependence in AAs. *LRP1B* encodes a low-density lipoprotein family receptor, whose expression is down-regulated in chronic cocaine-treated mouse brain. *ISL1* encodes a protein that regulates insulin gene expression, and has been reported to be differentially methylated in sperm of cocaine-preferring mice and differentially translated in motor neurons. *PELO* encodes a protein with a conserved nuclear localization signal. These findings include both previously identified and novel genes and pathways involved in substance dependence which may contribute to stimulant dependence risk.

3357W

Identifying sub-phenotypes from hidden structure in polygenic SNP scores. J. Yuan, B. Lai, H. Xing, A. Lamy, I. Pe'er. Computer Science, Columbia University, New York, NY.

While genome-wide association studies have in recent years discovered numerous associations between genetic variants and increased risk for disease phenotypes, GWAS analysis methods typically assume that phenotypic labels are homogeneous, with no underlying sub-structure arising from distinct disease etiologies. This assumption may not be true for diseases such as schizophrenia, which has been suggested to comprise multiple hidden genetic disorders despite presenting outwardly as a single disease. We propose to discover hidden sub-phenotypes from polygenic SNP scores by applying a logistic regression method in which the labels are not provided. Instead, the labels are inferred along with effect sizes distinguishing the sub-phenotypes, by taking advantage of the underlying heterogeneity in individual genotype data of cases. We demonstrate through simulation the robustness of this method in comparison to traditional clustering methods, which are more easily confounded by the numerous small effects characteristic of polygenic traits. For simulations of 30,000 genotypes of cases and a per-SNP heritability of .04 percent, our method discovers SNP effects and associates them with the correct sub-phenotypes with a consistently greater than 95 percent accuracy. We also present preliminary results on GWAS data of schizophrenia obtained from the Psychiatric Genomics Consortium. We further extend the scenario of heterogeneity to quantitative traits, showing the presence of heterogeneous cases to be detectable when evaluating the likelihood of correlations between associated SNPs. Finally, we combine a layer of quantitative phenotypes with a dichotomous outcome through hierarchical models. We demonstrate by simulation that with samples of at least 50,000 genotypes, this method recovers associations at both the hidden and observed phenotype layers with an R-squared of greater than 0.60. When heterogeneity of the dichotomous phenotype arises at the quantitative layer, we show it is also detectable under such parameters. We believe these methods may be applied widely across existing GWAS summary statistics to enhance understanding of the genetic etiology of these associations.

3358T

Inferring association between trait and single protein/protein pathways based only on GWAS summary statistics. S. Bacanu, C. Chatzinakos, B. Webb, B. Riley, V. Vladimirov, K. Kendler. Virginia Commonwealth University, Richmond, VA.

Lately, much attention was paid to “imputed” transcriptomic methods, i.e. using summary statistics to infer association between the trait and gene expression (GE) of gene/pathway of genes. However, there are various levels of RNA degradation between transcription and protein synthesis, which is assumed to greatly influence downstream cellular processes. Thus, proteomics is a more biologically informative measure for trait analysis and, thus, more powerful when available. We thus propose JEPEGMIX2-P, a proteomics software that is based on our already developed Gene-level Joint Analysis of functional SNPs in Cosmopolitan Cohorts 2 Pathway (JEPEGMIX2) transcriptomic software. The proposed method is an $O(n)$ procedure, uses only summary statistics to test the association between trait and protein concentration (PC) of single proteins or protein pathways, as estimated from protein Quantitative Trait Loci (pQTLs). Due to its practical use of predicted PCs, the test is practically a Mendelian Randomization statistic, which avoids the counter-causation and confounders encountered by most of the higher powered observational studies. Being written in C++, JEPEGMIX2-P takes less than 5 days to compute all proteins, protein pathways statistics. We applied JEPEGMIX2-P to summary statistics from Psychiatric Genetics Consortium data sets. This application suggests a big role for protein pathways with PCs that are regulated by several microRNA. .

3359F

Significance testing for allelic heterogeneity. Y. Deng, W. Pan. University of Minnesota Twin Cities, Minneapolis, MN.

It is useful to detect allelic heterogeneity (AH), i.e. the presence of multiple causal SNPs in a locus, which for example may guide developing new methods for fine mapping and may determine how to interpret an appearing epistasis. In contrast to Mendelian traits, the existence and extent of AH for complex traits had been largely unknown until the recent publication of Hormozdiari et al. (2017, *AJHG* 100: 789–802.). Hormozdiari et al. (2017) proposed a Bayesian method, called causal variants identification in associated regions (CAVIAR), and uncovered wide-spread AH in complex traits. However, there are several limitations with CAVIAR. First, it assumes a maximum number of causal SNPs in a locus, typically up to 6, to save computing time; this assumption, as to be shown, may influence the outcome. Second, its computational time can be too demanding to be feasible since it examines all possible combinations of causal SNPs (under the assumed upper bound). Finally, it outputs a posterior probability of AH, which may be difficult to calibrate with a commonly used nominal significance level. Here we introduce an intersection-union test (IUT) based on a joint/conditional regression model with all the SNPs in a locus to infer AH. We also propose two sequential IUT-based testing procedures to estimate the number of causal SNPs. Our proposed methods are applicable to not only individual-level genotypic and phenotypic data, but also GWAS summary statistics. We provide numerical examples based on both simulated and real data, including a large-scale schizophrenia (SCZ) and a high-density lipoprotein (HDL) GWAS summary datasets, to demonstrate the effectiveness of the new methods. In particular, for both the SCZ and HDL data, our proposed IUT not only was faster, but also detected more AH loci than CAVIAR. Our proposed methods are expected to be useful in further uncovering the extent of AH in complex traits.

3360W

GAMBITS: An improved method for genetic association testing of symptom and questionnaire data in psychiatric genetic studies. A.M. Holleman^{1,2}, K.A. Broadaway³, R. Duncan³, L.M. Almli⁴, B. Bradley^{4,5}, K.J. Ressler⁶, D. Ghosh⁷, J.G. Mulle^{2,3}, M.P. Epstein^{2,3}. 1) Department of Epidemiology, Emory University, Atlanta, GA; 2) Center for Computational and Quantitative Genetics, Emory University, Atlanta, GA; 3) Department of Human Genetics, Emory University, Atlanta, GA; 4) Department of Psychiatry and Behavioral Sciences, Emory University, Atlanta, GA; 5) Mental Health Service Line, Department of Veterans Affairs Medical Center, Atlanta, GA; 6) Department of Psychiatry, McLean Hospital, Harvard Medical School, Belmont, MA; 7) Department of Biostatistics and Informatics, Colorado School of Public Health, Aurora, CO.

Genetic studies of psychiatric disorders often deal with phenotypes that are not directly measurable. Instead, researchers rely on multivariate symptom data from questionnaires and surveys like the PTSD Symptom Scale (PSS) and Beck Depression Inventory (BDI) to indirectly assess a latent phenotype of interest. Researchers subsequently collapse such multivariate questionnaire data into a univariate outcome to represent a surrogate for the latent phenotype. However, when a causal variant is only associated with a subset of collapsed symptoms (which is a likely phenomenon for phenotypically heterogeneous psychiatric disorders), the effect will be challenging to detect using the univariate outcome. We propose a more powerful method to analyze multivariate symptom data called Gene Association Method for Broader Integration of Tests and Scores (GAMBITS). GAMBITS compares similarity in multivariate symptom-scale data from questionnaires to similarity in common genetic variants across a gene. We use simulated data to demonstrate that GAMBITS provides substantially increased power over standard approaches that collapse questionnaire data into a single surrogate outcome (i.e., univariate linear regression and kernel machine regression), while properly controlling Type I error. GAMBITS is also computationally efficient and can scale easily to large-scale psychiatric GWAS projects. We illustrate our approach using GWAS data from the Grady Trauma Project, and identify two study-wise significant genes; one gene (*SIRPA*) associated with PSS and another gene (*ZHX2*) associated with BDI. Both genes were not identified using standard univariate techniques.

3361T

Life is pain: Fibromyalgia as a nexus of multiple liability distributions.

A. Moscati¹, L.K. Davis², C. Arnaiz², G.N. Nadkarni¹, J.H. Cho¹, R.J.F. Loos¹, K.S. Kendler³. 1) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA; 2) Division of Genetic Medicine, Department of Medicine, Vanderbilt University, Nashville, TN, USA; 3) Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, 23298, USA.

Background: Fibromyalgia syndrome (FMS) is a condition characterized by widespread pain, tiredness, concentration problems, and other ailments. This diffuse complex of symptoms overlaps with other disorders, which can make diagnosis challenging. FMS also displays pronounced comorbidity with a number of diagnoses, but little research has focused on shared etiology between them. This study aims to identify and estimate phenotypic and genetic relationships that FMS may have with other diseases. **Methods:** Using data from the BioMe and BioVU biobanks (N = 19,558 and 64,208, respectively), ICD-10-defined FMS (M79.7, which had a PPV of 0.92 based on a clinician-validated subset of cases) was used to predict an extensive array of other ICD-defined conditions (1,814 phenotypes in BioMe) in a phenome-wide association analyses (PheWAS). GWAS summary statistics for depressive symptoms (DEP), neuroticism (NEU), and subjective well-being (SWB) were used to construct polygenic scores, which were correlated with FMS to estimate the genetic relationships in an initial European subset of BioMe (N = 2026). **Results:** PheWAS showed that FMS was robustly associated (all reported results having p-values < 1e-15) with psychiatric diagnoses, with odds ratio (OR) for major depressive disorder > 7 (OR_{anxiety disorder} > 7), autoimmune conditions (OR_{rheumatoid arthritis} > 20, OR_{lupus} > 10), and joint degeneration (OR_{spondylosis} > 14, OR_{osteoarthritis} = 6). Shared etiology between FMS and psychiatric traits was weakly suggested using polygenic scores; FMS was positively correlated with the polygenic scores for DEP and NEU, and negatively correlated with the score for SWB at all thresholds (p-values ranged from 0.02 to 0.77, due to limited N in the subset). **Conclusions:** Our findings suggest that FMS has a phenotypic relationship with internalizing disorders, with autoimmune conditions, and with joint degeneration. The polygenic score results are slightly suggestive of a genetic relationship with DEP & NEU (negative with SWB) as well, though further refinement of this analysis in the full sample of both biobanks, while incorporating autoimmune and joint degeneration polygenic scores, is underway. The results also indicate, due to the extent of sharing, that FMS should be treated as seriously as arthritis or major depression, and should not be dismissed as malingering. Progress in research on the condition could, indeed, benefit multiple branches of medicine.

3362F

Describing the genetic architecture of psychiatric disorders from population scale genealogies. A.J. Schork^{1,2}, W.K. Thompson³, T.M. Werge^{1,2}, A. Buil^{1,2}, . for the iPSYCH Consortium². 1) Institute for Biological Psychiatry, Roskilde, Denmark; 2) The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Copenhagen, Denmark; 3) Department of Family Medicine and Public Health, University of California, San Diego.

The Danish national medical register system represents an essentially complete aggregation of hospital discharge diagnoses for all persons living in Denmark since 1968. Through a central person register which links parents to children through a unique person identifier it is possible to reconstruct population scale genealogies. In this work, we have reconstructed the genealogies of 90,000 individuals genotyped as part of the iPSYCH case-cohort ascertained for the study of major psychiatric disorders. The genealogies of these 90,000 individuals include more than 1,500,000 individuals constituting more than 15,000,000 pairs of genetic relatives and span up to six generations. Integrating these data resources into a recently develop framework for efficient mixed model calculation, we estimate heritability of and genetic correlation among psychiatric and neurological disorders using one of the largest single cohorts. Further, we take advantage of the fact that our genealogies are seeded with genotyped subjects to compare the predictive performance of estimated genetic values calculated from the genealogies to those estimated via SNP-based polygenic risk scores.

3363W

Paternal age, birth order, and early-onset in sporadic schizophrenia cases. S. Wang¹, P. Hsiao², L. Yeh³, C. Liu⁴, C. Liu⁴, T. Hwang⁴, M. Hsieh⁴, Y. Chien⁴, Y. Lin⁴, S. Chandler⁵, S. Faraone⁶, S. Glatt⁶, M. Tsuang⁶, H. Hwu^{2,4,7}, W. Chen^{2,7,8}. 1) Department of Public Health and Department of Occupational Safety and Health, China Medical University, Taichung, Taiwan; 2) Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan; 3) Department of Healthcare Administration, College of Medical and Health Science, Asia University, Taichung, Taiwan; 4) Department of Psychiatry, College of Medicine and National Taiwan University Hospital, National Taiwan University, Taipei, Taiwan; 5) Center for Behavioral Genomics, Department of Psychiatry; & Institute for Genomic Medicine, University of California San Diego, La Jolla, California, USA; 6) Departments of Psychiatry and Behavioral Sciences and Neuroscience and Physiology, Medical Genetics Research Center, SUNY Upstate Medical University, Syracuse, New York, USA; 7) Institute of Brain and Mind Sciences, College of Medicine, National Taiwan University, Taipei, Taiwan; 8) Department of Public Health, College of Public Health, National Taiwan University, Taipei, Taiwan.

Background: Advanced paternal age has been associated with both increased risk of schizophrenia and earlier onset of schizophrenia. The leading hypothesis postulated to explain the paternal age effect is the increased *de novo* mutations with men's age. Among patients with schizophrenia, the birth order of the affected proband have been considered as a proxy for *de novo* mutations, and older paternal age is associated with higher proband's birth order. Nevertheless, it cannot be ruled out that such an association could be spurious because both paternal age and birth order are correlated with *de novo* mutations. In sporadic schizophrenia cases, a higher proband's birth order indicated a greater cumulated number of healthy senior siblings in the family, thus rendering a higher birth order proband a lower familial risk. Hence, we expected that the paternal age effect would be stronger in probands with higher birth order. This study aimed to examine whether the birth order would modify the effect of paternal age on early-onset of schizophrenia based on a large collection of schizophrenia patients. **Methods:** A total of 3093 schizophrenia patients from 3008 families were recruited from the Schizophrenia Trio Genomic Research in Taiwan (S-TOGET) project. Among the participants, 2923 without apparent family history upon brief clinical screening. Probands with a paternal age of <20 were excluded (n=14). We tested the association between paternal age and early-onset (<18) of schizophrenia in 2558 families with an affected offspring and multiple biological siblings (range 2-9). **Results:** Per 1-year increase in paternal age, there was an increased odds of being early-onset (OR=1.05, p=0.05 for third-born, and OR=1.11, p=0.01 for fourth- or later- born). However, there was no association of paternal age with early-onset in patients with lower birth order (OR=1.03, p=0.10 for first-born, and OR=1.02, p=0.29 for second-born). After adjusting for maternal age and sibship size, the interaction between birth order and paternal age reached a marginal statistical significance (p=0.07). **Conclusions:** The birth order of sporadic schizophrenia patients can modify the effect of paternal age on early-onset of schizophrenia. The paternal age effect is stronger in probands with a higher birth order, who may need more cumulative effect from *de novo* mutations from father to develop schizophrenia.

3364T

Association of genetic profile for substance abuse potentials and urinary illicit drug detections. T.J. Nappi, M. Nakano, J. Ying, C. Sailey. Molecular Testing Labs, 14401 SE 1st St. Vancouver, WA. 98684.

Introduction: The *ANKK1*, *COMT*, *DRD2*, and *OPRM1* genes have been studied due to their functions within the dopaminergic system. In fact, single nucleotide polymorphisms (SNPs) of *ANKK1* gene are associated with disruption of the homeostatic balance of dopamine reception. The *COMT* gene polymorphisms modulates dopamine metabolism. *DRD2* and *ANKK1* genes share a partial genomic region where a polymorphism Taq1A, also known as *ANKK1* variant rs1800497, is associated with substance abuse. Thus, the aims of this study was to evaluate the association of abusive compounds in urine and the patients' genetic profile. Additionally, a meta-analysis of addiction-risk associated genes and apparent substance abuse was introduced to provide evidence supporting the association. **Methods:** De-identified data from our database was prepared by performing a criteria-based query for all results with paired pharmacogenomics and toxicological testing. Genetic polymorphisms for *ANKK1* rs1800497, *COMT* rs34680, *DRD2* rs1799978/rs6277, and *OPRM1* rs1799971 were merged with paired testing for grouped narcotics and illicit stimulants: opioids/opiates, 6-acetylmorphine, amphetamines, and fentanyl. The meta-analysis was performed using the NCBI PubMed to aggregate meta-data associated with addiction-related SNPs. **Results:** Three hundred eighty-two individuals' toxicology and pharmacogenomics data were analyzed for gene variants and positivity rates of narcotics and illicit stimulants, however, genotype and allelic analysis for all addiction related genes did not reach statistical significance. The meta-analysis including this study and published 36 clinical studies exhibited the association between *ANKK1* variants and addiction was significant, yet the impact of the variant was small (odds ratio of 1.182, $p < 0.001$). Although previous clinical studies were found to be the most impactful, this meta-analysis of this study found this variant's numerical with the averaged odds ratio of 1.182 ($p < 0.001$). **Conclusion:** A clear association of addiction-related gene profiles and illicit drug use was not obtained in this study. The meta-analysis data remains to confirm the association of *ANKK1* as a genetic marker related to addiction. However, this association of *ANKK1* rs1800497 was not as strong as previously reported. Because polygenic interaction has been suspected for an increased susceptibility of drug addiction, further analysis of multiple gene variants paired with the illicit stimulants are needed.

3365F

Integration of mQTL data and enhancer-promoter interactions with GWAS summary results identifies novel schizophrenia-associated genes. C. Wu, W. Pan. Division of Biostatistics, University of Minnesota, Minneapolis, MN.

Although genome-wide association studies (GWAS) have identified thousands of variants associated with complex diseases, most identified variants are located outside gene coding regions and a biological interpretation of their function is largely unknown. By noting that enhancer-promoter interactions play important roles in regulating gene expression, a new gene-based method called E+G (Wu and Pan, Genetics, early online, 2018) has been proposed to both boost statistical power and offer biological insights by integrating enhancer-promoter interactions with GWAS association results. On the other hand, methylation quantitative trait loci (mQTLs) and expression quantitative trait loci (eQTLs) provide orthogonal ways of functionally annotating SNPs for complex traits and common diseases (Gamazon et al., Mol Psych 18, 340-346, 2013). Although some new gene-based methods, such as transcriptome-wide association study (TWAS) (Gamazon et al., Nat Genet 47, 1091-1098, 2015; Gusev et al., Nat Genet 48, 245-252, 2016), have been proposed to integrate eQTL data with GWAS results, the topic of integrating mQTL data with GWAS summary results remains largely untouched. Here, we propose integrating enhancer-promoter interactions and mQTL data with GWAS summary results to enhance mechanistic interpretability and boost statistical power. Through an application to two large-scale schizophrenia (SCZ) GWAS summary datasets, denoted SCZ1 and SCZ2, based on about 20,000 and 70,000 subjects respectively, we demonstrate that the proposed method could maintain correct Type 1 error rates and identify some novel SCZ-associated genes (containing no significant SNPs nearby). 80% of the significant genes identified by the smaller SCZ1 dataset were confirmed to be significant by the larger SCZ2 dataset, validating the high reproducibility of the proposed method. For the larger SCZ2 dataset with about 70,000 subjects, our proposed method identified 12 novel genes, such as *FAM177A1* and *CREB1*, which could not be uncovered by TWAS, E + G, and the standard gene-based method. We conclude that our proposed method is potentially useful, complementary to TWAS, E+G, and other existing standard methods.

3366W

Examining X chromosome associations in sex-specific characteristics of bipolar disorder. W.A. Jons¹, C.L. Colby¹, S.L. McElroy², M.A. Frye³, J.M. Biernacka^{1,3}, S.J. Winham¹. 1) Department of Health Sciences Research, Mayo Clinic, Rochester, MN; 2) Lindner Center of HOPE/University of Cincinnati, Cincinnati, OH; 3) Department of Psychiatry and Psychology, Mayo Clinic, Rochester, MN.

Bipolar disorder (BD) is a highly heritable genetic condition affecting 3% of males and females. However, important sex-differences exist in the characteristics of BD. For example, females with BD are at a greater risk of rapid cycling and cycle acceleration and more likely to engage in binge-eating behaviors. Conversely, males with BD more often have a comorbid substance use disorder. Despite the fact that there have been large-scale genome-wide association studies of BD, few studies have specifically explored the role of the X chromosome in BD or in its sex-related characteristics. This is in part due to the statistical challenge posed by X chromosome inactivation (XCI), an epigenetic process that silences one of the two copies of the X chromosome in females in order to have comparable gene expression to males; however, 10-15% of genes on the inactivated X chromosome are expressed (escape from XCI). In a sample of 957 patients from the Mayo Clinic Bipolar Disorder Biobank, we examined the association of over 14,000 X chromosome SNPs with sex-associated BD traits. These include bipolar illness characteristics such as presence of Type I vs. Type II disorder, rapid cycling, cycle acceleration, and increasing severity of mood episodes over time, and comorbidities such as substance use disorder, binge eating behavior, suicide attempts. For each SNP, we used a logistic regression model assuming additive genetic effects and adjusted for sex. We coded SNP effects in two ways: the Clayton approach that accounted for XCI, and the PLINK approach that assumes expression from both copies in females (escape from XCI). We then used a novel, combined statistical approach that takes into account prior knowledge about which regions of the inactivated X are silenced. No SNPs were significant after Bonferroni correction ($p < 4 \times 10^{-6}$), but yielded top SNPs of biological interest. Top SNPs include rs1013536 with rapid cycling ($p = 6.2 \times 10^{-4}$); this SNP lies in the WWC3 gene, which encodes a protein involved in the WNT signaling pathway that is implicated in neurodevelopment and BD. Additionally, two intergenic SNPs (rs6522637 and rs6418114) may be associated with risk of substance use in BD ($p < 5 \times 10^{-5}$), and are near PCDHX11, a protocadherin-encoding gene important in neural development. This work illustrates the use of our novel statistical approach to study X chromosome associations. Further work is needed to validate these findings in larger samples of BD patients.

3367T

CNV association with neurodevelopmental phenotypes in a Finnish population cohort. E.C. Saarentaus¹, M. Kurki^{1,2,3}, L. Urpa¹, A.S. Havulinna^{1,4}, M. Perola⁴, V. Salomaa⁴, M. Pelttonen⁴, O. Pietiläinen^{2,5}, M. Daly^{1,2,3}, A. Palotie^{1,2,3}. 1) Institute for Molecular Medicine Finland FIMM, Helsinki, Uusimaa, Finland; 2) Stanley Center for Psychiatric Research, The Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA; 3) Psychiatric and Neurodevelopmental Genetics Unit, Department of Psychiatry, Massachusetts General Hospital, Boston, Massachusetts, USA; 4) National Institute for Health and Welfare, Helsinki, Finland; 5) Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, USA.

Copy Number Variations (CNVs) are associated with syndromic and severe developmental disorders, such as intellectual disability, autism and schizophrenia. Commonly considered high-impact, CNVs are nevertheless observed in the general population. This presents a diagnostic challenge in evaluating their clinical impact. The size of a CNV is an important criterion in establishing the etiologic role of a variant. To investigate the effect of CNVs on general health, the risk for neurodevelopmental phenotypes, and socioeconomic and educational attainment in the Finnish population, we utilized longitudinal health record data together with genotype information from 22,996 individuals in the FINRISK population cohort. CNVs were identified from genotyped samples (Illumina HumanCoreExome DNA MicroArray) utilizing joint analysis with the PennCNV and iPsychCNV algorithms. In our preliminary analysis, carrying a large ($>1\text{Mb}$) deletion (124 cases, 0.54% of cohort) increased the risk for a neurodevelopmental or neuropsychiatric phenotype by up to half (16.1% vs 10.7%, $p = 0.059$). Compared to non-carriers, the risk ratio for intellectual disability was 7.1 for deletions and 5.6 for duplications. Self-reported education was low for 46% of large deletions carriers, compared to 33% of non-carriers ($p = 2.8 \times 10^{-3}$). Duplications (157 cases, 0.68% of cohort) were not associated with lower educational attainment (35% vs 33%; $p = 0.55$). Large deletions were also associated with lower income (OR=1.6 [1.1–2.3], $p = 0.017$). This effect was more marked for large deletions overlapping syndromic regions (OR=3.3 [1.1–11.1], $p = 0.019$), ID associated genes (OR=10.9 [2.4–100.7], $p = 1.6 \times 10^{-4}$), and genes with a high probability for loss-of-function intolerance (OR=1.9 [1.0–3.7], $p = 0.038$). 81% of both large deletion and large duplications carriers had no neurodevelopmental diagnosis or prolonged use of antipsychotic medication. We observed no difference in self-reported general health between non-carriers and carriers of any of the CNV types. This highlights a low penetrance of even large non-syndromic CNVs. In conclusion, our results show that despite a significant disease-risk burden associated with megabase-sized copy number variants in the general population, the majority of carriers have no neurodevelopmental disorder. However, a significant burden remains for lower socioeconomic status and educational attainment.

3368F

Excess of rare, inherited variation in children with developmental disorders. K.E. Samocha¹, T. Singh², P. Danecek¹, E.J. Gardner¹, G. Gallone¹, P.J. Short¹, P. Constantinou¹, H.V. Firth^{1,3}, J.C. Barrett¹, M.E. Hurles¹ on behalf of the Deciphering Developmental Disorders study. 1) Wellcome Sanger Institute, Hinxton, United Kingdom; 2) Broad Institute, Cambridge, USA; 3) East Anglian Medical Genetics Service, Cambridge University Hospitals NHS Foundation Trust, Cambridge Biomedical Campus, Cambridge, United Kingdom.

As part of the Deciphering Developmental Disorders (DDD) study, we collected exome sequences and detailed clinical phenotypes from ~13,500 children with undiagnosed developmental disorders as well as from their parents. We have identified a genetic diagnosis for ~35% of these children, with the vast majority of diagnoses being explained by a *de novo* variant. In order to understand the potential genetic contributions in the remaining cases, we sought to evaluate the burden of rare, inherited variation. We jointly called the exome data from DDD with ancestry-matched controls from the INTERVAL and UK10K cohorts (total n = 37,898). Across the ~10,000 parent-child trios in DDD, we found a significant burden of inherited protein-truncating variation compared to controls, which is concentrated in genes previously associated with developmental disorders ($p < 10^{-6}$) and constrained genes (those intolerant of protein-truncating variation; $p < 10^{-6}$). Children with parents that have similar clinical features have a disproportionately greater excess of inherited variation; conversely, children with a likely diagnostic *de novo* variant show no such burden in any gene set tested. We further investigated the distribution of rare, deleterious variants within each individual. There were significantly more case individuals than control individuals with two or more protein-truncating variants in constrained genes (chi-squared $p < 10^{-6}$); however, we found no evidence of an increase in biparental inheritance. Additionally, we explored whether the genes that harboured a burden of inherited variation were the same as those with a significant excess of *de novo* variants. Finally, we stratified cases by their polygenic risk score for educational attainment to explore the interaction between common and rare inherited variation in developmental disorders. Overall, we have identified a significant excess of rare, inherited variants in children with developmental disorders, which points to complex, oligogenic patterns of inheritance.

3369W

A multi-trait GWAS meta-analysis on refractive errors. M.S. Tedja¹, X. Han², V.J.M. Verhoeven¹, N. Eriksson³, N. Furlotte³, N. Amin¹, C.M. Van Duijn¹, S. MacGregor², C.C.W. Klaver^{1,4}, Consortium for Refractive Error and Myopia (CREAM). 1) Erasmus Medical Centre, Rotterdam, Zuid-Holland, Netherlands; 2) Queensland Institute of Medical Research, Brisbane, Queensland, Australia; 3) 23andMe, Inc., Mountain View, CA, USA; 4) Radboud Universiteit Nijmegen, Nijmegen, Gelderland, Netherlands.

Purpose: Myopia is a complex genetic trait and its morbidity will become a public health issue in the near future. A recent large GWAS meta-analysis on refractive error (RE) of the CREAM and 23andMe (N=160,420) yielded 161 independent associated loci. These loci explain only 8% of the phenotypic variance. We now increased the sample size with the UK Biobank and performed a multi trait GWAS meta-analysis on Spherical Equivalent (SphE) and Age of Spectacle Wear (AFSW) in 526,984 participants. **Methods:** We first performed a multi trait analysis of GWAS (MTAG) on Spherical Equivalent (N_{UKBioBank}=92,766, N_{CREAM}=55,352) and Age of First Spectacle Wear (AFSW; N_{UKBioBank}=350,305; N_{23andMe}=104,293) to take into account the overlapping samples in the UK Biobank (N_{overlap}=75,732). Subsequently, we selected index variants from the genome wide significant (GWS) genetic variants per trait using Plink --clump based on position (250 kb) and LD structure ($r^2 < 0.1$). Finally, we performed gene-based tests using MAGMA and DEPICT gene set enrichment and tissue enrichment on both traits. **Results:** The GWAS meta-analysis of SphE and AFSW yielded 383 and 406 loci, respectively. Combined, thus considering the overlap in loci (>92%), we found 414 loci associated with RE. MAGMA revealed an additional 134 genes significantly associated with RE. Subsequently, the DEPICT gene set enrichment for SphE yielded 303 enriched gene sets and 385 for AFSW, of which 68% overlapped. A total of 409 unique gene sets were found replicating all previously found pathways, with "abnormal retina morphology" as most significant enriched gene set in both traits ($P_{\text{SphE}} = 1.04 \text{ e-}11$; $P_{\text{AFSW}} = 9.10 \text{ e-}13$). Finally, tissue enrichment analysis revealed "Retina Sense Organ" ($P_{\text{SphE}} = 1.13 \text{ e-}7$; $P_{\text{AFSW}} = 1.15 \text{ e-}8$; both $\text{FDR} < 0.01$) as enriched tissue. **Conclusions:** This is the largest GWAS meta-analysis to date. It confirms previously proposed light dependent pathways and revealed new pathways for myopiagenesis. Future studies should focus on their functionality and impact on RE phenotypes.

3370T

Evidence for rare variants linked to myopia in Amish families. J.E.

Bailey-Wilson¹, A.M. Musolf¹, C.L. Simpson^{1,2}, L. Portas³, F. Murgia³, Q. Li¹, E. Ciner⁴, D. Stambolian⁵. 1) Computational and Statistical Genomics Branch, NIH/NHGRI, Baltimore, MD; 2) Dept of Genetics, Genomics and Informatics, University of Tennessee Health Science Center, Memphis, TN; 3) Institute of Population Genetics, CNR, Li Punti, Sassari, Italy; 4) Salus University, Elkins Park, PA; 5) Dept of Ophthalmology, University of Pennsylvania, Philadelphia, PA.

Myopia is a complex ocular disorder that affects about 25% of Americans. Because the Amish are a founder population with lower near work exposure than the general US population, families with multiple affected family members have a higher chance of containing highly penetrant variants of large effect. Here, we performed linkage of myopia with exome enriched array genotypes for 349 patients from 43 extended Amish families. Affection status was based on mean spherical equivalent in Diopters (D): affected (≤ -1 D), unaffected (≥ 0 D) or unknown (< 0 D, > 1 D). Two discrete types of two-point parametric linkage analysis were performed: variant-based and gene-based. Variant-based analysis tested for linkage between the phenotype and individual SNPs using TWOPOINTLODS while gene-based analysis tested for linkage between the phenotype and haplotypes of rare variants located within a particular gene (SEQLINKAGE and MERLIN). We assumed an autosomal dominant model with a disease allele frequency of 1% and a 90% penetrance for carriers and 10% phenocopy rate. Both the variant-based and gene-based tests identified chromosomal loci that were highly suggestive or significant. These locations varied across different families, which is not surprising for a complex trait like myopia. However, because each linkage was unique to a family and many of the linkages were present along long linked haplotypes (which would be expected in family data) it is likely that these linkage regions are carrying variants of high effect on disease risk. Our most interesting linked haplotypes were present at 1p36.22-32.2, 12q14.2-21.1, 3p14.1-q13.13, and 4p14-q12. The signal at 5p was the highest (LOD = 3.3) and was genome-wide significant (rare for a single family). Two regions have shown linkage in other studies: *MYP14* and *MYP3* at 1p36 and 12q21; thus our results serve to confirm these published linkages. All of the linked haplotypes contained rare variant(s) (minor allele frequency ≤ 0.05) that were part of the linkage. We have identified several linked regions to myopia in Amish extended families. Each family contained different linked regions and the high LOD scores within the different families suggest the possibility of highly penetrant, large risk effect genes. These regions contain good candidate genes, such as *GUCA1C* (3q13), which is known to be expressed in retinal photoreceptors. Targeted sequencing is planned on the linked haplotypes to elucidate the causal variant(s).

3371F

Gene-based analysis with multiple longitudinal glaucoma related phenotypes shows evidence for genetic factors in primary open angle glaucoma progression. A. Athanas¹, M. Christopher², Y. Choi³, A. Chan⁴, L. Zangwill⁵, J. Liebmann⁶, C. Girkin⁷, R. Feldman⁸, H. Dubiner⁹, Y. Chen⁶, K. Taylor⁶, X. Guo⁶, J. Rotter⁶, R. Ayyagari², R. Weinreb³, N. Schork^{2,3,4,5}.

1) Biomedical Informatics, University California San Diego, San Diego, CA; 2) Hamilton Glaucoma Center, Shiley Eye Institute, Department of Ophthalmology, University of California San Diego, La Jolla, California; 3) Departments of Psychiatry and Family Medicine and Public Health, University of California San Diego, La Jolla, California; 4) Human Biology, The J. Craig Venter Institute, La Jolla, California; 5) Department of Quantitative Medicine and Systems Biology, The Translational Genomics Research Institute, Phoenix, Arizona; 6) Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute and Department of Pediatrics, Harbor-UCLA Medical Center, Torrance, CA; 7) Department of Ophthalmology, University of Alabama at Birmingham, Birmingham, AL; 8) Bernard and Shirlee Brown Glaucoma Research Laboratory, Harkness Eye Institute, Columbia University Medical Center, New York, New York; 9) Ruiz Department of Ophthalmology, University of Texas Health Science Center, Houston, TX; 10) Eye Care Center Management, Inc., Marrow, GA.

Purpose. To evaluate the relationship between genetic variants obtained and multiple glaucomatous visual field progression phenotypes. **Methods.** Whole Genome Sequence (WGS) data was collected at an average read depth of 30x using the Illumina X10 technology on 1,234 glaucoma patients. We focused our analyses on a subset of glaucoma patients who had longitudinal visual field (VF) test results available (1 – 19 years) allowing us to assess glaucoma VF loss. These individuals were analyzed in two separate ancestry groups, European (N = 600), and African Ancestry (N = 634). 6.2 Million SNPs were selected for analysis based on filters for minor allele frequency, Hardy-Weinberg equilibrium, and total genotype call rate across all loci interrogated. Two measures for global VF, and each of the six Garway-Heath VF (GHVF) were chosen as primary outcome metrics. Progression for individuals with at least three measures (N=476, N=264) were calculated from least squares regression for each of the fourteen VF metrics. We tested SNP associations against disease progression using regression analysis techniques. Results from each GWAS on global and GHVF phenotypes were combined within genes and genetic regions to create region based significance scores. **Results.** In the European ancestry population four genes previously connected with glaucoma onset were associated with glaucoma progression (TLR4 P=7.74E-07, CDKN2B-AS1 P=6.33E-06, ARHGEF12 P=8.78E-05, LMO7 P = 2.58E-05). In addition, a gene linked with a related eye disease but not linked to glaucoma was also strongly associated (CNTNAP2 P=3.85E-07). In the African ancestry population several onset genes were also associated with glaucoma progression (TMTC2 P=3.04E-08, TGFB3 P=2.16E-06, CDKN2B-AS1 P=1.46E-05). CNTNAP2 (P=1.67E-06) was strongly associated with progression in the African population as well. **Conclusions.** Our data shows that known glaucoma onset genes that harbor variants are associated also linked with disease progression in addition to a wide variety of Glaucoma-related phenotypes. Two separate populations show strong signal for a novel progression candidate gene.

3372W

Rare variants associated with age-related macular degeneration in the Amish. A.R. Waksman^{1,2}, J.N. Cooke Bailey³, R.P. Igo, Jr.³, Y.E. Song³, R. Laux³, D. Fuzzell³, S. Fuzzell³, L.D. Adams⁴, L. Caywood⁴, M. Prough⁴, W.K. Scott⁴, M.A. Pericak-Vance⁴, J.L. Haines^{1,2,3}. 1) Department of Genetics and Genome Sciences, Case Western Reserve University, Cleveland, OH; 2) Institute for Computational Biology, Case Western Reserve University, Cleveland, OH; 3) Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH; 4) Hussman Institute for Human Genomics, University of Miami, Miami, FL.

Age-related macular degeneration (AMD) is a leading cause of blindness in the world. While 52 independent variants have recently been reported by the International Age-Related Macular Degeneration Genomics Consortium (IAMDGC) to be associated with AMD, about one-third of AMD heritability is still unexplained. To identify novel variants and loci for AMD, we analyzed Illumina HumanExome chip data from Amish populations in Ohio and Indiana. The Amish are members of an isolated, founder population that resulted from the settlement of a few hundred Swiss Anabaptists in North America nearly 300 years ago. Consequently, they may be enriched in genetic variation that is rare in the general Caucasian population and has been undetectable with traditional case-control genome-wide association studies. We obtained genotype data for 87 individuals with AMD, 79 individuals without AMD, and 15 related individuals with unknown AMD affection status. Of the 242,901 SNPs on the exome chip, 38,142 variants passed extensive quality control and were polymorphic in our 175 Amish samples. Using the Anabaptist Genealogy Database, we constructed a 2,118 person all-connecting path pedigree for these 175 Amish individuals that was used to calculate their expected kinship coefficients and elucidate family structure. We performed affecteds vs. unaffecteds association testing in our samples using the RObust Association-Detection Test for Related Individuals with Population Substructure (ROADTRIPS) software. We identified four variants associated with AMD at Bonferroni corrected significance ($p < 1.30 \times 10^{-6}$). They included missense variants in a member of the lipocalin family involved in small molecule transport (*LCN9*, $p = 9.54 \times 10^{-12}$), a regulator of a telomere elongation helicase (*RTEL1*, $p = 1.99 \times 10^{-8}$), and a cell growth regulator (*CGRRF1*, $p = 7.96 \times 10^{-7}$) as well as a synonymous variant in a gene that encodes a guanylate kinase-associated protein involved in protein-protein interactions at synapses and circadian entrainment (*DLGAP1*, $p = 2.31 \times 10^{-7}$). These variants are low-frequency to rare (MAFs < 0.02) in our Amish cohort as well as in the Genome Aggregation Database (gnomAD). They have not been previously associated with AMD and are not in linkage disequilibrium with any of the 52 AMD associated variants identified by the IAMDGC. The results obtained from these analyses nominate novel variants and loci for AMD in the Amish that warrant additional analyses to understand their roles in AMD etiology.

3373T

Selecting causal risk factors from high-throughput experiments using multivariable Mendelian randomization. S. Burgess^{1,2}, V. Zuber¹. 1) MRC Biostatistics Unit, University of Cambridge, Cambridge, United Kingdom; 2) Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom.

Modern high-throughput experiments provide a rich resource to investigate causal determinants of disease risk. Mendelian randomisation (MR) is the use of genetic variants as instrumental variables to infer the causal effect of a specific risk factor on an outcome. Multivariable MR is an extension of the standard MR framework to consider multiple potential risk factors in a single analysis model using genetic variants that are each associated with at least one of the risk factors, under the assumption that all causal pathways from the genetic variants to the outcome operate via the risk factors in the model. However, current implementations of multivariable MR use standard linear regression and hence perform badly with large numbers of risk factors. Here, we propose a novel approach to multivariable MR that performs risk factor selection to identify causal risk factors. Our approach uses Bayesian model averaging to quantify the evidence that a set of risk factors influences the outcome of interest. We show that our method is able to detect true causal risk factors reliably even when the candidate risk factors are highly correlated. This allows the identification of causal drivers of disease risk from high-throughput platforms that measure large numbers of related risk factors, such as -omics platforms or imaging data. We illustrate our approach by analysing publicly-available summarized data on metabolite measurements to find likely causal risk factors for age-related macular degeneration.

3374F

Association of variants on the X chromosome with age-related macular degeneration. M. Grunin, J.L. Haines, International Age-related Macular Degeneration Genomics Consortium. Case Western Reserve University, Cleveland, OH.

In genome wide association studies (GWAS), X chromosome variants have not been fully investigated. Sex-specific effects and X-specific quality control (QC) for GWAS are needed to examine these effects. Previous work identified 52 autosomal variants associated with AMD via the International Age-related Macular Degeneration Genomics Consortium (IAMDGC), but did not focus on X. Therefore, we aim to investigate variants in the X chromosome for association with AMD. We genotyped 29,269 European individuals (M/F:10,404/18,865; control/AMD:12,087/14,273) using the IAMDGC custom exome chip. X-specific QC was performed on both marker and individual level using XWAS 3.0 and Plink 1.9 before imputation with the Michigan Imputation Server using the Haplotype Reference Consortium r1.1 panel with X-specific imputation. Imputation generated a total of 1,221,623 variants for X alone. X-specific QC sensitive to gender was performed with XWAS, Plink, and R/Bioconductor post imputation for a final set of 597,585 variants. Age and informative PCs, along with subphenotypes, were used as covariates for logistic association analyses with Fishers correction. Gene and pathway analysis was performed with VEGAS, GSEASNP, DAVID, and mirPath. Via logistic association with sex correction, variants in or near the genes *SLITRK4*, *ARHGAP6*, *FGF13* and *DMD* were nominally associated with AMD ($P < 1 \times 10^{-4}$; Fishers combined-corrected). Via association testing for the subphenotypes of choroidal neovascularization (CNV) and geographic atrophy (GA), variants in *DMD* were associated with GA patients ($P < 1 \times 10^{-4}$; Fishers combined-corrected). Via gene based analysis with VEGAS, several genes were associated with AMD ($P < 0.05$, both truncated tail strength and truncated product P) including *SLITRK4* and *BHLHB9*. Pathway analysis using GSEASNP and DAVID showed genes associated with nervous system development (FDR- $P: 0.02$), and blood coagulation (FDR- $P: 0.03$). In addition, several variants in the region of a microRNA (miR) were associated with AMD ($P < 0.05$, truncated tail strength and truncated product P). Via DIANA mirPath analysis, downstream targets of these miRs show association with brain disorders and fatty acid elongation ($P < 0.05$). Therefore, X chromosome variants may show an association with AMD pathogenesis, and these variants may be linked to novel pathways. Further analysis is needed to confirm these results and to understand their biological significance and relationship with development of AMD in worldwide populations.

3375W

Discovery of three novel risk loci for age-related macular identification by genome-wide association analysis of data from the VA Million Veteran Program. R.P. Igo^{1,2}, C.W. Halliday³, D.C. Crawford^{1,4,5}, T. Hadji⁶, P. Greenberg^{7,8}, J.M. Sullivan⁹, S.J. Fliesler⁹, S.M. Damrauer^{10,11}, W.C. Wu^{12,13}, P.E. Konick¹⁴, N.S. Peachey^{2,15}, S.K. Iyengar^{1,2,4}, VA Million Veteran Program. 1) Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH; 2) Louis Stokes Cleveland VA Medical Center, Cleveland, OH; 3) Center of Innovation in Long Term Services and Supports, Providence VAMC, Providence, RI; 4) Genetics and Genome Sciences, Case Western Reserve Univ, Cleveland, OH; 5) Institute for Computational Biology, Case Western Reserve University, Cleveland, OH; 6) Ophthalmology and Visual Sciences, Case Western Reserve University, University Hospitals Eye Institute, Cleveland, OH; 7) Ophthalmology, Providence VA Medical Center, Providence, RI; 8) Ophthalmology, Alpert Medical School, Brown University, Providence RI; 9) Research Service, VA Western NY Healthcare System, Buffalo, NY; 10) Vascular and Endovascular Surgery, University of Pennsylvania, Philadelphia, PA; 11) Surgery, Corporal Michael Crescenz VA Medical Center, Philadelphia, PA; 12) Cardiology, Medical Service, Providence VA Medical Center, Providence, RI; 13) Cardiology, Alpert Medical School, Brown University, Providence, RI; 14) Psychiatry, Case Western Reserve University, Cleveland, OH; 15) Cole Eye Institute and Department of Ophthalmology, Lerner College of Medicine, Cleveland Clinic, Cleveland, OH.

Age-related macular degeneration (AMD), a leading cause of blindness, has a substantial and complex genetic burden. The Million Veteran Program (MVP) offers a unique opportunity to identify new genetic determinants of AMD. MVP has genomic data on a custom Affymetrix Biobank chip available for > 460,000 Veterans of diverse ethnic backgrounds. We used Veterans Administration (VA) electronic health records within the MVP database to identify AMD patients; (ICD9/ICD10 codes for AMD from at least two eye clinic visits within the VA system and age > 50 years) and controls (either no ICD codes for AMD or no eye clinic visits, age > 65 years). Our genomewide association study (GWAS) on 312,204 common genetic variants (minor allele frequency > 0.5%) included 17,818 AMD cases and 142,522 controls of European descent, adjusting for sex and 10 principal components. We identified 12 genomewide significant (GWS) loci reported previously (Fritsche et al. 2016, *Nat Genet* **48**, 134), and loci at *OCA2/HERC2*, *CLUL1* and *COL10A1*. The index SNP rs12913832 (dominant OR = 0.850 in "brown-eye" A allele, $p = 3.2 \times 10^{-22}$), in intron 86 of *HERC2*, is in an enhancer controlling *OCA2* expression, a major determinant of eye color. *OCA2* controls melanosome maturation. In the retinal pigment epithelium (RPE), eumelanin focuses light and detoxifies free radicals. Association at rs12913832 was abolished when self-reported eye color (blue vs. light/dark brown) was included as a covariate, but was nominally significant (OR = 0.749, $p = 0.015$) within the subset of brown-eyed Veterans. We examined 26 previously reported GWS variants in 12 candidate loci for iris color available in the Affymetrix Biobank panel, or a close surrogate ($r^2 > 0.5$). Markers in four loci—*HERC2*, *OCA2*, *TYR* and *NPLOC4/HGS*—were significantly associated with AMD after correction for multiple testing (nominal $p < 0.002$). The association at *NPLOC4/HGS*, and at one rare variant in *OCA2* (rs1800414), remained significant when either rs12913832 genotype (dominant in A) or self-reported eye color (blue vs. light/dark brown) was included in the model as a covariate. These results establish a link between blue eye color and increased AMD risk in U.S. Veterans of European descent, and also suggest roles for genes controlling pigmentation in AMD independent of iris color. For example, lighter-eyed individuals produce less eumelanin in the RPE and choroid, and focal hyperpigmentation of the RPE is associated with choroidal neovascularization.

3376T

Disorders of sexual differentiation, the implications of life at early diagnosis: Clinical case presentation of true hermaphroditism in a male XY.

F. Cuellar Lopez¹, J.M. Aparicio Rodriguez², H. Tovar Gonzalez³. 1) Pediatric Urology; 2) Genetics; 3) Psychology Hospital Para El Niño Poblano, San Andrés Cholula Puebla, MexicoO.

Children have right to benefit from the decisions taken at the appropriate time. The medical team recommend, but parents eventually or patients decide. Complete and adequate physical examination of the patient is fundamental at birth by a health care professional to carry out the diagnosis and treatment by a multidisciplinary medical team and thus offer a better quality of life, physical, social and mental to the affected patients. In this study a case of a patient with disorder of sexual differentiation of the type 46, XY true Hermaphroditism with early detection is observed, however the patient management was late as a result of bad decisions by parents and counselors. However the treatment approach was successful although it was delayed. Disorders of sexual differentiation "the sex of a person must be specified not only to the basis of a single criterion, but multiple criteria".

3377F

Association of low-frequency variants in regulatory regions with non-syndromic orofacial clefts.

J.R. Shaffer¹, J. LeClair², J.C. Carlson¹, E. Feingold¹, C.J. Buxo³, K. Christensen⁴, F.W.B. Deleyiannis⁵, L.L. Field⁶, J.T. Hecht⁷, L. Moreno-Urbe⁸, I.M. Oriol⁹⁻¹⁰, C. Padilla¹¹, A.R. Vieira¹, G.L. Wehby⁶, J.C. Murray⁶, S.M. Weinberg¹, M.L. Marazita¹, E.J. Leslie¹². 1) University of Pittsburgh, Pittsburgh, PA; 2) Boston University; 3) University of Puerto Rico; 4) University of Southern Denmark; 5) University of Colorado; 6) University of British Columbia; 7) McGovern Medical School and School of Dentistry UT Health at Houston; 8) University of Iowa; 9) Federal University of Rio de Janeiro; 10) INAGEMP (National Institute of Population Medical Genetics) Brazil; 11) University of the Philippines Manila; 12) Emory University.

Background: Genome-wide scans have been successful in identifying genetic variants associated with risk of non-syndromic orofacial clefts (OFCs), notably common variants (i.e., occurring at frequencies of 1% or higher). These studies have shown that functional OFC risk alleles tend to be located in non-coding, regulatory elements and cumulatively explain only part of the heritability of OFCs. Complementary approaches are needed to discover low-frequency variants, which may account for some of the "missing" heritability. Here we report a scan of low-frequency variants located within putative craniofacial enhancers with the goal of identifying novel sets of OFC risk variants and implicating new regulatory elements in OFC pathogenesis. **Methods:** A multi-ethnic sample including 1995 cases of cleft lip with or without cleft palate (CL/P), 221 cases with cleft palate (CP) only, and 1626 unaffected controls was used in this analysis. Putative craniofacial enhancers were identified from a literature search of ChIP-Seq studies in craniofacial tissues or cell lines. Of these, 119 putative enhancer regions had multiple low-frequency (0.01% to 1%) variants that we genotyped in participants using a custom Illumina panel. Two complementary statistical approaches, SKAT and CMC, were used to test association of the aggregated low-frequency variants across each enhancer region with CL/P and CP. **Results:** We discovered a significant association (SKAT p-value = 4.95×10^{-4}) between CP and a branchial arch enhancer, mm60, located on chromosome 3p13, near *FOXP1*. Additionally, we observed a suggestive association (CMC p-value = 0.001) between CL/P and a forebrain enhancer, hs1717, located on chromosome 9q22.33, which is near *FOXE1*. **Conclusions:** These findings demonstrate that low-frequency variants in craniofacial enhancer regions contribute to the complex etiology of non-syndromic OFCs. **Funding:** USA NIH grants R00-DE025060, X01-HG007485, R01-DE016148, U01-DE024425, R37-DE008559, R01-DE009886, R21-DE016930, R01-DE014667, R01-DE012472, R01-DE011931, R01-DE011948, U01-DD000295, R25-MD007607, K99-DE024571, S21-MD001830, U54-MD007587 and NSF grant DBI-1263020; Robert Wood Johnson Foundation AMFDP Grant 72429; Research Institute of the Children's Hospital of Colorado; Philippines Institute of Human Genetics, National Institutes of Health, University of the Philippines; Brazil FAPERJ grant E-26/20300/2017; Brazil CNPq grants 310772/2017-6, 400427/2013-3.

3378W

An evaluation of mating asymmetry estimation. *J. Hudson¹, M. Tsai^{2,3}, M.H. Roy-Gagnon³, K.M. Burkett¹.* 1) Department of Mathematics and Statistics, University of Ottawa, Ottawa, Ontario, Canada; 2) Department of Pediatrics, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan; 3) School of Epidemiology, Public Health and Preventive Medicine, University of Ottawa, Ottawa, Canada.

Population forces such as assortative mating or sex specific migration can cause mating asymmetry (MA) in human populations. If present, MA would result in differential mating type frequencies in case-parent trios, where the frequencies of the mother-father mating types differ from those of the reverse father-mother mating types. MA would thus confound tests of maternal genetic effects or imprinting effects. Therefore, a reliable estimator of MA levels is important for assessing how different levels of MA affect maternal genetics effect and imprinting tests. A mating asymmetry statistic named MaS was proposed by Bourgey et al. 2011 to estimate MA levels assuming Hardy Weinberg genotype proportions (HWP). However, MaS was shown to overestimate MA levels in small samples. We propose an alternate mating asymmetry statistic that does not assume HWP and a related likelihood ratio test (LRT) that measures deviations from expected symmetry in mate-pair frequency among the six informative mating types through the estimation of three asymmetry parameters. We use this statistic to estimate differential mating type frequencies from genome-wide single nucleotide polymorphism (SNP) in orofacial cleft case-parent trios from an international consortium (Beaty et al. 2010). We examine the effect of sample size on the statistic by using sub-samples of the data and simulated data. Upon repeated sampling of small sample sizes we note that there is little consensus between MaS values across sub-samples when the sample consists of fewer than 300 mate-pairs. The problem is also seen to a lesser extent with the LRT; it is less susceptible to bias with small sample size than the MaS. We also investigate alternate statistics that take into account sample size. Removing the HWP restrictions leads to different MaS values; however the differences do not change any conclusion of a moderate-to-high MA SNP. The LRT has no assumption of HWE and so it correlates better with the new statistic. We observed high values of the mating asymmetry statistic with non-significant likelihood ratio tests, indicating an overestimation of MA levels by the statistic in smaller samples or when allele frequencies are small. We also found that results from a maternal genetic effect test were positively correlated with the LRT. Our results highlight the importance of reliable estimation and testing of MA to understand its impact on maternal genetic effect testing.

3379T

Maternal and child folate pathway gene-environmental interaction effects on orofacial clefts. *M.C. Tsai^{1,2}, J. Hudson³, J. Little¹, K.M. Burkett³, M.H. Roy-Gagnon¹.* 1) School of Epidemiology and Public Health, University of Ottawa, Ottawa, Ontario, Canada; 2) Department of Pediatrics, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan; 3) Department of Mathematics and Statistics, University of Ottawa, Ottawa, Canada.

Orofacial clefts (OFC) are among the most common congenital malformations with multiple factors contributing to their occurrence. Several genes in the folate metabolic pathway have been found to be associated with the occurrence of OFC. However, reported effects of folate supplementation during pregnancy on OFC risk are inconsistent. This study aimed to jointly investigate the effects of child and maternal gene-environmental interaction on OFCs. We accessed data from an international consortium (Beaty et al. 2010) through dbGAP. For this analysis we included 730 European and 762 Asian case-parent trios with cleft lip with or without palate and non-missing information on maternal vitamin supplements. We selected 624 SNPs within 29 genes involved in the folate metabolism pathway, including components of cross-membrane transport, cytosolic and mitochondrial metabolism. Using log-linear models implemented in Haplin, we tested child and maternal genetic effects and their interaction with reported maternal vitamin supplement use. We used the minimum p values across the SNPs to evaluate the effects of the overall folate pathway and its subdividing components. We found significant effects of folate metabolic pathway components on OFCs (minimum p-values < 0.01). Across the two ethnicities, the cytosolic metabolism MTRR gene showed significant main genetic and gene-environmental interaction effects in both maternal and child genotypes (minimum p-values < 0.006). We also found an association with SNP rs514933, located in the folate cross-membrane transport gene FOLR2, in the Asian cohort with maternal multivitamin supplements, with relative risk estimate (95% confidence interval) of 0.55 (0.35-0.87) and 0.61 (0.61-0.96) for the child and mother genotypes, respectively. This association was not observed in those without maternal vitamin supplements (maternal and child gene-environmental interaction p values < 0.01). Our findings indicate that maternal folate supplement use interacts with the genotype of the mother and the genotype of the child in affecting risk for OFC. Using pathway-level gene-environment analysis helps consolidate the role of folate metabolism genes. .

3380F

Association between distance from hospital birth and patient's home with morbimortality outcomes in Bogotá and Cali in years 2015 - 2016. C. Tovar-Sánchez¹, G. Botta-Méndez¹, C. López-Burbano², S. Puentes-Mahecha³, J. Holguín⁴, P. Hurtado-Villa², I. Zarante¹. 1) Institute of Human Genetics, Pontificia Universidad Javeriana, Bogotá; 2) Faculty of Health Sciences, Pontificia Universidad Javeriana, Cali; 3) Secretaria de Salud de Bogotá, Congenital Malformations Surveillance Programme of Bogotá; 4) Secretaria de Salud Pública Municipal de Cali, Congenital Birth Defects Surveillance Programme of Cali.

CMSP: Congenital Malformations Surveillance Programme, CBDSP: Congenital Birth Defects Surveillance Programme, AGA: Appropriate for gestational age, SGA: Small for gestational age, WGA: weight for gestational age. **Objective:** Determine the distances from the residency to the birth hospital in patients with congenital defects and their association with WGA and mortality in the cities of Bogotá and Cali in the years 2015-2016. **Methods:** The SIVIGILA database of congenital defects was extracted for the cities of Bogotá and Cali in years 2015 - 2016. All the residence addresses of the patients were searched and confirmed using the Google Maps Software, subsequently the distance in km between the place of residence and the hospital of birth was established, the longest distance and the shortest distance in vehicle were calculated and an average of both was performed. The average distances for each city was calculated with them standard deviation, the Odds ratio was calculated for mortality and WGA according the distance. **Results:** In Bogotá from a total of 1.503 addresses, 961 were located and calculated the distance. In Cali from a total of 644 addresses, 282 were located and calculated the distance. The mean distance from the residency to the hospital in Bogotá was 12.85 ± 6.4 km, and in Cali the mean distance was 8.15 ± 3.5 km. In Bogotá 89% and in Cali 79% lived farther than 5 km from the hospital. In Bogotá was found an OR of 0.96 (CI [0.55; 1.65]) for early mortality and OR of 1.03 (CI [0.77; 1.39]) for WGA. In Cali was found an OR of 0.81 (CI [0.40; 1.62]) for early mortality and OR of 1.12 (CI [0.54; 2.30]) for WGA. **Discussion:** The average distance between the residencies to birth hospital is very high compared to the literature, where approximately 70% of the patients lived less than 5 km from the hospital. There is no association between the distance from the birth hospital to the residency and the early mortality or WGA in patients with congenital malformations in the cities of Bogotá and Cali in the years 2015-2016, however it will be interesting to determine the relationship of the distance with chronic outcomes as secondary complications. **Conclusions:** The far distances found in this study, are due to our model of patients health system assurance, among other possible explanations. Although we did not found association between distances to the hospital with mortality, it's necessary to study other outcomes, counting that our cities have heavy traffic.

3381W

Association between prevalence of Down syndrome and maternal ages in the cities of Bogotá and Cali, Colombia in the years 2001-2018. C.M. López-Burbano¹, C. Tovar-Sánchez², J. Holguín³, S. Puentes-Mahecha⁴, I. Zarante², P. Hurtado-Villa¹. 1) Faculty of Health Sciences, Pontificia Universidad Javeriana, Cali; 2) Institute of Human Genetics, Pontificia Universidad Javeriana, Bogotá; 3) Secretaria de Salud Pública Municipal de Cali, Congenital Birth Defects Surveillance Programme of Cali; 4) Secretaria de Salud de Bogotá, Congenital Malformations Surveillance Programme of Bogotá.

Objective: Determine the association between maternal age and the prevalence of Down syndrome (DS) in the cities of Bogotá and Cali between the years 2001 to 2018. Determine the temporal prevalence of DS in mothers older than 35y. **Methods:** The database of the Birth Defects Surveillance Programme of Bogotá and Cali, was used in the years 2001-2018 to determine the association of maternal ages and prevalence of DS. The patients were grouped by five-year periods and the prevalences of each group were calculated. ORs were calculated for mothers over 35 years of age. Additionally, temporary prevalence were calculated in patients over 35 years of age. All the prevalence rates $\times 10.000$ were calculated with their respective limits by Poisson Method and ORs with their CI. **Results:** 534.843 births were evaluated, 472.645 in the city of Bogotá and 62.198 in Cali in the years 2001-2018. 691 cases of DS were found, 607 in Bogotá and 84 in Cali. Was calculated a prevalence rate of DS in patients < 30 years of 6,75, in 30 - 34 years 10,90, in 35 - 39 years 35,53 and in > 40 years 122,39, with a global prevalence of $12,9 \times 10.000$ newborns. It was found for $\geq 35y$ an OR of 7,18 (CI [6,16-8,38]). Additionally was calculated the temporal prevalences of DS in $\geq 35y$ in years 2006-2018, finding no temporal variation with an average of 51.9×10.000 newborns. **Discussion:** The age of the mother is the main factor that has been shown to increase the risk of having a baby with DS. This risk increases every year, especially after the mother is 35 years. Younger women are more likely to have babies so the majority of total number of babies with DS is in women under 35 years. However, more women are waiting to have a child and births in mothers older than 35 years has tripled. The prevalence of DS in > 40 years is 19 times the prevalence of < 30 years. **Conclusions:** The global prevalence of DS found in this cohort is similar at the prevalence reported at the literature. In those studies they argued that its results are due to terminations of pregnancies and increasing prenatal diagnosis. In our country, we must focus on the search of the causes of influencing this prevalence in order to contribute with accurately information for health system planning.

3382T

Perinatal mortality associated with congenital anomalies of the central nervous system in Colombia from 2005 to 2014. *M. Sierra¹, J. Rumbo¹, A. Salazar¹, K. Sarmiento², F. Suarez³, I. Zarante^{2,3}.* 1) Facultad de Medicina. Pontificia Universidad Javeriana. Bogotá D.C., Colombia; 2) Departamento de Ciencias Fisiológicas. Pontificia Universidad Javeriana. Bogotá D.C., Colombia; 3) Instituto de Genética Humana. Pontificia Universidad Javeriana. Bogotá D.C., Colombia.

Introduction: Currently in Colombia, congenital anomalies are the second cause of death in children under one year of age. Among these, heart birth defects and Central Nervous System (CNS) anomalies are attributed with the higher mortality. **Objective:** To describe the frequency of mortality during the first year of life attributed to CNS anomalies in the Colombian territory during the period 2,005-2,014. **Methodology:** The Integral System of Social Protection (SISPRO) a national database that provides specific information related to mortality, was used for establishing children rates of death associated with anomalies of the CNS. The ICD10 nomenclature was used to group the CNS pathologies. Six age-related mortality groups were generated: mortality from neonate to 7 days of life, from 7 to 28 days, 28 days to 3 months, 3 months to 6 months and mortality after 6 months. The frequency of mortality was established for each group of age, type of anomaly, year of presentation and country geographical distribution. **Results:** 4,706 infant deaths were reported in the first year of life attributable to CNS anomalies, anencephaly was found as the leading cause of death in fetal age (51.8%) and in children under 7 days (33%). Hydrocephalus was the leading cause after 7 days of age (32.9%). Regarding georeferencing, the highest number of anomalies were registered in the capital city Bogotá (15.6%), followed by the department of Antioquia (13.6%) and Department Valle del Cauca (9.2%). **Conclusion:** neural tube defects like anencephaly are leading cause of fetal and neonatal death in Colombia, the distribution of cases affects principal cities in the country. Implications for public health, folic acid fortification, and prevention of CNS birth defects are discussed.

3383F

Changing frequencies of the main autosomal trisomies in the NIPT era: Population-based data from the Czech Republic. *A. Šipek Sr^{1,2,3}, V. Gregor^{1,2}, A. Šipek Jr⁴, J. Jirova⁵, J. Klaschka^{6,7}, M. Maly^{6,8}.* 1) Department of Medical Genetics, Thomayer Hospital, Prague, Czech Republic; 2) Department of Medical Genetics, Prenatal Sanatorium, Prague, Czech Republic; 3) Institute of Biology and Medical Genetics, 3rd Faculty of Medicine, Charles University, Prague, Czech Republic; 4) Institute of Medical Biology and Genetics, 1st Faculty of Medicine, Charles University, Prague, Czech Republic; 5) Institute for Health Information and Statistics, Prague, Czech Republic; 6) Institute of Computer Science of the Czech Academy of Sciences, Prague, Czech Republic; 7) Institute of Biophysics and Informatics, 1st Faculty of Medicine, Charles University, Prague, Czech Republic; 8) National Institute of Public Health, Prague, Czech Republic.

Introduction: Down, Edwards and Patau syndrome are the most common chromosomal anomalies diagnosed during the prenatal diagnostics. The major goal of this population-wide study was to identify any possible trends that can be associated with the gradual implementation of non-invasive prenatal genetic testing in the Czech Republic. **Methods:** The population-based data on prenatally diagnosed cases were obtained from the National Registry of Congenital Anomalies of the Czech Republic (2012-2016 time period). This registry is run by the Institute of Health Information and Statistics of the Czech Republic, the reporting is compulsory by the law and the registry has population-wide coverage. The proportion of the number of the three major autosomal trisomies was compared to the proportion of all other chromosomal aberrations that were identified during prenatal diagnostics. The annual number of live births in the Czech Republic was used as a denominator. The ratios were subjected to the logistic regression. Statistical evaluation was carried out by the statistical software R. **Results:** Down, Edwards and Patau syndromes together presented 62.6% of all prenatally diagnosed chromosomal aberrations. The overall frequency (per 10 000 live births) of major autosomal trisomies have increased from 30.86 in 2012 to 37.41 in 2016 (Poisson regression: $p = 0.014$). Also the proportion of major autosomal trisomies increased from 57.22 % (of all prenatally diagnosed aberrations) in 2012 to 66.09 % in 2016 (logistic regression: $p < 0.001$). **Discussion:** We confirmed that the overall frequency of major autosomal trisomies (and their proportion) is increasing. The possible explanations are the improving methods of prenatal diagnostics and also the increasing average maternal age in the Czech Republic. **Acknowledgements:** Supported by Ministry of Health of the Czech Republic, grant nr. AZV 17-29622A and RVO project: "Thomayerova nemocnice – TN, 00064190".

3384W

Associated anomalies in cases with anorectal anomalies. C. Stoll, Y. Alembik, B. Dott, M.P. Roth. Genetics, Faculté de Médecine, Strasbourg, France.

Anorectal anomalies (ARA) are a common type of congenital anomaly. The etiology of ARA is unclear and its pathogenesis is controversial. Cases with ARA often have other non-ARA associated congenital anomalies. The purpose of this study was to assess the prevalence and the types of these associated anomalies in a defined population. The associated anomalies in cases with ARA were collected in all livebirths, stillbirths and terminations of pregnancy during 29 years in 387,067 consecutive births in the area covered by our population-based registry of congenital malformations. Of the 202 cases with ARA, representing a prevalence of 5.21 per 10,000, 100 (49.5%) had associated anomalies. Cases with associated anomalies were further classified into groups with chromosomal abnormalities, non chromosomal recognized conditions, and non syndromic multiple congenital anomalies (MCA). There were 7 (3.3%) cases with chromosomal abnormalities including 2 trisomies 18, and 31 (15.3%) nonchromosomal recognized dysmorphic conditions including 17 cases with VACTERL association and 2 cases with OEIS sequence (Omphalocele, Exstrophy of bladder, Imperforate anus, Spinal defect). Sixty two (30.7%) of the cases had MCA. Anomalies in the urogenital, the musculoskeletal, the cardiovascular, the digestive, and the central nervous systems were the most common other anomalies. The anomalies associated with ARA could be classified into a recognizable malformation syndrome or pattern in 38 out of 100 cases (38 %). This study included special strengths: each affected child was examined by a geneticist, all elective terminations were ascertained, and the surveillance for anomalies was continued until 2 years of age. In conclusion the overall prevalence of associated anomalies, which was close to one in two cases, emphasizes the need for a thorough investigation of cases with ARA. A routine screening for other anomalies may be considered in infants and in fetuses with ARA.

3385T

Polygenic risk score predicts mild cognitive impairment and Alzheimer's disease significantly better than APOE in ADNI dataset. G. Leonenko¹, M. Shoaib², R. Pither³, R. Sims⁴, S.M. Laws⁴, O. Oshota³, P. Daunt³, A. Gibson², K. Banks³, G. Davidson³, J. Hardy², J. Williams^{1,5}, V. Escott-Price^{1,5}. 1) Cardiff University, Cardiff, United Kingdom; 2) UCL Institute of Neurology, London, United Kingdom; 3) Cytox Ltd, UK, Oxford, United Kingdom; 4) Edith Cowan University, and Cooperative Research Centre (CRC) for Mental Health, Perth, Australia; 5) Dementia Research Institute, Cardiff, United Kingdom.

Alzheimer's disease (AD) is the most common neurodegenerative disorder, affecting millions of people worldwide. The pathological process begins decades before the AD onset and so early identification of adults at elevated risk for AD is crucial. Mild cognitive impairment (MCI) is an intermediate phenotype that in some individuals progresses to AD, and may indicate the first manifestation of AD. Based on genome-wide association powerful data, polygenic risk score (PRS) analysis has been successfully applied to AD. The aim of this study is to examine whether PRS based upon AD genomic data can predict the MCI and AD in the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset. ADNI is a longitudinal study that assesses clinical, imaging and genetic data during aging in MCI, AD and control samples. 724 individuals with available genotypic information were analysed (MCI- 424, AD- 50 and 250-cognitively normal individuals), respectively. Whole-genome sequencing data were used which was preliminary filtered according to standard quality control methods. The International Genomics of Alzheimer's Project (IGAP) consortium summary statistics were used as an independent discovery dataset for the construction of PRS in the ADNI data. Prediction modelling in MCI/controls and AD/controls samples was performed taking age, sex, *APOE* *e2* and *e4* risk alleles, and PRS as predictors. The highest prediction accuracy was achieved using full model, for both AD and MCI status in the ADNI data, the area under the curve AUC=81% for AD and AUC=75% for MCI. In addition, the prediction accuracy was considerably higher when compared to the model based upon *APOE* risk alleles only for both MCI and AD (increase in AUC 15%, $p < 2 \times 10^{-16}$ and 10%, $p = 1.6 \times 10^{-4}$, respectively). The results show that PRS improves the prediction accuracy over and above *APOE*. Furthermore, the results demonstrate the ability of the AD PRS to differentiate individuals with MCI from those who are cognitively normal. As expected, however, the prediction accuracy for identifying individuals with MCI is slightly lower than for AD cases. These findings could potentially facilitate AD clinical trials.

3386F

Genome-wide causal studies on Alzheimer's disease. *H. Qin¹, R. Jiao², M. Xiong².* 1) Tulane University, School of Public Health and Tropical Medicine, New Orleans, LA; 2) University of Texas, Health Science Center, Houston, TX.

Despite significant progress in dissecting the genetic architecture of complex diseases by association analysis, understanding the etiology and mechanism of complex diseases remains elusive. Using association analysis as a major analytic platform for genetic studies of complex diseases is a key issue that hampers the theoretic development of genomic science and its applications in practice. Causal inference represents a promising avenue for dissecting mechanisms of complex diseases. Most genomic, epigenomic and image data are observational data. Many confounding variables are not or cannot be measured since performing experiments is unethical or infeasible in many cases. The unmeasured confounding variables would invalidate the most traditional causal inference analysis. As such, most genetic studies inferring causal relations must be based on observational data. Despite its fundamental role in science, engineering and biomedicine, the traditional causal inference from observational data alone is unable to identify unique causal-effect relations among variables. These "ununique" causal solutions seriously limit their translational application. In the past decade, causal inference theory is undergoing exciting and profound changes from discovering only up to Markov equivalent class to identifying unique causal structure. The newly developed causal inference principles with unique causal solutions would open up a new era of "causal revolution" and provide powerful tools for changing the paradigm of genetic and epigenetic studies from association analysis to causal inference. In this report, we will present nonlinear additive noise models (ANMs) for genome-wide causal studies (GWCS) of a qualitative trait which can uniquely discover the causal relationships. To examine the feasibility and utility of GWCS, we applied the ANM to Alzheimer's Disease (AD) dataset with 1,904 individuals and 3,482,609 common SNPs in 51,034 genes. A total of 107 causing common SNPs in 36 genes were identified, 17 of them including CLE-C16A, PCDH15, RARS, PCDHGA, GUCY1B3, and CLSTN2 were reported to be involved in AD disease in the literature. .

3387W

Change the paradigm of genetic studies of multiple traits from association analysis to causation analysis. *T. Xu, M. Xiong.* School of Public Health, University of Texas Health Science Center at Houston.

The widely used methods for dissecting genetic structures of multiple traits are multivariate association analysis. Despite significant progress in dissecting the genetic architecture of complex diseases by association analysis, the overall contribution of the new identified genetic variants is very small and a large fraction of genetic variants is still hidden. It is well known that a large proportion of association loci do not provide disease causing signals. There is also increasing evidence that a large proportion of causal loci cannot be derived from a set of association loci. There are urgent need to develop statistical methods for identifying causal loci. The purpose of this abstract is to summarize our works on developing methods for discovery of variables that cause variation of multiple phenotypes. Specifically, we utilize a causal discovery trace method to test causality between two sets of continuous variables. Trace method can be used to identify the direction of causation, if exists, of two linearly dependent multivariate variables. We propose permutation test as a possible approach to calculating P-values for testing causality between two sets of continuous variables. The developed trace method for causal analysis was applied to test causality between gene expression pathways and cognitive assessments. Data were downloaded from Alzheimer's Disease Neuroimaging Initiative (ADNI). Both simulation and application on ADNI data is provided. We also apply Hotelling's test to compare results among causation and association studies. Results were very encouraging. A total of 8 pathways were identified to cause the development of Alzheimer's disease.

3388T

Martingale residuals of Cox model reveal genomic associations with lifespan. P.K. Joshi¹, P.R.H.J. Timmers¹, D.W. Clark¹, T. Esko², Z. Kutalik³, K. Fischer², J.F. Wilson¹. 1) University of Edinburgh, Edinburgh, Scotland, United Kingdom; 2) Institute of Social and Preventive Medicine, University Hospital of Lausanne, Lausanne, Switzerland; 3) Estonian Genome Center, University of Tartu, Tartu, Estonia.

Introduction: Cox models¹ are the natural framework for measuring the effects of genetic variants on survival, but also provide a potential to improve disease onset modelling over logistic case-control methods. However, genomic data requires a very large number of models to be fit, causing run time issues, and specialist genomic software such as Bgenie², and BOLTMM³ have not implemented Cox modelling. We have recently used martingale residuals of the cox model⁴ to convert survivorship into a quantitative trait, amenable to conventional genomic analysis and found 4 genome-wide significant associations⁵. **Aim:** to demonstrate the properties of the Martingale residuals, in the estimation of hazard ratios, in particular, their mean and variance, the consequent power and to demonstrate their application in real data. **Method:** properties such as unbiasedness of the hazard estimate, its standard error, and consequent power relative to full model fit were established by review of established literature, direct proof, and simulation of lifespan with survivorship following the Gompertz-Makeham law and proportionate hazards. **Results:** Martingale residuals must be adjusted inversely proportional to the proportion dead, and then give unbiased estimates of the hazard ratio. The variance of the residuals also depends on the proportion dead. Nonetheless, power is close to that of using full model fit. Thirty-five genome-wide significant hits for survival are obtained using this method from meta-analysis of UK Biobank and previous published data. **Conclusion:** Martingale residuals of the cox model are the natural framework for modelling disease onset and survival in large scale genomic data. 1 Cox, D. R. Regression Models and LifeTables. J. R. Stat. Soc. Ser. B Stat. Methodol. 34, 187 (1972) 2 Bycroft, C. et al. Genome-wide genetic data on ~500,000 UK Biobank participants. bioRxiv (2017). 3 Loh, P. R. et al. Efficient Bayesian mixed model analysis increases association power in large cohorts. Nat Genet 47, 284290, doi:10.1038/ng.3190 (2015) 4 Therneau, T. G., PM; Fleming, TR. Martingale Based residuals for Survival Models. Biometrika 77, 147160 (1990) 5 Joshi, P. K. et al. Genomewide metaanalysis associates HLA-DQA1/DRB1 and LPA and lifestyle factors with human longevity. Nat Commun 8, 910, doi:10.1038/s41467017-009345 (2017).

3389F

Gene methylation co-regulation network analysis of mortality in the elderly. J.B. Lund¹, J. Baumbach², K. Christensen¹, Q. Tan^{1,3}. 1) Department of Public Health, Unit of Epidemiology and Biostatistics, Odense, Denmark, Denmark; 2) Wissenschaftszentrum Weihenstephan, Technical University of Munich, Freising-Weihenstephan, Germany; 3) Unit of Human Genetics, Department of Clinical Research, University of Southern Denmark, Odense, Denmark.

Mortality is a hot topic within genetics and epidemiology and is especially a focus in older population studies. While remaining blurry, today's technologies enable us to explore much of these uncharted lands of mortality in greater detail and find causes of mortality through means of epigenetics e.g. histone modifications or methylation-based data. In this study, we explore methylation data and aims to find mortality associated gene-clusters to help shed light on mortality in the older populations. Based on large-scale mortality data on the Lothian birth cohorts of older people (LBC1921, LBC1936, N=1,425), we summarized DNA methylation levels at the promoter regions (TSS200, 1stExon, TSS1500) collected using Illumina 450K bead chip arrays. We then perform weighted-gene methylation correlation network analysis, in order to elucidate gene-clusters with significant mortality association, utilizing Cox proportional hazards models and the clusterProfiler R-package. By analysis of gene promoter methylation levels, 27 gene-modules were discovered whereof 19 significant gene-modules ($p < 0.05$, $p = 0.00014$ to 0.044) were identified through means of their eigengene (first principal component) with respect to mortality. The overall module significance found using Kruskal Wallis Test reported a P-value of $3.5E-282$. Using the top genes ranked by Cox models, with criterion $p < 0.05$ ($N=3,535$), we filtered modules based on intramodular connectivity correlated with module membership ($p < 0.05$) and minimum size of 20 genes, yielding a total of 12 gene clusters. Processing each gene cluster by dividing the cluster into sets of positive and negative associated genes based on hazard ratio with respect to mortality, based on previously mentioned Cox models. We applied each set to clusterProfiler for the discovery of KEGG pathways, Gene Ontology (GO) terms and Gene Set Enrichment Analysis (GSEA). Among the significantly ($FDR < 0.05$) revealed items are several pathways related to gene ontologies, extracellular matrix, and intra/extracellular cell parts; And gene sets involved in diabetes among other things. Examining gene promoter methylation enables the possibility of gene co-regulation analysis and its associated functions. We detected a rich flora of highly significant gene functions and pathways in older people with high correlation between gene items and mortality.

3390W

Association of APOE haplotype with cognitive function in a healthy elderly cohort. M. Riaz¹, R. Sebra², J. Ryan¹, A. Huq^{1,3}, R. Wolfe¹, R. Woods¹, E. Storey¹, RajC. Shah⁴, A. Murray⁵, E. Schadt⁴, J. McNeil⁴, P. Lacaze⁵, ASPREE Investigator Group, ASPREE Healthy Ageing Biobank. 1) Department of Epidemiology and Preventive Medicine, School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia; 2) Icahn Institute and Dept. of Genetics & Genomic Sciences at Mount Sinai School of Medicine, New York, NY, United States; 3) Department of Genetic Medicine, Royal Melbourne Hospital, Melbourne, Australia; 4) Department of Family Medicine and Rush Alzheimer's Disease Center, 2468Rush University Medical Center, Chicago, IL, USA; 5) Berman Center for Outcomes and Clinical Research, Minneapolis Medical Research Foundation, Minneapolis, MN, USA.

ASPIRIN in Reducing Events in the Elderly (ASPREE) was a randomized clinical trial of daily low-dose aspirin for primary prevention in 16,703 Australians aged over 70 years and 2,411 Americans aged over 65 years. ASPREE examined the impact of aspirin on a range of clinical outcomes over a five-year period and is now entering a new phase as a longitudinal observational study of healthy ageing, with follow up continuing for a planned further 6.5 years. At baseline, median age of ASPREE participants was 74 years with all confirmed to have no serious CVD history, no prior diagnosis of Alzheimer's disease (AD) and Modified Mini-Mental State Examination (3MS) score ≥ 78 . Approximately 15,000 ASPREE participants provided consent and biospecimens for genetic analysis through ASPREE Healthy Ageing Biobank. To date, we have performed DNA sequencing on 11,541 samples using a targeted 'super panel' covering > 750 genes used in clinical testing including pan-cancer, cardiovascular and neurological gene coverage. Here we present genotyping results for *Apolipoprotein E* (*APOE*) gene, a major cholesterol carrier in the brain and strong genetic risk factor for AD. We stratified on *APOE* genotype; $\epsilon 3/\epsilon 3 = 7127$ (61.7%), $\epsilon 3/\epsilon 4 = 2448$ (21.2%), $\epsilon 2/\epsilon 3 = 1479$ (12.8%), $\epsilon 2/\epsilon 2 = 64$ (0.6%), $\epsilon 1/\epsilon 3:\epsilon 2/\epsilon 4 = 254$ (2.2%) and $\epsilon 4/\epsilon 4 = 169$ (1.5%). We identified 64 $\epsilon 2$ homozygotes and 169 $\epsilon 4$ homozygotes, both rare in the population. We observed slightly higher learning and memory measures (HVLT-R) in $\epsilon 2/\epsilon 2$ and $\epsilon 2/\epsilon 3$ (HVLT-R mean = 8.0, SD = 2.8) compared to $\epsilon 3/\epsilon 3$ (HVLT-R mean 7.0, SD = 2.8) which remained after controlling for smoking, gender, age, body mass index, years of education, LDL and alcohol use ($P = 0.02$, adjusted difference in means = 0.17, 95% confidence interval (CI) = 0.02 to 0.32). We investigated 46 *APOE* $\epsilon 4/\epsilon 4$ individuals aged > 75 years and observed a similar difference in episodic memory (HVLT-R mean = 6.0, SD = 3.0) compared to 2871 $\epsilon 3/\epsilon 3$ at age > 75 years (HVLT-R mean = 7.0, SD = 2.9) which was explained by controlling for all above mentioned covariates ($P = 0.11$, adjusted difference in means = -0.67, 95% CI = -1.50 to +0.15). Whole genome sequencing (WGS) of *APOE* $\epsilon 4/\epsilon 4$ and $\epsilon 2/\epsilon 2$ is underway to accompany the extensive longitudinal clinical and phenotypic data already available and that will be further extended in future. This uniquely ascertained and well-phenotype elderly population provides a significant opportunity to further examine the role of *APOE* in ageing and health.

3391T

Integrative network analysis identifies relationships between metabolomics, genomics, and risk factors for Alzheimer's disease. B.F. Darst¹, Q. Lu², S.C. Johnson^{3,4,5}, C.D. Engelman^{1,3,5}. 1) Department of Population Health Sciences, University of Wisconsin School of Medicine and Public Health, Madison, WI; 2) Department of Biostatistics & Medical Informatics, University of Wisconsin, Madison, WI; 3) Wisconsin Alzheimer's Institute, University of Wisconsin School of Medicine and Public Health, Madison, WI; 4) Geriatric Research Education and Clinical Center, Madison, WI; 5) Alzheimer's Disease Research Center, University of Wisconsin School of Medicine and Public Health, Madison, WI.

Integrative multi-omics approaches could be useful for the identification of underlying mechanisms that contribute to complex diseases, such as Alzheimer's disease (AD). We performed an integrated network analysis to investigate relationships between genomics, metabolomics, and AD risk factors using the Wisconsin Registry for Alzheimer's Prevention (WRAP). Analyses included 1,111 Caucasian participants with whole blood expression for 11,376 genes (imputed from genome-wide genotyping using PrediXcan), 1,097 fasting plasma metabolites, 364 fasting cerebral spinal fluid metabolites, and 19 AD risk factors. After adjusting each of the 12,856 variables for potential confounders, residuals were used to test all 82,631,940 possible pairwise Spearman correlations. Inter-omic correlations meeting a Bonferroni-adjusted $P < 6.1 \times 10^{-10}$ were used to develop an undirected graphical network, which included 908 edges (correlations) and 532 nodes (variables). Although no genes were directly correlated with AD risk factors, many genes were indirectly linked to AD risk factors through plasma metabolites. One correlation between *CPS1* and glycine has been reported in numerous previous investigations, as have correlations between glycine, body mass index (BMI), waist-hip ratio (WHR), and homeostatic model assessment for insulin resistance (HOMA-IR), which were also identified in our network. Mediation analyses revealed that although direct associations between *CPS1* and these three risk factors could not be detected, there were strong mediation effects such that lower *CPS1* expression led to higher plasma glycine levels, which led to lower BMI, WHR, and HOMA-IR (unadjusted $P < 2.0 \times 10^{-16}$ for each of the three outcomes). Thus, *CPS1* may have protective effects for certain AD risk factors. These results demonstrate the feasibility of performing integrative omics analyses and the potential utility of other hypotheses generated by this network.

3392F

Inferring the rate of aging from quantitative traits using machine learning methods. J. Ding¹, E.D. Sun¹, T.J. Butler¹, G.R. Abecasis², F. Cucca³, I.G. Goldberg¹, L. Ferrucci¹, D. Schlessinger. 1) National Institute on Aging, 251 Bayview Boulevard, Suite 100, Baltimore, MD; 2) Department of Biostatistics and Center for Statistical Genetics, University of Michigan Ann Arbor, MI; 3) Institute of Genetic and Biomedical Research, National Research Council, Monserrato, Cagliari, Italy.

It is widely thought that individuals age at different rates. A method that measures physiological age independently of chronological age could therefore be a first step to understand relevant mechanisms; but searches for individual biochemical markers of physiological age have had limited success. In this study, we assessed the extent to which an individual's physiological age could be determined as a composite score inferred from a broad range of biochemical and physiological data. Data were collected in population studies in Sardinia ("SardiNIA") and Tuscany ("InCHIANTI"). We used machine learning strategies on data for ~6,000 Sardinian participants, who ranged in age from 12 to 81. The best predictive models were determined from multiple combinations of dimensionality reduction, classification, and regression algorithms. They reached very strong correlations ($R > 0.9$) between predicted and actual ages, and showed relative stability in successive visits of the same individuals ($R > 0.5$). Similar results were also seen for participants in the InCHIANTI study. We then defined an Effective Rate of Aging (ERA) for each participant, a continuous trait measured as the ratio of an individual's predicted age to his/her chronological age. The inference that individuals have a characteristic rate of aging is supported by findings that in the SardiNIA cohort, the inferred values of ERA showed genetic heritability of 40%. This has been sufficient to initiate genome-wide association studies that identify genetic variants influencing the rate of aging in genes that affect telomere length and metabolic activity.

3393W

Testing for pleiotropy using functional linear regression: An application with neurodegeneration endophenotypes. A.J. Dugan, D.W. Fardo, O.A. Vsevolozhskaya. Department of Biostatistics, University of Kentucky, Lexington, KY.

Because the brain directly affects cerebrospinal fluid (CSF), CSF measures are promising biomarker endophenotypes for neurological diseases such as Alzheimer's disease (AD) (Kauwe et al., 2014). Several CSF analytes associated with AD neuropathology also have genetic associations and are known to be highly correlated (Kim et al., 2011). This apparent pleiotropy is of interest. Existing methods for the joint testing of multiple related phenotypes only allow for the testing of single variants. We propose using penalized splines to fit a smooth function, $G(t)$ (where t indexes a genetic variant's position over a genetic region), to the SNP-level allele counts over genetic regions. When this smoothed function, $G(t)$, is treated as the outcome in a reverse-regression model, pleiotropy may be tested by including multiple phenotypes in the model as predictors. We propose the use of simultaneous confidence bands to allow for the testing of associations at both the genetic-region and SNP levels without the need to correct for multiple testing. We will apply this approach to data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) to test for pleiotropy with CSF analytes over genetic regions known to be associated with AD.

3394T

A fast family-based quantitative trait association test for nuclear and extended pedigrees for the analysis of whole genome sequence data. Z. Zhang, S. Leal. Baylor College of Medicine, Houston, TX.

Whole genome sequence data is being used to elucidate the genetic etiology of complex traits. One strategy is to perform association study using families, because pathogenic variants within families are likely to have larger effect sizes than those for sporadic cases. Although there are several family-based association methods for binary traits, there are only a couple for analyzing quantitative traits, e.g. famSKAT and FFBSKAT. We recently extended the Rare variant (RV)-generalized disequilibrium test (GDT) to analyze quantitative traits RV-QGDT. This method analyzes rare variants in aggregate and can be used for any pedigree type including with missing data. We evaluated speed, type I error and power for nuclear and extended pedigrees (10 members). We found that RV-QDGT is five times faster than famSKAT and comparable to FFBSKAT. FFBSKAT and famSKAT give almost identical results, because they implement the same algorithm. We evaluated power for $\alpha=2.5 \times 10^{-6}$ for a variety of pedigree types. For 1000 sib-pairs power was 52.5% (FFBSKAT) and 69.2% (RV-QGDT) for no missing data and when 50% of the founders are missing sequence data power is >15% higher for RV-QGDT than FFBSKAT. For 1000 multiplex families, the power gain is 6% for RV-QGDT compared to FFBSKAT when none of the family members are missing data and increases to 12.5% when 50% of the parents or grand-parents are missing their sequence data. We will further demonstrate the power and utility of RV-QGDT method using simulated and late-onset Alzheimer's disease families, to elucidate genetic factors underlying complex disease etiology.

3395F

Fine-mapping and colocalization at the *IL1RL1* locus to identify asthma susceptibility effector genes. S. Ghosh¹, D.N. Mayhew², Q. Lu³, J.D. Hoffman¹.

1) Department of Genetics, GlaxoSmithKline, Collegeville, PA; 2) Department of Computational Biology, GlaxoSmithKline, Collegeville, PA; 3) Target and Pathway Validation, GlaxoSmithKline, Collegeville, PA.

Many genes are associated with asthma susceptibility. Of importance are the dual, strong associations in both *IL33* ligand (chr 9p24.1) and its receptor sub-unit (*IL1RL1*) regions (2q12.1). We performed extensive genetic fine mapping around the *IL1RL1* locus using UK Biobank phenotype and genotype data to identify asthma susceptibility effector genes as well as relationships to other markers of asthma susceptibility, peripheral eosinophil and neutrophil counts at this locus. Asthma cases and controls were defined using the nurses-interview questionnaire (Field 20002) and HES ICD10 codes falling under the J45 tree. Neutrophil and eosinophil counts on all UK Biobank samples underwent a residual rank-normalization process for downstream analyses. Asthma, eosinophil, and neutrophil genome-wide association studies were carried out using BOLT-LMM. Post GWAS, we carried out approximate conditional analysis using GCTA for each phenotype. Colocalization was then performed for each asthma signal with GTEx version 7 lung eQTL data. Colocalization was additionally completed across asthma, eosinophil and neutrophil count signals. We identified 64k asthma cases and 424k controls and 436,579 subjects with both neutrophil and eosinophil counts for analysis. Approximate conditional analysis identified two conditionally distinct independent associations with asthma led by index variants rs10197862 (1.1×10^{-73}) and rs10179654 (1.31×10^{-15}). For eosinophil counts, we observed 5 conditionally distinct GWAS significant associations with index variants rs9807989, rs11676124, rs1041973, rs1921622, rs6543122. In the neutrophil GWAS we observed one signal with index variant rs13019066. We observed strong colocalization of GTEx lung tissue eQTLs for *IL1RL1* and *IL18R1* with the rs10179654 conditional association in asthma ($H_4 > 0.90$). We also observed colocalization of one of the eosinophil count signals lead by rs1041973 with the rs10179654 lead asthma variant. We do not observe colocalization of neutrophils with either asthma association. Our results highlight 2 independent genetic associations for asthma and colocalization of one of the associations with both *IL1RL1* and *IL18R1* expression as well as eosinophil counts. However, our findings suggest that the asthma signals are independent of the neutrophil associations at the locus and only partially overlapping with the eosinophil signals. Further dissection of the secondary asthma signal in asthma susceptibility is ongoing.

3396W

Constructing homogeneous subgroups for phenome-wide association studies via the K-means algorithm.

A.R. Hsieh¹, C.J. Lee², C.S.J. Fann². 1) Graduate Institute of Biostatistics, China Medical University, Taichung, Taiwan; 2) Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan.

The accelerated growth of genomic and phenotypic data has provided an unprecedented opportunity for studying complex trait genetics. Phenome-wide association study (PheWAS) is a genotype-to-phenotype approach to investigate the association between a genetic variant and a wide spectrum of phenotypes, i.e. phenome. A major challenge of PheWAS is the endogenous heterogeneity of controls due to shared biological processes underlying phenotypes and high dimensionality of datasets consisting of genotypic data and a plethora of biological measurements. One widely used dimension reduction approach is defining case/control status by grouping and counting ICD-9 codes. However, co-morbid diseases can share similar defective biological processes and thus a similar genetic background. These confounding factors are poorly accounted for when using ICD-9 criteria in choosing cases and controls. Thus heterogeneity in control groups is highly likely to dilute statistical power when performing PheWAS. Here we propose a strategy based on machine learning method to reduce the heterogeneity problem by considering both disease status and other clinical manifestations. Our proposed strategy combines a synthetic oversampling technique with K-means clustering to divide samples into subgroups. For controls, we classify these individuals by the similarity in their phenome via a K-means clustering algorithm. The cases are synthetically oversampled to produce a balanced dataset i.e. equal numbers of cases and controls. When performing PheWAS, controls are selected only if they are not in the subgroup where the phenotype of interest is included, i.e. case-control PheWAS comparisons will be carried out within each given phenotype class. An automation procedure is created from clustering steps to oversampling when different phenotypes are analyzed in PheWAS. As a proof of concept, we tested our PheWAS approach on 7427 individuals from the Taiwan Biobank dataset. By using hypertension as the phenotype, we validated three out of 5 published hypertension-associated SNPs ($p < 0.05$). In comparison, none of these SNPs were found by using original data. The results show that our proposed strategy can improve the performance of PheWAS. Our findings highlight a new way to mine BIG DATA that can help to uncover pleiotropic genetic effects in gene mapping studies.

3397T

Replication of prior GWAS catalog variants within the eMERGE Network PGRNseq Cohort.

I.B. Stanaway¹, J.E. Jackson², A.S. Gordon¹, L.J. Rasmussen-Torvik³, D. Carrell⁴, J.M. Erwin¹, T.O. Hall¹, M.R. Palmer¹, E.A. Rosenthal¹, S.V. Driest⁵, C. Prows⁶, M.S. Williams⁶, S. Pendergrass⁷, E. Leary⁸, G. Hripacsak⁹, F. Mentch⁹, J. Denny⁹, H. Hakonarson⁹, G.P. Jarvik¹, D.R. Crosslin¹, eMERGE Network. 1) University of Washington, Seattle, WA; 2) Vanderbilt University, Nashville, TN; 3) Northwestern University, Chicago, IL; 4) Kaiser Permanente, Seattle, WA; 5) Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 6) Geisinger Research, Rockville, MD; 7) Marshfield Clinic, Marshfield, WI; 8) Columbia University, New York, NY; 9) Children's Hospital of Philadelphia, Philadelphia, PA.

Electronic medical records (EMR) provide a mechanism to systematically evaluate GWAS significant phenotype-SNP pairs for replication. We used the PGRNseq panel of ~82 genes in 9007 participants from the Electronic Medical Record and Genomics (eMERGE)-PGx project to examine associations of 66 SNPs (MAF>0.01) with phenotypes similar to those in the NHGRI GWAS Catalog. We created a list of matched phenotypes between the GWAS catalog and eMERGE EMR for SNPs where the GWAS entry matched one or more EMR codes. For example, uric acid level is a risk factor for gout. The GWAS entries for rs2231142 and "Uric Acid" levels were matched with three EMR code descriptions containing "Gout". We considered 1021 tests with matched EMR defined case-control phenotypes for associations using logistic regression while controlling for the first two principal components of ancestry, sex, and eMERGE site. We report 5 SNPs with associations ($p < 0.05/66$ SNPs; see Table) considering the traits for each SNP are correlated, and 5 significant SNP-trait pairs, when adjusting for multiple contrasts stringently for all tests considered ($p < 0.05/1021$; **bolded**). We conclude that in this moderate sample size, related phenotypes from the EMR are associated with variants previously identified by GWAS.

eMERGE	Original GWAS	rsID	Gene	p	OR	95%CI	Cases	Controls
Gout	Uric Acid	rs2231142	ABCG2	7x10 ⁻⁹	1.7	1.4-2	592	5035
Gout2	Uric Acid	rs2231142	ABCG2	3x10 ⁻⁸	1.6	1.4-1.9	660	4967
Gout3	Uric Acid	rs2231142	ABCG2	7x10 ⁻⁷	1.8	1.4-2.3	284	5343
OCA	CVD	rs854572	PON1	4x10 ⁻⁶	1.6	1.3-1.9	251	7223
CAOwl	CVD	rs854572	PON1	1x10 ⁻⁵	1.6	1.3-2	226	7139
Alopecia	Hair	rs11150606	PRSS53	7x10 ⁻⁵	2	1.4-2.9	410	6955
PHC	Phenotypes	rs9923283	intergenic	1x10 ⁻⁴	1.5	1.2-1.8	514	7661
	Platelet Count							
Thrombocytopenia	Platelet Count	rs9923283	intergenic	4x10 ⁻⁴	1.5	1.2-1.8	415	7760
Anemias	Reticulocyte Count	rs3811445	TRIM58	7x10 ⁻⁴	0.88	0.8-0.9	2084	6091

Phenotypes Gout2=Gout and other Crystal Arthropathies Gout3=Gout Arthropathy OCA=Occlusion of Cerebral Arteries CAOwl= Cerebral Artery Occlusion w/ Infarction CVD=Cardiovascular Disease PHC=Purpura and other Hemorrhagic Conditions.

3398F

Gene-based kernel machine test for longitudinally measured binary phenotype. *W. Wu, Z. Wang.* Department of Biostatistics, Yale School of Public Health, New Haven, CT.

Gene-based tests have become popular for the identification of rare variants that are associated with disease traits. Longitudinal repeated measures have been increasingly used in Genome-wide association studies for they not only provide the temporal development of traits of interest but also increase statistical power. However, most of the gene-based tests for longitudinal phenotypes exist so far are only for continuous outcomes, and thus are not suitable for discrete outcomes which are frequently used in the study of recurrent diseases and substance abuse. In this paper, we propose a group of region-based tests that is capable for performing tests for binary longitudinal phenotypes. The tests extend the longitudinal kernel machine test (LSKAT) under the generalized linear mixed effect model framework which incorporates several random effects to take into account the within-subject correlation and temporal fixed effects to describe temporal development of traits. In the simulation studies, by comparing this test with several other existing tests, we demonstrated that this test achieves consistently high power under different settings and remain robust when the covariance structure is misspecified. We illustrate our method using data from the Framingham Heart Study to explore the association of repeated measures of BMI/obesity with rare and common variants.

3399W

Application of ComPaSS-GWAS to genome-wide association analysis of methylmalonic acid concentration, serum vitamin B12 concentration and holo-transcobalamin in the Trinity Student Study population. *C.M. Justice¹, J.A. Sabourin¹, A.J.M. Sorant¹, L.C. Brody², J.E. Bailey-Wilson¹, A.F. Wilson¹.* 1) Computational and Statistical Genomics Branch, National Human Genome Research Institute, NIH, Baltimore, MD; 2) Gene and Environment Interaction Section, National Human Genome Research Institute, NIH, Bethesda, MD.

ComPaSS-GWAS is a method based on Complementary Pairs Stability Selection that uses a regression based genome-wide association study (GWAS) on each half of a randomly split sample to identify corroboration between these two independent halves and then aggregates the results over many random sample splits. This corroboration aspect of ComPaSS-GWAS can be taken to be an alternative to an independent replication study, when no replication population is available. Furthermore, aggregating over many random splits prevents the loss of power that occurs when a sample is split only once in order to create a discovery and replication sample. A marker is considered as significantly associated if it is corroborated in GWAS at a critical value of 1×10^{-3} in both halves of the split and corroborated across both halves in at least 60 percent of the splits. ComPaSS-GWAS was applied to the Trinity Student Study (TSS) to validate the method. The TSS is composed of 2,232 healthy young Irish adults genotyped for 757,533 SNPs and phenotyped for over 30 quantitative traits [Molloy et al., 2016]. Associations between several genetic loci and methylmalonic acid concentration (MMA), serum vitamin B12 concentration (SB12) and transcobalamin bound vitamin B12 (HoloTC) have been previously published, for which the replication cohort consisted of 5,186 subjects aged > 60 years participating in the Trinity, Ulster, Department of Agriculture (TUDA) Study [Molloy et al., 2016; Velkova et al., 2017]. ComPaSS-GWAS identified all the regions (or genes) that have been previously reported. For MMA, sixty-six SNPs had a ComPaSS-GWAS score greater or equal to 0.6, while twenty-nine SNPs and twenty-six SNPs reached this threshold for HoloTC and SB12, respectively. All of these SNPs also obtained a GWAS p -value < 5×10^{-8} for the analysis of the TSS discovery cohort as a whole. There were four additional SNPs that had a GWAS p -value < 5×10^{-8} , but these did not achieve a ComPaSS-GWAS score ≥ 0.6 . However, all four of these SNPs were located in genetic regions for which other SNPs in close proximity did meet the ≥ 0.6 threshold.

3400T

Unravelling genetic regulation of Immunoglobulin G N-glycosylation and its involvement in complex traits and diseases. L. Klaric^{1,2}, Y. Tsepilov^{2,4}, M. Mangino⁵, T. Esko^{6,7,8}, P. Salo⁹, S.J. McGurnaghan¹, F. Vuckovic², H.M. Colhoun^{1,10}, M. Dunlop¹¹, M. Perola⁸, K. Fischer⁸, O. Polasek¹², J.F. Wilson^{1,13}, T. Spector⁶, Y.S. Aulchenko^{3,4}, C. Hayward¹, G. Lauc^{2,14}. 1) Human Genetics Unit, Medical Research Council, Institute of Genetics and Molecul, Edinburgh, United Kingdom; 2) Genos Glycoscience Research Laboratory, Zagreb, Croatia; 3) Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Science, Novosibirsk, Russia; 4) Novosibirsk State University, Novosibirsk, Russia; 5) Department of Twin Research and Genetic Epidemiology, King's College London, London, United Kingdom; 6) Estonian Genome Center, University of Tartu, Tartu, Estonia; 7) Broad Institute of the Massachusetts Institute of Technology and Harvard University, Cambridge, USA; 8) Division of Endocrinology, Boston Children's Hospital, Cambridge, USA; 9) Genomics and Biomarkers Unit, Department of Health, National Institute for Health and Welfare (THL), Helsinki, Finland; 10) Department of Public Health, NHS Fife, Kirkcaldy, United Kingdom; 11) Colon Cancer Genetics Group, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, United Kingdom; 12) Department of Public Health, Faculty of Medicine, University of Split, Split, Croatia; 13) Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, United Kingdom; 14) Faculty of Pharmacy and Biochemistry, University of Zagreb, Zagreb, Croatia.

Glycosylation, attachment of complex carbohydrates to surface of molecules, is a very common post-translational modification of proteins and can have a major impact on their function. Glycosylation of Immunoglobulin G (IgG) influences its interaction with effector cells of the immune system and was associated with many physiological states. Even though the main enzymes of glycosylation are well established, little is known about their genetic regulation or role in diseases. To address these questions, we performed genome-wide association analyses (GWAS) of 77 IgG N-glycosylation traits in eight European populations (discovery $N=8090$, replication $N=2388$) using HapMap2 imputed genotypes and assessed pleiotropy with complex traits and diseases using summary-level Mendelian randomisation analysis. We discovered twenty-seven significantly associated loci, 12 of which replicate previous findings, including 4 glycosylation enzymes *MGAT3*, *FUT8*, *ST6GAL1*, *B4GALT1*, while 15 are novel. We prioritised genes in associated loci based on pleiotropy with gene expression, coding region variants and enrichment in relevant gene-sets. For 9 genes we found evidence of a non-synonymous amino acid change, 4 were pleiotropic with expression in B-cells and 11 with expression in peripheral blood, while the remaining genes were selected based on the enrichment in antibody synthesis related pathways and cells. Capitalising on the omics nature of our data, we estimated the similarity of glycome-wide association estimates and proposed how associated genes connect in a functional network that regulates main glycosylation enzymes. Using summary level Mendelian Randomisation analyses we found evidence of pleiotropy with autoimmune and inflammatory diseases in six IgG N-glycosylation loci. In one locus we also showed that IgG N-glycosylation is pleiotropic with both expression of *ORMDL3*, *GSDMB* and *IKZF3* genes in B-cells and peripheral blood and risk for inflammatory bowel disease, rheumatoid arthritis, cirrhosis, asthma and allergy. With this study we expanded the network of genes involved in glycosylation of immunoglobulin G providing us with further insights how these molecules could be involved in complex human diseases.

3401F

Leveraging functional enrichment improves polygenic prediction accuracy. C. Marquez Luna¹, S. Gazal^{2,3}, P. Loh^{3,4}, N. Furlotte⁵, A. Auton⁵, A. Price^{1,2,3}, 23andMe Research Team. 1) Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA; 2) Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA; 3) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; 4) Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA; 5) 23andMe Inc., Mountain View, CA, USA.

Genetic variants in functional regions of the genome (e.g. coding, conserved, regulatory) are enriched for complex trait heritability. We developed a new method for polygenic prediction, LDpred-funct, that leverages trait-specific enrichments inferred by stratified LD score regression under the baseline-LD model (Gazal et al. 2017 Nat Genet), which includes binary functional annotations as well as continuous-valued annotations modeling LD-dependent causal architectures. LDpred-funct first analytically estimates posterior mean causal effect sizes, accounting for LD between variants and statistical noise and using per-SNP heritability inferred under the baseline-LD model as a Gaussian prior (to avoid overfitting, the baseline-LD model is fit using training data only). LDpred-funct then uses cross-validation within validation samples to regularize causal effect size estimates in bins of different magnitude, improving prediction accuracy for sparse architectures. We applied LDpred-funct to 16 highly heritable traits in the UK Biobank. We used association statistics from British-ancestry samples as training data (avg $N=365K$) and samples of other European ancestries as validation data (avg $N=22K$), to minimize potential confounding due to shared population stratification or cryptic relatedness between samples. We analyzed 6 million genotyped and imputed SNPs with minor allele frequency greater than 1% that passed quality control filters. LDpred-funct attained a +27% relative improvement in prediction accuracy across the 16 traits compared to the best method that does not use functional data (avg LDpred-funct prediction $R^2=0.173$; highest $R^2=0.417$ for height), consistent with large improvements in simulations with real genotypes and phenotypes generated under the baseline-LD model. Omitting the regularization step reduced the relative improvement to +18% across the 16 traits, which is entirely attributable to leveraging functional enrichment and LD-dependent architectures using the baseline-LD model. For height, meta-analyzing training data from UK Biobank and 23andMe cohorts (total $N=1107K$) increased the prediction R^2 to 0.430; the moderate improvement despite the much larger sample size is consistent with the 30% higher heritability in UK Biobank. Our results show that leveraging functional enrichment substantially improves polygenic prediction accuracy, bringing polygenic prediction of complex traits closer to clinical utility.

3402W

Identity-by-descent mapping in a large electronic health record-linked DNA biobank. *L.E. Petty, H. Chen, J.E. Below.* Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN.

Since the early GWAS era, there has been considerable attention given to the significant remaining disparity between the estimated heritability of many complex traits and the variance explained by known loci for these traits, or so-called "missing heritability". Large-scale DNA biobanks present a unique opportunity to bridge the gap between family- and population-based methods for gene discovery by leveraging patterns of genomic sharing due to distant relatedness. To identify the effects of genomic segments shared due to relatedness on the human phenome, we applied identity-by-descent (IBD) mapping in BioVU, Vanderbilt University Medical Center's large ($n > 280k$) DNA biobank linked to extensive electronic health records. We identified identity-by-descent shared segments using GERMLINE and created clusters of three or more individuals with a shared haplotype using DASH. We then performed association testing of these clusters with phenotype. Preliminary analysis of body mass index for a subset of 6,120 individuals genotyped on the Illumina Multi-Ethnic Global Array, including 815,371 variants after QC, identified several significantly associated clusters with shared haplotypes spanning regions of the genome harboring genes known to impact body mass index (*CADM2*, *CDC20/PTPRF*, *ANGPTL4*), as well as several genome-wide significant regions that are potentially novel. These preliminary findings demonstrate the utility of this approach in a regionally ascertained DNA biobank, a population likely to be enriched for distant relatedness, both as a method for validating known gene-trait associations identified through GWAS and for discovery of novel regions that may bear less common variants that GWAS is not powered to detect. This motivates application of identity-by-descent mapping to all BioVU samples and to the other phenotypes ascertained through electronic health records. Results will be presented for the full available BioVU sample, phenome-wide.

3403T

ComPaSS-GWAS reduces the type I error rate of a quantitative trait GWAS when the normality assumption of regression residuals is violated. *J.A. Sabourin, A.F. Wilson.* Genometrics Section, CSG, NHGRI, NIH, Baltimore, MD.

When a quantitative phenotype is not adequately transformed in a regression based GWAS, the residuals of the fitted model may violate the normality assumption of the residuals and result in an inflation of the type I error rate. Previous work has shown that the amount of inflation when the normality assumption is violated is affected by at least two factors. The first is the degree of non-normality of the residuals, where more non-normality results in more inflation; the second is the minor allele frequency (MAF) of the SNP being tested, where the lower the MAF, the higher the inflation. ComPaSS-GWAS is a new method based on Complementary Pairs Stability Selection that uses a regression based GWAS aggregated over many random sample splits for internal corroboration. This approach has been shown to ameliorate the loss of power seen in a traditional two-stage GWAS approach with a single split, and have similar power to a traditional GWAS on the full sample. In this study, the type I error rate for ComPaSS-GWAS is investigated for different degrees of non-normality of the residuals for SNPs in different MAF ranges to evaluate the amount of inflation relative to a traditional GWAS. Simulations were performed based on the GAW19 data (1764 unrelated individuals, 73,488 SNPs). One thousand replicates of non-genetic traits following a standard normal distribution were simulated to identify parameters for the traditional GWAS and ComPaSS-GWAS that achieved equivalent type I error rates on common variants. One thousand replicates of non-genetic traits following a gamma distribution with shape and scale parameters of 3 and 20 respectively (moderate non-normality) and a gamma distribution with shape and scale parameters of 1.2 and 2 respectively (substantial non-normality) were simulated and used to evaluate the inflation of type I error rates for different ranges of MAF. For the gamma traits with a moderate degree of non-normality, the traditional GWAS had minor inflation in SNPs with MAF between 1% and 5% and substantial inflation for SNPs with MAF $< 1\%$ (consistent with previous findings). Results for the gamma distribution with substantial degree of non-normality were similar, but with a larger amount of inflation. ComPaSS-GWAS had only minor inflation for SNPs with MAF $< 1\%$. These results suggest that ComPaSS-GWAS is more robust to the violation of normality than the traditional GWAS.

3404F

Estimating the autocorrelation of causal SNP effect size magnitudes as a function of genomic distance. A.P. Schoech^{1,2}, L. O'Connor^{1,2}, O. Weissbrodt^{1,2}, N. Patterson², Y. Reshef^{1,2}, A.L. Price^{1,2}. 1) Harvard T.H. Chan School of Public Health, Boston, MA; 2) Broad Institute of Harvard and MIT, Cambridge, MA.

Common diseases and complex traits are highly polygenic, but the location of most causal genetic variants remain unknown. It is possible that these causal variants systematically cluster together, i.e. that causal variants tend to be located close to other causal variants. Allelic heterogeneity, i.e. multiple causal variants at a locus, as well as functional regions that are enriched for causal variants, strongly suggest that this is the case. However, it is currently unknown to what extent and on which length scale along the genome causal effect size magnitudes are correlated. We introduce a method that estimates the correlation between the squared causal effect sizes of nearby SNPs, as a function of their genomic distance. Our method is not confounded by linkage disequilibrium (LD), which causes nearby SNPs to have correlated squared marginal effect sizes due to tagging. The main idea behind the method is that the product of squared association statistics of two neighboring SNPs increases on average if the causal SNPs that they tag have correlated absolute effects. We regress the product of the squared association statistics on the sum of fourth moment LD values of the regression SNPs to other linked SNP pairs that are a specified distance apart. We show that the resulting regression coefficients can be used to infer the correlation of squared causal SNP effect sizes. We evaluated our method using LD patterns in UK Biobank and simulated traits, obtaining unbiased estimates under a range of genetic architectures. The method is computationally scalable to biobank sized data sets (hundreds of thousands of individuals and millions of SNPs) and only requires LD data and association summary statistics. We also developed a related method that obtains unbiased estimates of the correlation of *signed* SNP effects as a functions of the genomic distance between the SNPs, which occurs when one type of alleles, e.g. derived alleles, locally have systematically positive or negative effects on the trait due to natural selection. Estimating the strength and length scale of SNP effect autocorrelation along the genome can shed light on important questions about complex trait architecture and fine mapping, but the method also allows for using different distance measures than genomic distance, e.g. being located in the same topologically associated domains (TADs) or in genes that are connected in a gene network.

3405W

Using identity by descent to improve genotype imputation. M. Abney, A. El Sherbiny. Human Genetics, University of Chicago, Chicago, IL.

Genotype imputation often results in many genotypes being poorly imputed, particularly in studies where the individuals are not well represented in standard reference panels. Here, we introduce Kinpute, a novel method and software package that can dramatically improve the quality of imputed genotypes. Kinpute uses identity by descent (IBD) information---due to either recent, familial relatedness or distant, unknown ancestors---to extract additional information from the reference panel that can be used to infer genotypes in the study sample. Kinpute is designed to use a within-study reference panel and works in conjunction with other genotype imputation software to improve the quality of the imputed genotypes. In addition, Kinpute can select an optimal set of study individuals to sequence as the within-study reference panel. We ran Kinpute on a data set with many related individuals following imputation with Impute2. Across essentially all genotype classes, Kinpute raised the R^2 of the imputed dosage with the true genotype compared to Impute2. For instance, for common SNPs with high INFO score (> 0.4), for those genotypes that had low imputation certainty (maximum genotype probability of 0.9), Kinpute improved the R^2 of 0.32 from Impute2 to 0.68. This class of genotypes represented approximately 10% of all genotypes. Even for common SNPs with high INFO and high genotype certainty, Kinpute improved the 0.90 R^2 from Impute2 to 0.93 (60% of all genotypes). When IBD is present in the sample, a method such as Kinpute can substantially improve the genotype imputation of sequence data in a population that is not well represented in standard reference panels.

3406T

Estimating and accounting for unobserved covariates in high dimensional correlated data. *C.G. McKennan, D. Nicolae.* Department of Statistics, University of Chicago, Chicago, IL.

An important phenomenon in high dimensional biological data is the presence of unobserved covariates that can have a significant impact on the measured response. When these factors are also correlated with the covariate(s) of interest (i.e. disease status), ignoring them can lead to increased type I error and spurious false discovery rate estimates. There have been a number of methods developed by the statistics and genetics community to estimate and account for these latent factors in downstream analyses when samples are assumed independent, including Surrogate Variable Analysis (SVA), Removing Unwanted Variation (RUV) and Confounder Adjusted Multiple Testing (CATE). However, many datasets have complex sample correlation structure, which includes data with multiple samples from the same individual under different treatment conditions or from different tissues, as well as data from related individuals (e.g. twin studies). Unsurprisingly, standard latent factor adjustment methods tend to produce spurious and inaccurate estimates for data with non-trivial sample correlation structure. For example, methods to estimate the number of latent factors, K , like parallel analysis used in SVA and bi-cross validation used in CATE tend to estimate K to be on the order of the sample size, which drastically reduces the number of degrees of freedom and power in downstream analyses. This motivated us to develop CorrConf, a provably accurate and computationally efficient method to estimate the number and column space of unobserved covariates in data with correlated observations. Our method (to the best of our knowledge) is the first subspace estimation procedure designed specifically for data with complex correlation structure, and we prove that our estimates for the effects of interest are asymptotically equivalent to the generalized least squares estimator when the latent factors are observed. We apply our method to both longitudinal DNA methylation data to identify ancestry specific methylation patterns that are conserved from birth to early childhood, as well as gene expression data to identify anthracycline response eQTLs in cardiomyocytes.

3407F

On performing gene expression imputation and gene-level association analysis in multiethnic or admixed populations. *A.I. Konkashbaev¹, E.R. Gamazon^{1,2}.* 1) Vanderbilt University Medical Center, Nashville, TN; 2) Clare Hall, University of Cambridge, Cambridge, United Kingdom.

Transcriptome-wide association analysis, such as implemented in PrediXcan and FUSION, has been utilized to identify numerous disease-associated genes. The approach estimates the genetically determined component of gene expression using GWAS data and a transcriptome panel measured only in a reference dataset. Imputation models are generated using local genetic variation to predict the molecular traits; the genetically determined phenotype is then used to map trait-associated genes. However, the impact of ancestry on this high-dimensional analysis remains to be clarified. Here we present an approach to building imputation models in a multiethnic or admixed population. Using GTEx skeletal muscle ($n=60$) and NIGMS Human Variation Panel ($n=81$) African American data, as reference transcriptome panels, we compare several approaches – penalized regression, random forest, polygenic score – for control of type I error rate and statistical power in a heterogeneous population. We then define a statistic, analogous to the traditional fixation index F_{st} for genetic variants, that measures population differentiation at the gene-level. We provide application of our ancestry-aware models to publicly available GWAS datasets. Our study has important methodological implications for the application of transcriptome-wide association analysis to current GWAS data.

3408W

The susceptibility of genome-wide association studies to recent fine-scale population structure. K.A. Rand¹, J.M. Granka¹, E.P. Sorokini¹, M. Zhang¹, J. Byrnes¹, E.L. Hong¹, K.G. Chahine², C.A. Ball¹. 1) AncestryDNA, San Francisco, CA; 2) AncestryDNA, Lehi, UT.

Large, heterogeneous genetic datasets have identified associations of thousands of single nucleotide polymorphisms (SNPs), sometimes subtly, with hundreds of human diseases and traits. Numerous methods have been developed to account for confounding effects of population structure unavoidable in such datasets; these include linear mixed models (LMMs) and the inclusion of population-level principal components (PCs) in fixed models. However, the susceptibility of GWAS to subtle, fine-scale population structure unidentifiable from PCs or F_{ST} has not been fully studied. Here, we address whether ultra-fine scale population structure, such as from post-colonial U.S. history or recent British history (Leslie Nature 2015), could produce an inflation of false positive GWAS associations, even for common variants. To do so, we use simulated phenotype data and genome-wide SNP data from over 1 million AncestryDNA customers who have consented to research. Among these individuals, we identify recent fine-scale population structure using community detection algorithms across a network of identity-by-descent (Han Nature Communications 2017) that corresponds to history within the last several hundred years and are driven by both spatial and cultural factors such as religion. We simulate phenotypes under various null and alternative models, as well as under models where latent environmental variables associated with sub-populations affect the phenotype to varying degrees. Under both balanced and imbalanced sampling across fine-scale populations, we examine results of GWAS obtained from various methods. The susceptibility of current GWAS approaches to fine-scale structure is dependent upon the degree of genetic differentiation among identified sub-populations, as well as the degree of sampling bias. In cases where latent environmental variables affecting the phenotype are related to fine-scale population structure, accounting for such structure using LMMs markedly improves GWAS accuracy. This highlights that even when underlying confounders are unknown, accounting for fine-scale structure can lessen effects of hidden environmental covariates without compromising power. Furthermore, our results suggest that accounting for fine-scale structure is even more critical when studying rare variant associations.

3409T

Selection and cis-epistasis in human transcriptome biology. A.B. Faucon, P.D. Evans, E.R. Gamazon, X. Zhong, L.K. Davis, N.J. Cox. Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN.

A recent BioRxiv paper [1] described the observation that — for a subset of genes — the most deleterious alleles are disproportionately observed to show reduced expression. One hypothesis for such an observation is that for these genes, there are fitness consequences to the deleterious variants in heterozygotes, and consequently, deleterious alleles found on haplotypes with lower than average expression may be more likely to survive and be characterized in transcriptome studies. Deleterious alleles found on haplotypes with lower expression may be more likely to be characterized in population-based transcriptome studies (GTEx) and provide a lens through which to study epistatic selection between coding and regulatory variation in human evolution. There are a variety of corollaries to this hypothesis that might be considered as well, including some related to the expectation that deleterious alleles on highly expressed haplotypes may yield more penetrant phenotypes. Therefore, we hypothesize that deleterious alleles on highly expressed haplotypes, through increased penetrance, may provide an analytic tool through which to understand disease associations in large-scale EHR / biobanks. Conceptually, this can be understood with a simple framework of a plot with deleteriousness on the x-axis and genetically determined expression on the y-axis. The region with high deleteriousness and low expression (or low deleteriousness and high expression) is reflective of epistatic selection and can be studied in transcriptome studies. On the other hand, we can do a PheWAS of the region with high deleteriousness and high genetically determined expression to investigate disease mechanisms that may arise from "disrupting" what is "expected" in reference transcriptome panels. We used phased data from GTEx to determine coding variants on haplotypes with differential expression but focused on genetically predicted expression of the haplotypes, rather than a single eQTL. Next, for those high CADD variants with which we see the most significant enrichment/depletion, we utilize PrediXcan applied to a large-scale electronic health records tied to a biobank (BioVU) to determine associated phenotypes. We hypothesize that for genes with differential expression in tissues, we will be more likely to see more deleterious variants on low expressing haplotypes, but deleterious variants on high expressing haplotypes will be associated with more phenotype. 1. Castel et al. BioRxiv (2018): 190397.

3410F

Mixed-linear model adjustment recalibrated with LD structure improves statistical power for the detection of significant ieQTLs. *N. Patel¹, W.S. Bush¹, A.E. Fish², J.A. Capra².* 1) Institute for Computational Biology, Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH; 2) Vanderbilt Genomics Institute, Department of Biological Sciences, Vanderbilt University, Nashville, TN.

With the availability of the large-scale reference datasets containing genotype and gene expression data of hundreds of individuals, it is now possible to quantify the influence of various genetic components on gene expression. While the bulk of genetic variation in gene expression appears to be additive, several studies have investigated dominant and interaction effects. Our prior work suggests the approaches used to detect interaction eQTLs (ieQTLs) are susceptible to various statistical, biological and technical limitations, which likely limited their power and clouded their biological interpretation. We have previously described a novel approach for ieQTL analysis based on mixed linear model regression within the cis-regulatory region. When we applied this approach to the 1119 false positive interactions identified in a discovery set of 210 individuals genotyped as part of the HapMap Phase II, we were able to eliminate all but six of them. However, from simulation studies, we estimated the statistical power for this approach to be too low for interactions with medium and large effect sizes. Here, we present a modified method that accounts for the LD structure of the cis-regulatory region to remove variants near the ieQTLs that are in strong LD ($r^2 > 0.4$) when building the Genetic Relationship Matrices. This approach provides similar ability to eliminate confounded models/interactions compared to the ones obtained from a simple mixed linear model analysis for the false positives in the HapMap dataset, with a significantly improved statistical power. We also apply this method to the 3319 random interactions near 10 genes derived from the Genotype-Tissue Expression (GTEx) whole blood data set and find 18 ieQTLs to be significantly associated with the expression.

3411W

Genetic regulation of transcriptional response in African American hepatocytes upon multiple drug treatments. *M. Perera¹, Y. Zhong¹, T. De¹, Y. Xu², C. Alarcon¹.* 1) Northwestern University, Chicago, IL; 2) University of Chicago, Chicago, IL.

Drug-Drug interaction and inter-individual differences in drug response are long-standing issues in medicine and are key goals for precision medicine. However, drug perturbation experiments have been limited to immortal cell lines which do not reflect the true physiology and gene expression of the liver, a key organ for drug metabolism. Here, we genotyped 60 primary human hepatocytes derived from donors of African American (AA) ancestry and treated them with 6 known inducers of drug metabolism (Omeprazole, Phenobarbital, Dexamethasone, Carbamazepine, Phenytoin and Rifampicin). We measured the transcriptome before and after treatment. We performed pair-wise differentially expressed (DE) gene analysis and demonstrated that Phenobarbital and Omeprazole have the strongest response and Carbamazepine has the lowest response in terms of the number of DE genes. Commonly overexpressed DE genes are enriched for targets of a transcriptional factor, *TCF11*, which suggests its role as a general factor in modulating drug metabolism. DE gene analysis revealed the similarity between Phenytoin and Carbamazepine, which may result from their common therapeutic application on epilepsy. We identified genetic associations with drug-induced gene expression (di-eQTLs) and found a bimodal sharing pattern of di-eQTLs across drug treatments, with the majority of di-eQTLs found uniquely in one condition ($n = 11,809$) or common to all 7 conditions ($n = 11,601$). Genes of unique di-eQTLs to drug induction are enriched for cytochrome P450 family, suggesting differential transcriptional regulation of drug metabolism. Interestingly, *LRP12* which mediates the metabolism of lipophilic molecules is under genetic regulation only with drug treatments (not found in the untreated condition). With Meta-Tissue, we mapped eQTLs across all conditions in meta-analysis approach and identified eQTLs (e.g. rs368942766 with *DYNLRB2*) with large effect size heterogeneity ($I_2 > 15$). We finally evaluated condition-specific eQTLs with respect to their functional characteristics, genomic context, and relationship to disease-associated variants. In conclusion, this is the first study that characterizes signatures of regulatory changes that affect drug metabolism specifically in a minority population. These results demonstrate that both drug-specific and cross-drug mechanisms underlie drug metabolism induction, an important interaction pathway in pharmacology.

3412T

The impact of smoking on BMI depends on your genome. C. Amador¹, P.K. Joshi², J.F. Wilson², P. Navarro¹, C.S. Haley^{1,3}. 1) MRC Human Genetics Unit, University of Edinburgh, Edinburgh, EH4 2XU, United Kingdom; 2) Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh EH8 9AG, United Kingdom; 3) Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Edinburgh, EH25 9RG, United Kingdom.

Obesity has increased in the past decades and continues to increase almost universally. It is linked to health conditions such as type II diabetes, coronary heart disease and cancer, hence, pressure has increased to tackle a problem that has a large impact on healthcare systems. To tackle obesity efficiently, we require better understanding of its causes of variation in the population. Although obesity related traits are highly heritable (twin studies show a heritability for BMI of up to 60%) known causal variants have very small effects and account for a small proportion of the heritability. Obesogenic environments have been identified (e.g. lifestyle choices, caloric diets, physical inactivity) but their specific impact and potential interaction with genetic variation have not been properly quantified. We take a new approach and study the impact on BMI of genetic variation, jointly with two specific environmental variables (smoking and alcohol intake), and the interactions between the genome and these environments. To do this we estimate the total variance that genome-by-environment interactions (GxE) explain in a variance component analysis using similarity matrices for individuals based on their genomes and their environment (smoking and alcohol intake) and a genome-by-environment interaction similarity matrix. This method allows estimation of the effect of sharing both genes and lifestyles in addition to their independent effects. Using data from UK Biobank we estimated the heritability of BMI to be 50% and an effect of smoking on BMI of 1% of the variance. However, when including an interaction term, it explained an extra 2.5%, suggesting that smoking has a larger impact on BMI depending on the genetic variants carried. This value was larger when modelling females and males separately, with up to a 5% of the variance explained by genome-by-smoking interactions in males and 3% in females. Similar results were also observed for alcohol intake. GxE is a potential source of missing heritability and these results suggest the effect can be substantial with genome-by-environment effects explaining more variation than the corresponding environmental effects. A detailed description of these interactions can improve our understanding of trait architecture and prediction and potentially improve treatment and interventions to tackle obesity.

3413F

Phenome-wide analyses of pharmacomimetic variants in their genomic context: New opportunities for genetic validation of therapeutic targets in UK Biobank. J.D. Hoffman¹, K.J. Song¹, I. Tachmazidou¹, A. Pandey¹, M.G. Ehm¹, J.C. Whittaker¹, M.R. Nelson¹, T. Johnson¹, L.M. Yerges-Armstrong¹, R.A. Scott¹. 1) Department of Genetics, GlaxoSmithKline, Collegeville, PA; 2) Department of Computational Biology, GlaxoSmithKline, Collegeville, PA; 3) Statistical Platforms and Technologies, GlaxoSmithKline, Collegeville, PA.

The majority of drugs in clinical development fail to reach approval, with late stage failures mainly due to lack of efficacy or unforeseen adverse events. Genetic support for a target increases the likelihood of it being safe and efficacious and, as such, it is of interest to robustly evaluate genetic support for potential therapeutic targets by phenome-wide association testing of variants predicted to mimic pharmacologic target perturbation of a given target: "pharmacomimetic" (PM) variants. Such approaches enable evaluation of the primary indication hypothesis as well as possible alternative indications and adverse effects, but several downstream analyses are required to allow robust inference. Here, we evaluate the association of common and rare published PM variants for 29 targets approved or in clinical development, with up to 2413 phenotypes in 379,337 unrelated participants from UK Biobank. Where we identified associations reaching statistical significance ($p \leq 2.07 \times 10^{-5}$), we performed approximate conditional analyses and fine-mapping to evaluate the evidence for the index variant as credibly causal in its genomic context. Across the variants, we identified over 400 statistically significant associations with 157 phenotypes. In addition to replicating associations related to the primary indication, we identified associations with previously unreported traits representing potential repositioning opportunities or safety flags. For example, rs1837253 at *TSLP* associated with the primary indication of asthma, and also had a colocalized (Posterior-probability (PP)=0.99) association with COPD ($p=6.5 \times 10^{-8}$) in the same direction. Among several such examples, we found multiple associations with rs12916 (an eQTL at *HMGCR*, encoding the target of statins), apparently supporting the lipid-lowering and weight-increasing effects of statins. However, the cholesterol association ($p=3.0 \times 10^{-45}$) led by rs12916 did not colocalize with the BMI association ($p=7.25 \times 10^{-17}$; PP=0), which was led by a missense variant at the nearby *POC5* (rs2307111; $p=7.3 \times 10^{-35}$). Furthermore, rs12916 was not in the 95% credible set for the BMI association. We demonstrate the value of large-scale studies with access to complete genome-wide data to evaluate genetic evidence to support therapeutic target validation. In particular, we highlight the potential for incorrect inference without thorough evaluation of genomic context and the importance of such evaluations in future PheWAS analyses.

3414W

HLA association in Chronic Fatigue Syndrome/Myalgic Encephalopathy (CFS/ME) implicating the involvement of the immune system.

A. Lande^{1,2}, O. Fluge³, S.T. Flåm¹, O. Mella³, O.D. Saugstad⁴, T. Egeland^{2,5}, M.K. Viken⁵, B.A. Lie^{1,2,5}. 1) Oslo University Hospital, Oslo, Norway; 2) Institute of Clinical Medicine, University of Oslo; 3) Department of Oncology and Medical physics, Haukeland University Hospital, Bergen, Norway; 4) Department of Pediatric Research, Oslo University Hospital; 5) Department of Immunology, Oslo University Hospital.

Introduction Etiology and pathophysiology in Chronic fatigue syndrome/Myalgic encephalopathy (CFS/ME) are largely unknown and highly controversial. Immune dysregulations have been reported among patients, and autoimmune mechanisms have been hypothesized as being part of the etiology. We want to contribute to the evaluation of this hypothesis, by investigating Human leukocyte antigen (HLA) associations in CFS/ME. **Materials and methods** We performed high resolution HLA genotyping of classical HLA class I and II genes by next generation sequencing 309 adult, Norwegian CFS/ME patients, all diagnosed according to the 2003 Canadian criteria. 140 of the patients had severe or very severe disease. Allele frequencies for the loci *HLA-A*, *-B*, *-C*, *-DRB1*, *-DQB1* and *-DPB1* were compared to a Norwegian HLA reference panel of 4514 healthy individuals. **Results** In this material, we identified a positive association between CFS/ME and the allele *HLA-DQB1*03* (Allele frequency CFS/ME = 0.36, Allele frequency Controls = 0.32, OR = 1.20, p_{adj} = 0.032). **Discussion** Our result is in accordance with a previous study of 110 patients with Chronic fatigue immune dysfunction syndrome, where they also reported an association with *DQB1*03* (OR = 1.83). We are currently following up our analysis with higher resolution HLA genotypes and analyses of HLA haplotypes in a cohort comprising altogether 425 Norwegian CFS/ME patients, as well as stratifying the results for clinical parameters. Since HLA associations are well documented in most autoimmune diseases, a confirmed *HLA-DQB1* association points towards an involvement of the immune system in CFS/ME.

3415T

Integrative genomic analysis of pulmonary function leveraging eQTL resources from lung and whole blood.

J.N. Nguyen¹, A.B. Wyss², S.A. Christenson³, W. Tew⁴, M. Neighbors⁵, M. Grimbaldston⁶, S. Garudadr^{7,8}, S. Nerella⁹, P. Sakornsakolpat¹⁰, C.R. Farber¹¹, M.H. Cho¹², R.J. Kaner¹³, E.R. Bleeker¹⁴, D.A. Meyers¹⁵, W.K. O'Neal¹⁶, P.G. Woodruff¹⁷, L.V. Wain^{10,11}, J.I. Rotter¹², S.J. London¹², H.K. Im¹³, S.S. Rich¹⁴, D.J. Ledere^{14,15}, R.G. Barr^{14,15}, A. Manichaikul¹⁶. 1) Center for Public Health Genomics, University of Virginia, Charlottesville, VA; 2) Epidemiology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Research Triangle Park, NC; 3) Division of Pulmonary, Critical Care, Allergy, and Sleep Medicine, Department of Medicine, University of California San Francisco Medical Center, San Francisco, CA; 4) OMNI Biomarker Development, Genentech, Inc., South San Francisco, CA; 5) Cleveland Clinic Lerner College of Medicine, of Case Western Reserve University, Cleveland, OH; 6) Pulmonary and Critical Care Division, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 7) Department of Genetic Medicine, Weill Cornell Medical College, New York, NY; 8) University of Arizona College of Medicine, Tucson, AZ; 9) Marsico Lung Institute, University of North Carolina at Chapel Hill, Chapel Hill, NC; 10) Department of Health Sciences, University of Leicester, Leicester, UK; 11) National Institute for Health Research, Leicester Respiratory Biomedical Research Unit, Glenfield Hospital, Leicester, UK; 12) Institute for Translational Genomics and Population Sciences and Department of Pediatrics, Los Angeles Biomedical Research Institute at Harbor, University of California, Los Angeles Medical Center, Torrance, CA; 13) Section of Genetic Medicine, The University of Chicago, Chicago, IL; 14) Department of Medicine, College of Physicians and Surgeons, Columbia University, New York, NY; 15) Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY.

Lung function is an important measure of health and a predictor of morbidity and mortality in the general population. Genome-wide association studies (GWAS) have identified over 100 loci for lung function traits, opening the opportunity using summary-based integrative approaches to combine with publicly available eQTL data. One method is the S-PrediXcan approach. In applying such approaches, it is unclear whether to perform tissue-specific analyses using disease relevant tissues only, or if it can be beneficial to leverage eQTL resources from other tissues. We applied gene-based S-PrediXcan to integrate GWAS results for spirometric pulmonary function (FEV1, FVC, and FEV1/FVC ratio) results from the CHARGE consortium ($n > 90,000$) and UK BiLEVE ($n > 48,000$) with genetically-regulated gene expression (GReX) inferred using cis-eQTL SNP data from lung ($n = 333$) and whole blood ($n = 315$) in the Genotype-Tissue Expression (GTEx) project v7p. After combining S-PrediXcan (using a false discovery rate [FDR] < 0.05) with a Bayesian colocalization analysis for GWAS of pulmonary function traits in CHARGE, followed by replication in UK BiLEVE (Bonferroni p -value < 0.05), 74 GReX were identified that were associated with at least one lung function trait. Of the identified genes, 10 were identified through integration using both the GTEx lung and whole blood models, 29 were identified exclusively for lung tissue, and 35 were identified exclusively for whole blood. In follow-up analysis using RNA-seq on bronchial epithelial brushings from 129 participants in the Subpopulations and Intermediate Outcome Measures in COPD Study (SPIROMICS), we observed statistically significant associations of lung function with gene expression levels for 8 of the 74 genes based on a FDR q -value < 0.1 , of which 4 were identified from our integrative analyses with lung tissue (*PYGB*, *MMP15*, *DNAH5*, *TYRO3*) and 4 from whole blood (*PABPC4*, *STMN3*, *ZNF668*, *HM13*). *ZNF668*, *DNAH5*, *TYRO3*, *HM13* represent new loci, while *PYGB*, *PABPC4*, *MMP15*, and *STMN3* lie in regions previously identified by GWAS. While we identified distinct sets of genes through integration with lung vs. whole blood eQTL, equal numbers were validated through gene expression on bronchial brushings, suggesting further gains may be achieved by combining eQTL evidence across tissues.

3416F

Gene-based rare variant association tests for ancestry-matched

case-control data. C. Wang^{1,2}, B. Sun², Z. Wang³, M. Deng^{3,4,5}, H. Chen^{6,7}. 1) Department of Epidemiology and Biostatistics, School of Public Health, Tongji Medical School, Huazhong University of Science and Technology, Wuhan, Hubei, China; 2) Computational and Systems Biology, Genome Institute of Singapore, Singapore; 3) Center for Quantitative Biology, Peking University, Beijing, China; 4) LMAM, School of Mathematical Sciences, Peking University, Beijing, China; 5) Center for Statistical Sciences, Peking University, Beijing, China; 6) Human Genetics Center, Department of Epidemiology, Human Genetics and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, Texas, U.S.A; 7) Center for Precision Health, School of Public Health & School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, Texas, U.S.A.

With an increasingly large amount of human sequencing data available, analysis incorporating external controls becomes a popular and cost-effective approach to boost statistical power in disease association studies. To prevent spurious association due to population stratification, it is important to carefully match the ancestry backgrounds of cases and external controls. However, popular rare variant association tests based on logistic regression models, including the burden test, sequence kernel association test (SKAT) and the mixed effects score test (MiST), which is a hybrid version of burden and SKAT, are conservative for matched case-control samples when all matched strata have the same case-control ratio and might become anti-conservative when case-control ratio varies across strata. To account for the matching structure, we propose gene-based tests based on a conditional logistic regression (CLR) model, namely CLR-burden, CLR-SKAT, and CLR-MiST. We show that the CLR model coupled with ancestry matching is a general approach to control for population stratification. Through extensive simulations of population stratification and matching schemes, we demonstrate that both CLR-burden and CLR-SKAT are more powerful than standard burden test and SKAT respectively in ancestry-matched data while robustly controlling the type 1 error, and CLR-MiST is most powerful under a wide range of scenarios. Furthermore, because CLR-based tests allow for different case-control ratios across strata, a full-matching scheme can be employed to fully utilize available cases and controls to accelerate the discovery of disease genes.

3417W

Using maximal segmental score to estimate region-specific p-values in sequencing data for recessive diseases. C.C. Chang¹, I-B. Lian², J. Ott³, C.S.J. Fann¹. 1) Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan; 2) Graduate Institute of Statistics and Information Science, National Changhua University of Education, Taiwan; 3) Rockefeller University, New York, NY, USA.

Advances in high-throughput DNA sequencing technology have generated a great momentum for genomic data. For rare variant association test, the power for detecting low frequency or by traditional association methods is lower. Previously, we proposed Maximum Segmental Score method (MSS) that is capable of scanning genetic variants to identify regions most likely harboring disease gene with rare and/or common causal variants. The method uses aggregate scores to identify the region with possible disease association¹. If the disease-causing variant occurs with high frequency in the affected families, family studies may provide a significant boost in statistical power. One of the advantages of using samples from families of the same ethnical group is to reduce bias arisen from population stratification which can lead to false-positive results. Instead of the usual ways to do case-control studies, we compared a given case (or a few cases) with all controls and carried out such a comparison for each variant. In our method, there were many more controls than cases, therefore, permutation tests were carried out to obtain empirical p-values. We considered a sequence of p-values obtained by marker-wise tests. Euclidean distance was used to measure the difference between cases and controls and the mean values were compared between the two groups by a t statistic. Firstly, the p-values were transformed into a sequence of scores according to a specific scoring system. Secondly, MSS was applied to scan the sequence of scores to identify the region with the highest cumulative score. Finally, the tail probability was calculated. The segment with the highest score would suggest the likelihood of putative disease genes within the region. We applied our approach on two autosomal recessive diseases. Three multiple intestinal atresia cases were chosen from three families respectively and 30 controls drawn from UK10K data. The result found the TTC7A gene as being most significant. One osteogenesis imperfecta case was used along with 32 controls from UK10K data, the result showed that the BMP6 gene was most significant. Both genes had been published previously by method based on Hamming Distance Ratio (HDR)². These results demonstrate that our extended method of MSS is able to identify putative disease genes with very small numbers of cases using sequencing data for recessive diseases. 1. Ying-Chao Lin, et al. Genetic Epidemiology, 2012. 2. Atsuko Imai, et al. Scientific Reports, 2015.

3418T

Medium-coverage DNA sequencing in the design of the genetic association study. C. Xu, R. Zhang, H. Shen, H.W. Deng. Center of Genomics and Bioinformatics, Department of Global Biostatistics and Data Science, Tulane University, New Orleans, LA, 70112, USA.

DNA sequencing is a widely used tool in genetic association study. Sequencing cost remains a major concern in sequencing-based study, although the application of next generation sequencing has dramatically decreased the sequencing cost and increased the efficiency. The choice of sequencing depth and the number of samples will largely determine the final study investment and performance. Many studies have been conducted to find a cost-effective design of sequencing depth that can achieve certain sequencing accuracy using minimal sequencing cost. The strategies previously studied can be classified into two groups: 1) single-stage to sequence all the samples using either high ($>30\times$) or low ($<10\times$) sequencing depth; and 2) two-stage to sequence an affordable number of individuals at a high-coverage followed by a large sample of low-coverage sequencing. However, limited studies examined the performance of the medium-coverage ($10\times$ - $30\times$) sequencing depth for a genetic association study, where the optimum sequencing depth may exist. In this study, using a published simulation framework, we comprehensively compare the medium-coverage (MC) sequencing to the single- and two-stage high/low-depth sequencing in terms of the power and type I error of the variant discovery and association testing. We find, given certain sequencing effort, MC yields a close discovery power and better type I error control compared with the best scenarios using other design. However, MC is not as competent as other sequencing coverages with respect to the association power, especially for the rare variants and when the sequencing investment is limited.

3419F

Genetics and nicotine dependence: Heaviness of smoking index in European individuals. C. Batini, N. Shrine, Q. Wang, M.D. Tobin. Department of Health Sciences, University of Leicester, Leicester, UK.

Smoking is one of the leading risk factors of the top five causes of death worldwide, including heart and respiratory diseases, killing around 6 million people a year. Smoking behaviour is a very complex trait but, based on twin studies, it has been shown to have genetic heritability ranging between 35% and 75% depending on the specific phenotype considered, with nicotine dependence being the highest. Here we use the categorical Heaviness of Smoking Index (HSI) to quantify nicotine dependence in all UK Biobank (UKB) ever smokers for which data on both cigarettes per day and time to the first cigarette of the day were available, for a total of 34,752 UKB participants. Of these, 54% show low nicotine addiction, while high addiction was observed only in a minor group, and more frequently in men (9%) than women (4%). We present here a preliminary genome-wide association study of categorical HSI using a linear mixed model to test associations in 32,212 UKB individuals of European origin. Genome-wide significant associations ($p < 5 \times 10^{-8}$) highlight signals previously reported for other smoking behaviour traits: *DBH* for smoking cessation, *CHRNA3* for cigarettes per day and *CHRNA4* for nicotine dependence measured by the Fagerström Test. Associations passing a suggestive significance threshold ($p < 10^{-5}$) indicate novel loci on chromosomes 3, 11 and 22. Replication analyses, association testing in non-European ancestries in UKB and functional *in silico* follow-up will be carried out to help understanding the biological background of nicotine addiction and informing the identification of future drug targets for smoking cessation.

3420W

Universal Control Repository: A practical framework for aggregating control genotype data. D. Chen^{1,4}, K. Tashman^{1,4}, D. Palmer^{1,4}, K. Panayidou^{2,5}, T. Zheng³, C. Churchhouse^{1,4}, K. Roeder^{2,5}, A. Bloemendaal^{1,4}, B. Neale^{1,4}. 1) Stanley Center, Broad Institute, Cambridge, MA; 2) Department of Statistics, Carnegie Mellon University, Pittsburgh, PA; 3) Department of Statistics, Harvard University, Cambridge, MA; 4) Analytical and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA; 5) Computational Biology Department, Carnegie Mellon University, Pittsburgh, PA.

A major obstacle to the successful identification of genetic risk factors is the availability of large control samples for association studies. Genotyping is often restricted to cases to reduce costs, making analysis impossible without incorporating external controls. While many control resources are available through repositories such as dbGap, administrative and computational barriers to entry are substantial. Further, confounding due to both population stratification and technical biases across genotyping technologies require computational and statistical expertise. We build on the work of Bodea et al. AJHG (2016) to develop the UNiversal Control Repository Network (UNICORN) as a scalable framework for aggregating control samples while allowing users to perform ancestry-controlled GWAS without the exchange of individual-level genotype data. To this end, we have aggregated 600,000 control samples comprising 110,000 samples obtained from 23 collections within dbGaP, with a further 30,000 and 488,000 incoming from the PGC consortium and UK Biobank respectively. These were genotyped on a plethora of technologies including Illumina and Affymetrix arrays. At present the majority of controls are of European ancestry, but we have increasing numbers from other major ancestries, including around 40,000 African samples, 20,000 East Asian samples and 5,500 Latino samples. Merging data from multiple genotyping array platforms creates various challenges, which we address with a multi-stage pipeline. We first performed pre-imputation quality control at the cohort and platform levels. We then imputed the platform-level data using the Michigan Imputation Server, merged the results and performed further post-imputation QC. Finally, we used this merged data to identify, remove and re-impute additional poorly performing SNPs on a per-platform basis. The pipeline was implemented in Hail¹, scalable open source genetics software. This reproducible framework allows additional controls to be incorporated easily, and enables users to run the pipeline on case sets. UNICORN provides an aggregated and imputed global control resource, in which batch effects have been controlled and ancestry has been modeled. Our unified QC pipeline ensures reproducibility, extensibility and usability. We envision future routine use of this resource to boost power in case-control GWAS and to enable studies in situations where controls are lacking.

¹ <https://github.com/hail-is/hail>

3421T

Correcting index event bias in genome-wide association studies of prognosis or survival. F. Dudbridge¹, R.J. Allen¹, A.F. Schmidt^{2,3,4}, L.V. Wain¹, R.G. Jenkins⁵, A.D. Hingorani⁵, R.S. Pate⁵. 1) Department of Health Sciences, University of Leicester, Leicester, UK; 2) Groningen Research Institute of Pharmacy, University of Groningen, Groningen, The Netherlands; 3) Institute of Cardiovascular Science, Faculty of Population Health Sciences, University College London, London, UK; 4) Department of Cardiology, University Medical Centre Utrecht, Utrecht, The Netherlands; 5) Division of Respiratory Medicine, University of Nottingham, Nottingham, UK.

Following the completion of numerous genome-wide association studies of disease susceptibility, there is increasing interest in identifying genetic associations with disease prognosis, survival or other subsequent events. However, such studies are vulnerable to index event bias, by which selection of subjects according to their disease status creates biased associations if common causes of susceptibility and prognosis are not accounted for. This is a particular problem for polygenic traits and is also a potentially serious bias in Mendelian randomization studies. We propose a novel correction for index event bias using the residuals from the regression of genetic effects on prognosis on the corresponding effects on susceptibility. Assuming that direct effects on prognosis are independent of those on susceptibility, our approach yields correct type-1 error rates and unbiased estimates of prognosis effects. Under genetic correlation between susceptibility and prognosis, our approach has reduced type-1 error and bias compared to the standard unadjusted analysis. Our approach can have greatly increased power for SNPs with concordant direction of effects between susceptibility and prognosis. In a study of idiopathic pulmonary fibrosis, we resolved a paradoxical association with increased survival of the strong susceptibility gene *MUC5B*; under our adjusted analysis the risk allele had a highly significant association with decreased survival. In re-analysis of a study of Crohn's disease prognosis, our approach attenuated four genome-wide significant associations to a suggestive level. We recommend our approach as a standard analysis for genome-wide association studies of events subsequent to disease onset.

3422F

Genetic associations with sleep apnea in a European population. Y.

Huang, R. Gentleman, 23andMe Research Team. Research, 23andMe Inc, Mountain View, CA.

Background: Sleep apnea is characterized by uncontrollable pauses or repetitive periods of shallow breathing. It affects approximately 3% to 7% of adults and is associated with increased risk for diabetes, cardiovascular conditions as well as premature mortality. While obesity has been considered to be one major risk factor in the development and progression of sleep apnea, previous studies have also suggested inheritance and familial factors in both sleep apnea and obesity. The genetic basis of sleep apnea is still not yet well understood. **Objective:** To investigate associations between genetic risk loci, body mass index (BMI), and risk of sleep apnea in European population using the 23andMe database. **Methods:** We conducted a genome-wide association study using logistic regression with self-reported sleep apnea as the dependent variable. Covariates included age, sex, BMI, and ancestry-informative principal components (PC) 0-4. Multivariable logistic regression models adjusting for potential confounders, including diabetes and high blood pressure, were further used to determine the relationship between BMI and sleep apnea. **Results:** A total of 1,925,006 (case: 184,110, control: 1,740,896) Europeans who had consented to participate in research were included. Fifteen loci or genes have surpassed the genome-wide significance threshold ($P = 5 \times 10^{-8}$). rs2336715 (MSRB3, $P = 1.5 \times 10^{-12}$) and rs592333 (DLEU1-DLEU7, $P = 2.8 \times 10^{-12}$), associated with snoring in previous research, were indicated to be significantly related to sleep apnea. Although strongly associated with BMI or type II diabetes, rs531385 (SEC16B, $P = 2.9 \times 10^{-10}$) and rs17715065 (EBF1, $P = 1.2 \times 10^{-11}$) still remained at genome-level significance after adjusting for the BMI. Moreover, in the secondary analysis, people who reported being overweight (odds ratio (OR): 2.08, 95% confidence interval (CI): 2.05 - 2.12) or obese (OR: 6.00, 95% CI: 5.91 - 6.10) were more likely to have sleep apnea compared to those who were normal or underweighted after adjusting for confounders. **Conclusions:** Our findings highlighted sleep apnea susceptibility genes in the European population, especially those with higher BMI. Further research in various ancestry groups is needed to evaluate these potential relationships in more detail. .

3423W

Genetic associations to gut microbiome variation in the Flemish Gut

Flora Project. D.A. Hughes¹, J. Wang², R. Bacigalupe³, K. Wade¹, L. Corbin¹, N. Timpson¹, J. Raes³. 1) MRC Integrative Epidemiology Unit, Bristol Medical School, University of Bristol, Bristol, UK; 2) Institute of Microbiology, Chinese Academy of Science, Beijing, China; 3) VIB-KU Leuven Center for Microbiology, Rega Instituut, Leuven, Belgium.

The Flemish Gut Flora Project (FGFP) is a large, population-level, cross-sectional, microbiome collection derived from individuals living in Flanders, Belgium. Previous work from this sample has highlighted many associations between microbiota diversity and abundance to health, dietary and lifestyle choices, and stochastic biological outcomes like stool consistency. Here we expand on these observational results by testing for associations between genotype and microbiota phenotypes, including alpha and beta diversity, enterotype, as well as abundance estimates for some 88 common, taxonomic units. Given the variety and uncertainty in the field as how to best handle this zero-inflated ecological data, we explicitly evaluate a variety of normalization procedures, as well as test zero-inflated taxa in data in binary, presence absence generalized linear models. These data will be vital in (a) understanding how to best handle such complex data, (b) identifying loci that contribute to the heritability of gut microbiota flora diversity, (c) and in identifying markers that can be used in mendelian randomization analysis to infer causal relationships between microbiome diversity and other phenotypes of interest.

3424T

Genome-wide association study of longitudinal change in quantitative emphysema. W. Kim¹, M. Parker², M.H. Cho^{2,3}, E. Silverman^{2,3}, T.H. Beaty¹, COPDGene Investigators. 1) Department of Epidemiology, Johns Hopkins School of Public Health, Baltimore, MD; 2) Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, MA; 3) Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, MA.

Introduction: Genome-wide association studies (GWAS) have identified genetic factors associated with cross-sectional measures of quantitative emphysema, but the genetic determinants of longitudinal change in quantitative emphysema remain largely unknown. Our study aims to identify the genetic influences of emphysema progression measured by computed tomography (CT) imaging in current and ex-smokers from the COPDGene study, a longitudinal cohort enrolling Non-Hispanic White (NHW) and African American (AA) subjects between the ages of 45 and 80 years at baseline with a minimum 10 pack-year smoking history. **Methods:** Among subjects who completed baseline and 5-year follow-up visits, genetic and longitudinal CT data were available on 4,238 subjects (3,078 NHWs and 1,160 AAs). Two quantitative emphysema measures from CT imaging, percent emphysema and adjusted lung density, were considered for GWAS. Longitudinal change in quantitative emphysema was defined as the difference in emphysema measures between phase 1 and 2. Multiple linear regression was performed under an additive genetic model, stratified by race, using genome-wide markers imputed to the Haplotype Reference Consortium panel. Baseline measures of age, pack-years of smoking, smoking status and emphysema (percent emphysema and adjusted lung density, accordingly), sex and principal components of genetic ancestry were included as covariates. We also performed GWAS of baseline measure of quantitative emphysema to determine whether genetic determinants of change in emphysema are also associated with baseline emphysema severity. **Results:** We identified five genetic loci associated with change in percent emphysema at traditional genome-wide significance ($P < 5 \times 10^{-8}$), markers in or near *LOC100133091*, *ST6GALNAC3*, *TPK1*, and *SLC26A4* among AAs and *MAT2B* among NHWs. All associations had low minor allele frequency < 5%, and only the *MAT2B* locus passed a revised threshold for densely imputed data ($P = 5 \times 10^{-9}$). Notably, these genetic loci were not reported from previous GWAS of cross-sectional quantitative emphysema and were not significantly associated with baseline measure of quantitative emphysema. **Conclusions:** Our study identified one genetic locus significantly associated with longitudinal change in percent emphysema among NHW heavy smokers. Confirmation of our finding may help identify genes influencing progression of emphysema as distinct from genes influencing development of emphysema.

3425F

Three novel loci were identified for plasma acylcarnitine levels and their ratios in genome-wide association study. H. Li, S. Huo, X. Lin. Shanghai Institute For Biological Sciences, Cas, Shanghai, China.

Plasma acylcarnitine levels have been reported to be associated with metabolic disorders such as obesity and diabetes. However, the understanding of human genetic variation affecting plasma acylcarnitine levels remains limited. We performed a genome-wide association study of acylcarnitine levels and their ratios in 2865 unrelated Chinese Hans from the Study on Nutrition and Health of Aging Population in China, a population-based cohort study. We identified three genome-wide significant novel loci at *FADS1* (β : 0.252-0.481, $P \leq 6.16 \times 10^{-17}$, for associations with plasma levels of C20:4n-6, and acylcarnitine ratios including C14OH/C20:4n-6, C14:1OH/C20:4n-6, C16:1/C20:4n-6, C16:2/C20:4n-6, C18/C20:4n-6, C18:2/C20:4n-6, and C20/C20:4n-6, respectively), *ACADVL* (β : -0.272 - 0.356, $P \leq 9.32 \times 10^{-11}$, for association with acylcarnitine ratios including C3DC/C12, C6/C12, C6/C14, C6OH/C12, C6OH/C14, C8/C12, C8/C14, C10/C12, C12/C12OH, C12OH/C14, C14OH/C16:2, and C14:1OH/C16:2, respectively) and *ACSM5* (β : -0.263-0.257, $P \leq 9.95 \times 10^{-10}$ for associating with C2/C7DC and C7DC/C12OH ratio, respectively). We also confirmed six loci ($P \leq 9.89 \times 10^{-10}$) (*CPT2*, *ACADM*, *CPS1*, *MCCC1*, *ETFDH*, and *ACADS*) previously reported for associations with plasma levels of various components of acylcarnitines and their ratios in European populations. Notably, two novel independent association signals were observed at loci *CPS1* (rs715) and *CPT2* (rs12082305) for associations with plasma C6DC(rs715) and C16(rs12082305) levels ($P \leq 9.89 \times 10^{-10}$) and acylcarnitine ratios including C6/C16, C6OH/C16, C6DC/C7DC, C6DC/C16, C6DC/C10DC, C14OH/C18 and C14OH/C20 ($P \leq 7.99 \times 10^{-10}$). Our findings provide new insights into the genetic basis underlying carnitine metabolism.

3426W

Genome-wide association study in 400k individuals identifies 139 novel signals of association with lung function highlighting pathways and pleiotropy. N. Shrine¹, A.L. Guyatt¹, V. Jackson¹, A.M. Erzurumluoglu¹, C. Batini¹, N. Reeve¹, K. Song², P. Sakornsakolpat^{3,4}, B.D. Hobbs^{3,5}, M.H. Cho^{3,5}, D.P. Strachan⁶, A.P. Morris⁷, I.P. Hall⁸, M.D. Tobin^{1,9}, L.V. Wain^{1,9}, SpiroMeta consortium, Lung eQTL study. 1) Department of Health Sciences, University of Leicester, Leicester, UK; 2) Target Sciences - R&D, GSK Medicines Research Centre, Gunnels Wood Road, Stevenage, Hertfordshire, SG1 2NY, UK; 3) Chan-ning Division of Network Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA; 4) Department of Medicine, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand; 5) Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA; 6) Population Health Research Institute, St George's, University of London, London, UK; 7) Department of Biostatistics, University of Liverpool, Liverpool, UK; 8) Division of Respiratory Medicine, University of Nottingham, Nottingham, UK; 9) National Institute for Health Research Biomedical Research Centre – Respiratory theme, Glenfield Hospital, Leicester, UK.

Chronic obstructive pulmonary disease (COPD) accounted for 2.9 million deaths globally in 2016. Spirometrically-determined lung function is used in the diagnosis and grading of COPD. Genome-wide association studies (GWAS) of lung function quantitative traits are a more powerful alternative to COPD case-control studies for increasing the yield of associated genetic variants. The variants thus identified highlight the genes and pathways underpinning lung function and development of COPD potentially leading to novel preventive and therapeutic strategies for COPD. Through enhanced quality control and characterisation of spirometric measures of lung function in UK Biobank, completion of UK Biobank genotyping and expansion of the SpiroMeta Consortium, we undertook the largest genome-wide association study of quantitative lung function traits performed to date, comprising 400k individuals of European ancestry. We report 139 new independent signals of association with forced expiratory volume in 1s (FEV₁), forced vital capacity (FVC), FEV₁/FVC and peak expiratory flow (PEF), bringing the number of new and previously reported signals of association with lung function to 279. Fine-mapping of signals, together with gene and protein expression analyses with improved tissue specificity and stringency, has implicated new genes and pathways highlighting the importance of cilia development, TGFB-signalling, and elastic fibres in the aetiology of airflow obstruction. Many of the genes and pathways contain drug targets including those in development for indications other than COPD. Phenome-wide association testing of 279 lung function signals with 2,411 phenotypes in UK Biobank revealed genetic variants with apparent specificity of lung function effects and others with pleiotropic effects that might indicate shared biological pathways between different diseases. Investigating the nature of the pleiotropic effects of some of these variants associated with multiple traits may indicate the relevance of drug repurposing.

3427T

Impact of ancestral variation in the levels and genetics of n-3 and n-6 long chain polyunsaturated fatty acids in distinct Hispanic subgroups. C. Yang¹, L.M. Reynolds², T.D. O'Connor³, A.C. Wood⁴, R. Dutta⁵, M.C. Seeds⁶, I. Rucczinski⁷, R.A. Mathias⁸, R.N. Lemaitre⁹, S.S. Rich¹, F.H. Chilton¹⁰, A. Manichai-kul¹. 1) Center for Public Health Genomics, University of Virginia, Charlottesville, VA; 2) Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC; 3) Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD; 4) ARS/USDA Children's Nutrition Research Center, Department of Pediatrics, Baylor College of Medicine, Houston, TX; 5) Department of Urology, Wake Forest School of Medicine, Winston-Salem, NC; 6) Department of Internal Medicine/Section on Molecular Medicine, Wake Forest School of Medicine, Winston-Salem, NC; 7) Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD; 8) Department of Medicine, Johns Hopkins University, Baltimore, MD; 9) Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, Washington; 10) Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC.

Background. Genetic variants in the fatty acid desaturase (FADS) cluster are associated with LC-PUFAs levels and numerous molecular and clinical phenotypes. PUFA metabolism has had limited attention in non-Caucasian populations. In this study we examined the hypotheses that (a) Hispanic American ancestry is predictive of LC-PUFA levels, and (b) distinct patterns of genetic association are observed in different Hispanic subgroups. **Methods.** 1,225 Hispanic American participants from the Multi-Ethnic Study of Atherosclerosis (MESA) were assayed by gas chromatography for plasma phospholipid n-3 eicosapentaenoic acid (EPA), docosapentaenoic acid (DPA), docosahexaenoic acid (DHA) and n-6 (arachidonic acid, AA) PUFAs. Hispanic country of origin was inferred using participant-reported birthplaces for their parents and grandparents. Regression analysis was used to assess associations of LC-PUFA levels with percent Native American ancestry, computed using ADMIXTURE, and FADS cluster SNPs, available via genome-wide genotyping with imputation to the 1000 Genomes Phase 3. Regression models included covariate adjustment for age, sex, study site, four PCs of ancestry and batch. **Results.** The MESA Hispanic cohort represented the following subgroups: Central America (n=86), Cuba (n=47), Dominican (n=152), Mexico (n=644), Puerto Rico (n=169) and South America (n=100). Higher proportions of Native American ancestry associated with lower levels of all LC-PUFAs in participants of Central American, South American or Mexican origin compared to those of Dominican, Cuban or Puerto Rican origin. For example, each 10% increase in Native American ancestry associated with decreases of 0.40 and 0.14 percent of total fatty acids for AA ($p=2.0 \times 10^{-16}$) and DHA ($p=2.0 \times 10^{-16}$), respectively. In genetic analysis, there was a strong association between AA and rs97384 (FADS2, intronic, $p=1.2 \times 10^{-10}$) in the Central America subgroup, and between AA and rs174533 (MYRF, intronic, $p=1.5 \times 10^{-12}$) in the South America subgroup, with 78kb separating the two SNPs ($R^2=0.40$ for these two SNPs in 1000G AMR). EPA was associated with rs174536 (MYRF, intronic, $p=3.09 \times 10^{-24}$) in the Mexico subgroup, but did not demonstrate statistically significant associations in other subgroups. **Summary.** Statistically significant differences in LC-PUFA levels and genetic associations were observed across subgroups in the MESA Hispanic cohort, suggesting that certain Hispanic subgroups may have deficiencies in LC-PUFAs.

3428F

Genetic correlation between polypoidal choroidal vasculopathy and age-related macular degeneration. Q. Fan¹, C.Y. Cheng², C.P. Pang³, K.H. Park⁴, N. Yoshimura⁴. 1) CQM Duke-NUS University, Singapore, Singapore; 2) Singapore Eye Research Institute, Singapore Singapore; 3) Department of Ophthalmology and Visual Sciences, the Chinese University of Hong Kong, Hong Kong, China; 4) Department of Ophthalmology and Visual Sciences, Kyoto University Graduate School of Medicine, Kyoto, Japan; 5) Department of Ophthalmology, Seoul National University Bundang Hospital, Gyeonggi, Korea.

Polypoidal choroidal vasculopathy (PCV) is a subtype of age-related macular degeneration (AMD) and more frequently seen in East Asians. PCV has both common and distinct clinical manifestations with typical neovascular AMD (tAMD). We aim to examine the extent to the shared genetic variants between these two subtypes. We performed the meta-analysis of association in a total of 2,219 PCV and AMD patients and 5,275 controls of East Asian descent from the Genetics of AMD in Asians (GAMA) Consortium. PCV and tAMD are genetically highly correlated ($r_s = 0.69$). A total of eight known loci were significantly associated with PCV, from the SNP-based test and the gene-based tests. Weaker association for PCV was observed at *ARMS2-HTRA1*, compared with tAMD. Variants at *CFH*, *CETP*, and *VEGFA* exhibited different association signals in East Asians, in contrast to those in Europeans. Our data suggest a substantially shared genetic susceptibility for PCV and tAMD.

3429W

Genome-wide haplotype association studies: Comparison of novel methods. B. Laabs, I.R. König. Institut für Medizinische Biometrie und Statistik, Universität zu Lübeck, Universitätsklinikum Schleswig-Holstein, Campus Lübeck, Lübeck, Germany.

Standard genome-wide association studies (GWAS) based on single SNPs are known not to exploit the entire genetic information, thus possibly missing relevant genetic associations. Important gaps might be filled analyzing multiple SNPs simultaneously by the analysis of genetic interactions or haplotypes. The latter are particularly promising in that a haplotype not only contains the information of the SNPs themselves, but also of their interactions and unmeasured loci between them. One of the disadvantages of haplotype-based GWAS over SNP-based GWAS is that they need preprocessing to be applicable. For example, it is necessary for the most haplotype-based approaches to determine haploblocks which build different haplotypes. Another challenge for haplotype-based GWAS in the past was the handling of rare haplotypes. Most of the older methods avoided the resulting statistical problems by summing up all rare haplotypes in one category. However, especially these rare haplotypes could carry information of rare functional variants. In the last few years a number of new methods for haplotype-based GWAS have been proposed which are able to estimate the association with rare haplotypes. In this contribution, we compare some of the methods which are able to deal with rare haplotypes with regard to the power to detect functional haplotypes. Additionally we use one of the standard approaches (haplo.glm) that pools rare haplotypes using a fixed threshold as reference. We apply these methods to a GWAS data set with given pre-specified haploblocks and compare the findings with the results of earlier studies using haplotype-based GWAS.

3430T

Leveraging biological pathways and gene networks to understand the genetic architecture of diseases and complex traits. S.S. Kim^{1,2}, C. Dai³, F. Hormozdizadeh⁴, B. Geijn⁵, S. Gazal², Y. Park¹, L. O'Connor², P. Loh⁴, H. Finucane⁵, A.L. Price^{2,3,5}. 1) Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA; 2) Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA; 3) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA; 4) Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 5) Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA.

Recent studies have implicated many biological pathways that are enriched for disease loci and heritability, and have suggested that gene networks can help refine these signals. Here, we constructed a broad set of pathway, network, and pathway+network annotations and applied stratified LD score regression to 42 independent diseases and complex traits (average N=323K) to identify enriched annotations conditioned on all genes and on the baseline-LD model (Gazal et al. 2017 Nat Genet). We constructed annotations from 18,119 known biological pathways, including 100kb windows around each gene. We identified 156 pathway-trait pairs whose enrichment was statistically significant (FDR<5%) after conditioning on all other annotations, a stringent step that greatly reduced the number of pathways detected. These annotations had an average enrichment of 4.13 (s.e. 0.12). Most pathway-trait pairs were previously unreported, such as GABA-A receptor activity for autism, which is consistent with known downregulation of GABA receptors in autism cases. For each of 4 published gene networks, we constructed probabilistic annotations based on closeness centrality, a gene-specific measure of how close a gene is to other genes in the network. Each network annotation was significantly enriched: up to 1.37x (s.e. 0.05), meta-analyzed across traits. The enrichments were fully explained by correlations between network annotations and regulatory annotations from the baseline-LD model, validating the baseline-LD model and demonstrating the importance of accounting for regulatory annotations in gene network analyses. For each of the 156 enriched pathway-trait pairs, we constructed pathway+network annotations reflecting all-new sets of candidate genes linked to the input pathway via the input gene network, by applying a machine learning method that leverages topological patterns and network centrality metrics (Li et al. 2018 Nat Methods). For each of 4 gene networks, these pathway+network annotations were significantly enriched for the corresponding traits: up to 1.36 (s.e. 0.02), averaged across 156 pathway-trait pairs. The pathway+network enrichments were significant conditional on network annotations, but were largely explained by regulatory annotations from the baseline-LD model. In conclusion, both biological pathways and gene networks can be highly informative for disease and trait architectures, but the information in gene networks may be subsumed by regulatory annotations.

3431F

Previously ignored high-frequency common variants greatly contribute to heritability of a common disease. Y. Nagao^{1,2}. 1) Department of Clinical Genetics, Tokai University, Isehara, Kanagawa, Japan; 2) Department of Pediatrics, Takashimadaira Chuo General Hospital, Itabashi, Tokyo, Japan.

Genome-wide association studies are also called common variant association studies because almost all the identified single nucleotide polymorphisms belong to common variants. In the conventional studies the estimation of heritability for a disease was performed for the sets of variants with an assumption that each allele of a variant shows a co-dominant effect, and variants had been classified by minor allele frequency (MAF). In the published report I showed a method to calculate heritability of an individual polymorphism (h_p^2). Here I estimate h_p^2 of a variant that acts in three different manners; a co-dominant, a dominant and a recessive manner, utilizing the method. The results suggest that h_p^2 of a high-frequency common variant ($p > 0.5$) is too high to be ignored. Especially high-frequency variants which act in a recessive manner have the highest h_p^2 among all the variants with same odds ratio (OR). In the conventional studies a high-frequency variant had been treated as a variant that has a MAF of another allele and a reciprocal of OR, and it is thought that such variants account for about half of the total. However, contribution of variants to heritability had been evaluated only for those with OR larger than 1.0. A trial calculation of heritability for schizophrenia when considering high-frequency common variants suggest that it is estimated 1.7 times the conventional heritability explained by common variants. Missing heritability might be solved by re-examining previously ignored variants, i.e., those with OR less than 1.0 when classified by MAF.

3432W

Reconciling S-LDSC and LDAK models and functional enrichment estimates. A.L. Price^{1,2,3}, S. Gazal^{1,2}, H.K. Finucane². 1) Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA; 2) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; 3) Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA.

Recent work has highlighted the importance of accounting for LD-dependent genetic architectures in analyses of functional components of heritability, motivating the development of the LDAK model and the baseline-LD model used by S-LDSC. Although both models include LD-dependent effects, they produce different estimates of functional enrichment, e.g. 9.35x (± 0.80) enrichment for conserved regions using baseline-LD (Gazal et al. 2017 Nat Genet; average $N=96K$) vs. 1.34x (± 0.26) using LDAK (Speed et al. 2017 Nat Genet; average $N=7K$). Here, we perform formal model comparisons and empirical analyses using a combined model to reconcile these findings. We performed model comparisons using a likelihood approach. We compared an infinitesimal model with no LD-dependent effects ("GCTA model"), the LDAK model, and the baseline-LD model. For 16 highly heritable UK Biobank traits, we computed likelihoods using $N=20K$ samples and either $M=2.8M$ imputed 1000G SNPs or $M=4.7M$ imputed HRC SNPs, using enrichment parameters that we fitted (where applicable) using a disjoint set of samples. In the $M=2.8M$ analysis, LDAK outperformed GCTA (change in log likelihood (ΔLL)=334), consistent with Speed et al. 2017; however, in the $M=4.7M$ analysis, GCTA slightly outperformed LDAK ($\Delta LL=15$). Baseline-LD outperformed LDAK in both analyses ($\Delta LL \geq 644$), with $\Delta LL > 0$ for at least 15/16 traits. We obtained similar results for a model including only the LD and MAF annotations from baseline-LD ($\Delta LL \geq 439$ vs. LDAK) and for a model combining baseline-LD annotations with annotations constructed using LDAK weights (baseline-LD+LDAK; $\Delta LL \geq 761$). Model comparisons based on accuracy of out-of-sample polygenic predictions computed using each model as a prior (Marquez-Luna et al. ASHG abstract) produced similar findings. We determined that S-LDSC with the baseline-LD+LDAK model (unlike baseline-LD or LDAK) produces unbiased estimates in simulations both under the LDAK model and under the baseline-LD model. In analyses of 16 UK Biobank traits and 28 main functional annotations, functional enrichment estimates obtained using baseline-LD+LDAK (e.g. 7.51x (± 0.49) for conserved regions) were not significantly different from those obtained using baseline-LD (e.g. 8.11x (± 0.54)), but were significantly larger than those obtained using LDAK (e.g. 2.62x (± 0.17)). Our results support the use of the baseline-LD model and confirm the existence of functional annotations that are highly enriched for complex trait heritability.

3433T

Distinguishing maternal and fetal contribution to the heritability – a novel approach for complex pregnancy and early fetal phenotypes. A.K. Srivastava^{1,2}, J. Juodakis³, J. Bacelis⁴, J. Chen⁵, K. Teramo⁶, M. Hallman⁷, B. Jacobsson^{8,9}, L.J. Muglia^{1,2}, G. Zhang^{1,2}. 1) Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA; 2) Center for Prevention of Preterm Birth, Perinatal Institute, Cincinnati Children's Hospital Medical Center and March of Dimes Prematurity Research Center Ohio Collaborative, Cincinnati, OH, USA; 3) Department of Obstetrics and Gynecology, Institute of Clinical Sciences, Gothenburg, Sweden; 4) Department of Obstetrics and Gynecology, Sahlgrenska University Hospital Östra, Gothenburg, Sweden; 5) Division of Bioinformatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA; 6) Obstetrics and Gynecology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland; 7) PEDEGO Research Unit and Medical Research Center Oulu, University of Oulu and Department of Children and Adolescents, Oulu University Hospital, Oulu, Finland; 8) Department of Obstetrics and Gynecology, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden; 9) Department of Genetics and Bioinformatics, Area of Health Data and Digitalization, Norwegian Institute of Public Health, Oslo, Norway.

The last decade has witnessed several technical and theoretical advances towards solving the mystery of missing heritability for various complex traits/diseases. However, pregnancy outcomes such as gestational age and fetal growth measures at birth, conjointly defined by maternal and fetal genomes attained restricted expedition – possibly due to lack of appropriate approaches to avoid confounding by shared alleles between mother and fetus. Here, we propose a novel approach of treating mother-fetus pairs as a single analytical unit composed of three alleles - maternal transmitted, maternal non-transmitted and paternal transmitted. We generate three separate relatedness matrices using only maternal transmitted alleles, only maternal non-transmitted alleles and only paternal transmitted alleles. We fit the relatedness matrices concurrently in linear mixed model to identify the additive genetic variance attributed to each of three components. We test the validity and robustness of the approach using simulated genotype and phenotype data. Further, we apply the approach on empirical dataset (~6000 mother-fetus pairs) to estimate the heritability of gestational age and fetal anthropometric traits at birth. The study suggests that narrow-sense heritability of gestational age is mainly attributed to maternal transmitted (12.9%) and non-transmitted (10%) alleles whereas paternal transmitted alleles have negligible contribution to its heritability. On the contrary, narrow-sense heritability of fetal birth weight and birth length are mainly owed to maternal transmitted alleles (10-15%) whereas maternal non-transmitted and paternal transmitted alleles have negligible contribution. Further, this method can be extended to trios by fitting an additional relatedness matrix into the model, generated from paternal non-transmitted alleles. Our method provides an advancement to existing methods for the heritability estimation and detect explicit contributions of maternal and fetal alleles for duos and also, paternal alleles for trios.

3434F

Assessment of genome-wide significance of conditionally independent signals. S. Ghasemi^{1,2}, A. Teumer^{1,2}, T. Becker¹. 1) Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany; 2) DZHK (German Center for Cardiovascular Research), partner site Greifswald, Greifswald, Germany.

The presence of multiple independently associated SNPs within the same linkage disequilibrium (LD) region is a common phenomenon. In contrast to tests for primary association, conditional tests should not be applied on a genome-wide scale but are limited to the genomic regions surrounding primary GWAS index signals. Therefore, the established genome-wide significance level $\alpha = 5 \times 10^{-8}$ in conditional analyses is too strict and implies an unnecessary loss of power to detect secondary signals. We developed a "quasi-adaptive" method not only to take into account the reduced number of tests but to use R^2 and physical distance LD (R^2 , D) to the index SNP to prioritize surrounding SNPs for analysis. In this regard first, we identify surrounding LD region for each index SNP, then compute R^2 and physical distance D with the index SNP for each SNP, and construct a general priority function $f = f(R^2, D)$. Next, for each candidate SNP, we compute conditional p-values with GCTA. Our conditional analysis controls family-wise error rate at $\alpha = 0.05$. We spend this amount of α level over the candidate SNPs from the LD regions. The distribution, however, is not even. Instead, SNPs with higher priority will get a higher α value. To this purpose, we use a distribution function $G = G(\alpha, f)$ which assigns to each SNP its SNP-specific α -threshold. Finally, for each SNP, the GCTA p-values are compared to the SNP-specific α -threshold, in order to assess significance. In essence, SNPs with higher priority have to meet less stringent alpha-levels, or, in other words, more alpha is spent on SNPs with higher priority. As a proof-of-concept, we applied our method on a GWAS on thyroid-stimulating hormone (TSH) using the SHIP cohort ($n = 3603$) focusing on four published associations (Rawal *et al.* 2012): (PDE8B, CAPZB, NR3C2, LOC440389). We could find the signal rs16956373 genome-wide conditionally independent significant for the index SNP rs3813582 (LOC440389) that would not have been found with the standard approach (i.e. requiring $\alpha = 5 \times 10^{-8}$). We developed an algorithm with increased power for detecting secondary SNP signals and applied it successfully to a GWAS. Ongoing steps include replication of the findings in a larger GWAS and a simulation study to assess the overall experiment-wise type I error of our method.

3435W

A GWAS and candidate gene study for maternal nondisjunction of chromosome 21. J.M. Chernus¹, E.G. Allen², Z. Zeng³, E.R. Hoffmann⁴, T.J. Hassold⁵, E. Feingold^{1,3}, S.L. Sherman². 1) Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA; 2) Department of Human Genetics, Emory University School of Medicine, Atlanta, GA, USA; 3) Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA; 4) Center for Chromosome Stability, University of Copenhagen, Copenhagen, Denmark; 5) School of Molecular Biosciences and Center for Reproductive Biology, Washington State University, Pullman, WA, USA.

Maternal meiotic nondisjunction is the leading cause of human pregnancy loss, birth defects, and intellectual disabilities. Known risk factors include increasing maternal age and altered patterns of meiotic recombination. To better understand the genetic components of risk and the processes underlying errors in meiosis, we conducted the first genome-wide association study of maternal nondisjunction of chromosome 21. Our study group consisted of 2,186 participants genotyped on an Illumina SNP array, comprising 749 live-born children with standard, maternally-derived trisomy 21 and 1,437 of their parents. First, within each family, genotypes of the parents and child were used to determine the maternal origin of the trisomy and the stage of meiosis in which nondisjunction occurred (meiosis I (MI) or meiosis II (MII)). As our previous work has suggested that MI and MII nondisjunction may differ in etiology, we performed a novel set of four genome-wide scans designed both to identify variants associated with nondisjunction in either stage of meiosis and to distinguish variants with an effect specific to just one stage. In addition to annotating GWAS signals, we also investigated candidate genes and loci linked to chromosome dynamics during meiosis and loci associated with genome-wide recombination counts in humans. The candidate gene study showed statistically significant associations, including *RAD21L1*, *SMEK1*, *SYCP1*, *SYCP2*, *SYCE2*, and *MND1*. In the genome-wide studies, several suggestive ($p < 10^{-5}$), but not genome-wide significant, associations were observed in or near genes with plausible roles in relevant meiotic processes, such as *SLC39A10* (involved in zinc homeostasis, crucial to mouse oocyte maturation) and *MYO10* (thought to be required for meiotic cytokinesis). These results suggest for the first time that common variation in genes responsible for normal meiotic processes confers risk for nondisjunction.

3436T

Dissecting the causal relationship between maternal phenotypes and birth outcomes: A genetic score analysis in mother/infant pairs. J. Chen¹, A. Srivastava^{2,3}, J. Bacelis⁴, J. Juodakis⁵, K. Teramo⁶, M. Hallman⁷, B. Jacobsson^{8,9}, L. Muglia^{2,3}, G. Zhang^{2,3}. 1) Division of Biomedical Informatics, Cincinnati Children's Hospital, Cincinnati, OH; 2) Division of Human Genetics, Cincinnati Children's Hospital, Cincinnati, OH; 3) Center for Prevention of Preterm Birth, Perinatal Institute, Cincinnati Children's Hospital Medical Center and March of Dimes Prematurity Research Center Ohio Collaborative; 4) Department of Obstetrics and Gynecology, Sahlgrenska University Hospital Östra, University of Gothenburg, Gothenburg, Sweden; 5) Department of Obstetrics and Gynecology, Institute of Clinical Sciences, University of Gothenburg, Gothenburg, Sweden; 6) Obstetrics and Gynecology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland; 7) PEDEGO Research Unit and Medical Research Center Oulu, University of Oulu, and the Department of Children and Adolescents, Oulu University Hospital, Oulu, Finland; 8) Department of Obstetrics and Gynecology, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden; 9) Department of Genetics and Bioinformatics, Area of Health Data and Digitalization, Norwegian Institute of Public Health, Oslo, Norway.

Background: This study aimed to investigate the causal mechanisms underlying the observed associations between maternal phenotypes and birth outcomes using a parental-specific genetic score approach. **Methods and Findings:** We conducted a genetic score analysis using phenotype and genome-wide single nucleotide polymorphism (SNP) data of 8,067 mother/infant pairs from multiple birth cohorts collected from European countries and the United States. We built genetic scores using SNPs known to be associated with adult height, BMI, blood pressure and blood glucose level to index these maternal phenotypes respectively. To avoid the interference in causal inference due to the transmission of maternal alleles, we inferred parental transmission of the SNPs and constructed parental-specific haplotype genetic scores. Our haplotype-based genetic score association analysis showed that maternal non-transmitted haplotype score of height was significantly associated with longer gestational duration ($P=9.5E-5$), larger birth weight ($P=1.3E-6$) and birth length ($P=7.4E-3$); but the latter two birth size measures were more significantly associated with maternal and paternal transmitted height scores (all $P<1E-12$). Our results also showed that maternal transmitted BMI score was associated with birth weight ($P=6.4E-4$). In addition, both the maternal transmitted and the paternal transmitted haplotype scores of blood pressure were significantly associated with reduced birth weight ($P = 2.3E-3$ and $1.4E-3$, respectively), suggesting genetic variants associated with blood pressure have direct genetic effects in the fetuses. In contrast, the maternal non-transmitted haplotype score of blood glucose was significantly associated with increased birth weight ($P = 3.2E-5$) while the paternal transmitted haplotype score of blood glucose was significantly associated with decreased birth weight ($P = 2.5E-2$) indicating that the birth weight is positively causal influenced by the maternal blood glucose level but negatively influenced by blood glucose increasing alleles carried by the fetuses. **Conclusions:** Our results confirmed the causal impact of maternal height on gestational duration and also demonstrated the strong causal effect of maternal blood glucose level on fetal growth. In addition, our results also showed that genetic variants associated with adult height, BMI, glucose, and blood pressure have various direct genetic effects on fetal growth.

3437F

Exploring the effect of parental height on a newborn's birth weight. Y. Lee¹, H. Gjessing^{1,2}, A. Jugessur^{1,2}, P. Magnus¹. 1) Norwegian Institute of Public Health, Oslo, Norway; 2) University of Bergen, Bergen, Norway.

Parental height and newborn's birth weight are associated. To understand the underlying etiology, we used parental polygenic risk scores for height as instrumental variables. Genotype data from 10,496 complete newborn parent trios were available from the 'HARVEST' study, which is part of the Norwegian Mother and Child Cohort Study. We identified 697 SNPs in HARVEST that were genome wide significant ($p<5\times 10^{-8}$) for height, and these were used to generate a polygenic risk score (PRS). The birth weight (gram) of newborns was extracted from the Medical Birth Registry of Norway. Two stage least square regression was used, with and without adjusting for the newborn's own PRS. The effect of mother's height (cm) on newborn's birth weight (gram) with and without adjustment for the newborn's own PRS for height were 12.1 (95% CI: 7.7, 16.5) and 20.5 (95% CI: 16.9, 24.1) respectively. The father's height also increased the newborn's birth weight (12.8; 95% CI: 9.3, 16.2), but this effect substantially decreased after adjusting for newborn's PRS (3.3; 95% CI: -0.8, 7.3). Maternal height increases newborn's birth weight independently of the newborn child's PRS. By contrast, the effect of paternal height was attenuated after adjusting for the newborn child's PRS. These results, with some assumptions, suggest that maternal height is causally related to fetal growth.

3438W

Examining the role of reverse causality in Mendelian randomization. S.M. Lutz¹, J.E. Hokanson¹, C. Lange². 1) University of Colorado, Anschutz, Aurora, CO; 2) Harvard T.H. Chan School of Public Health, Boston, MA.

In the past few years, Mendelian Randomization (MR) has gained in popularity as a tool to examine if a phenotype of interest causes the outcome by using Single Nucleotide Polymorphisms (SNPs) as instrument variables for the phenotype of interest. MR assumes that the only path from the SNPs to the outcome is through the phenotype of interest (i.e. no pleiotropic relationships exist). Although, it is not always known which variable is the phenotype of interest and which variable is the outcome. For instance, if an investigator wants to use MR to examine the role of depression on BMI, it is not clear if BMI causes depression or if depression causes a change in BMI. In a scenario like this, investigators have run the MR analysis twice: 1) with BMI as the phenotype of interest and depression as the outcome and 2) with depression as the phenotype of interest and BMI as the outcome. It is not clear how reverse causality influences the type 1 error rate for MR, especially in the presence of pleiotropy. Through simulation studies, we examined the type 1 error rate for 4 popular MR methods when the role of the intermediate phenotype and outcome were reversed (i.e. reverse causality) with and without a pleiotropic relationship of at least one SNP on the outcome. When there is a pleiotropic relationship with at least one SNP and the outcome, the type 1 error rate is inflated for all MR approaches in the scenario of reverse causation. We apply this approach to the COPDGene study, a case-control study of Chronic Obstructive Pulmonary Disease (COPD) in current and former smokers to examine the role of methylation and cigarette smoking.

3439T

A Mendelian randomization study detects the distortive effects of health-related traits on food frequency questionnaire responses. C. McDonnell¹, E.J. Grzeszkowiak¹, J.F. Wilson^{1,2}, N. Pirastu¹. 1) Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Teviot Place, Edinburgh, Scotland, EH8 9AG; 2) MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, Scotland, EH4 2XU.

Background Nutrition is an important contributor to the development of obesity and chronic disease. Carrying out nutritional research is, however, difficult due to known bias in dietary questionnaires caused by misreporting of food intake by people with particular health conditions. Measuring this bias is challenging because it requires knowledge of true food consumption that is difficult to obtain in a large study. Our aim was to verify if measuring this bias is possible using Mendelian randomization to test the effect of BMI on reported food consumption. **Methods** We created genetic instruments for several exposures for which dietary advice is usually given (body mass index, LDL, HDL, & total cholesterol, diastolic and systolic blood pressure and triglycerides), by LD pruning of publicly available summary statistics. We ran association analysis of the selected traits on the FFQ traits available in UK Biobank (n~500,000) to use as outcomes. After removing SNPs with heterogeneous effects, MR was conducted with an inverse variance-weighted method or with MR Egger regression if the intercept of the regression was significantly different from zero. **Results** We identified significant results suggesting a causal effect of all health-related trait exposures on food intake. We report Benjamini & Hochberg-corrected p-values. Individuals genetically predisposed to a higher BMI reported lower: beef intake ($p=1.4 \times 10^{-3}$), fat content of milk ($p=1.2 \times 10^{-36}$) and addition of spread to bread ($p=8.4 \times 10^{-16}$), whilst reported increased intake of fresh fruit ($p=8.5 \times 10^{-4}$) and salad ($p=5.3 \times 10^{-4}$). We detected a similar pattern for total and LDL cholesterol, with individuals with genetically raised lipids reporting lower beef ($p=5.0 \times 10^{-2}$, $p=5.7 \times 10^{-3}$), fat content of milk ($p=5.7 \times 10^{-28}$, $p=5.7 \times 10^{-28}$), addition of spread to bread ($p=7.9 \times 10^{-4}$, $p=1.9 \times 10^{-3}$), along with cheese ($p=3.8 \times 10^{-10}$, $p=2.4 \times 10^{-9}$) and salt intake ($p=7.3 \times 10^{-5}$, $p=6.2 \times 10^{-4}$). There was an increased reporting of fresh fruit ($p=6.9 \times 10^{-4}$, $p=7.6 \times 10^{-4}$) and salad ($p=9.1 \times 10^{-3}$, $p=4.8 \times 10^{-3}$). Additionally, those with higher diastolic blood pressure reported a lower salt intake ($p=3.0 \times 10^{-2}$). **Conclusion** Our results suggest that the studied exposures are indeed influencing the reported food consumption. This corresponds with previous epidemiological studies on FFQ reporting but for the first time, we can give an estimate of this bias using a large sample size. This will be important for correctly estimating food-health relationships.

3440F

Phenome-wide Mendelian randomization study mapping the influence of the plasma proteome on complex diseases. J. Zheng¹, V. Haberland¹, D. Baird¹, A. Gutteridge², J. Dopierala², J. Staley¹, B. Sun⁴, J. Danesh⁴, H. Runz³, J. Maranville³, A. Butterworth⁴, J. Yarmolinsky¹, B. Elsworth¹, C. Laurin¹, V. Walker¹, J. Liu³, K. Estrada³, M. Nelson³, P. Haycock¹, G. Davey Smith¹, G. Hemani¹, R. Scott⁴, T. Gaunt¹. 1) MRC Integrative Epidemiology Unit, Bristol Medical School, University of Bristol; 2) Glaxosmithkline, Gunnels Wood Road, Stevenage, Hertfordshire, SG1 2NY, UK; 3) Biogen, 225 Binney Street, Cambridge, MA 02142; 4) MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK; 5) Celgene, 1 Longwalk Rd, Hayes, Uxbridge UB11 1DB.

Drug development is a lengthy, expensive process beset by high failure rates due to lack of efficacy or adverse safety signals. As such, approaches to predict safe and effective targets as early as possible in the drug development process are much needed. Genetic analysis of the plasma proteome provides a promising opportunity to validate novel or existing targets through prediction of efficacy or potential adverse effects and by guiding target-indication pairing. Here, we performed two sample Mendelian randomization (2SMR) to map the putative causal effects of thousands of plasma proteins against hundreds of human diseases. Combining GWAS summary data from four protein quantitative trait locus (pQTL) studies, we identified 1,449 plasma proteins with genetic instruments. To increase reliability of the instruments, we focused on those with no strong evidence of heterogeneity across studies and performed 2SMR of each protein on 130 diseases in MR-Base (163,604 tests, Bonferroni-corrected P threshold of 3.05×10^{-7}). For MR results that reached statistical significance, we further evaluated the findings by performing colocalization analysis to investigate whether the pQTL was also likely to be driving or in linkage disequilibrium with a distinct disease association. All analyses were performed in the TwoSampleMR and coloc R packages. Using 2SMR, we observed 866 associations across 105 diseases, 395 of which remained promising after heterogeneity test of instruments across pQTL studies and colocalization analysis. For example, MR using two independent cis-acting SNPs (rs11581607 and rs3762318) showed that reduced *interleukin 23 receptor* (IL23R) level was associated with decreased risk of inflammatory bowel disease (IBD) (OR=0.37, 95%CI=0.34 to 0.40, $P=1.92 \times 10^{-143}$). Colocalization analysis suggested both traits shared the same causal variants. An existing IL23R antagonist, Ustekinumab, demonstrated efficacy in reducing Crohn's disease activity index in a phase 2 trial, and is currently in phase 3 trials for IBD (www.clinicaltrials.gov). Our study identified thousands of putatively causal associations between the plasma proteome and the human phenotype. These associations are necessary but not sufficient to prove causality, as potential horizontal pleiotropy remains an alternative explanation. We will release the MR and colocalization analysis results to the public. We are further exploring the translational value of our findings for drug discovery and development.

3441W

One stop shop for X-inclusive whole-genome association scans. L. Sun^{1,2}, B. Chen¹, R.V. Craiu¹, B. Xiao³, J. Gong³, G. He³, N. Panjwani³, S. Blackman⁴, H. Corvol^{5,6}, G. Cutting^{7,8}, M. Drumn⁹, M. Knowles¹⁰, J. Rommens³, L. Strug^{3,2}.

1) Department of Statistical Sciences, Faculty of Arts and Science, University of Toronto, Toronto, ON, M5G 3G3 Canada; 2) Biostatistics Division, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, M5T 3M7, Canada; 3) Program in Genetics and Genome Biology, Research Institute, The Hospital for Sick Children, Toronto, ON, M5G 0A4, Canada; 4) Division of Pediatric Endocrinology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21287, USA; 5) Assistance Publique-Hopitaux de Paris (AP-HP), Hopital Trousseau, Pediatric Pulmonary Department; Institut National de la Sante et la Recherche Medicale (INSERM) U938, Paris 75012, France; 6) Sorbonne Universites, Universite Pierre et Marie Curie (UPMC) Paris 06, Paris 75005, France; 7) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21287, USA; 8) Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, Maryland 21287, USA; 9) Department of Pediatrics, School of Medicine, Case Western Reserve University, Cleveland, Ohio 44106, USA; 10) Marsico Lung Institute/UNC CF Research Center, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA.

Despite the call for X-inclusive whole-genome association scans in 2013 (Wise et al.), to date it is yet to be a common practice presumably due to the additional analytic complexities. Several statistical methods have been developed, each of which only partially address the X-chromosome considerations. For example, some allow for sex-specific allele frequencies, while others are designed to handle the possibility of random, skewed or no X-inactivation. Assumption of additivity, choice of reference allele and sex-by-SNP interaction analysis have received little attention to date. Through theoretical justifications, simulation and application results, we provide a unified regression-based framework that simultaneously resolves these X-chromosome issues. Unlike the autosomes, we show that the choice of reference allele can impact X-chromosome inference if sex-specific effects are not accounted for in an analysis. Further, the current standard practice that models the main effect of a SNP assuming additivity can be called into question. Adding a dominant term (equivalent to using the genotypic model) increases the degrees of freedom (d.f.), and is in general avoided due to expected power loss if the additivity assumption is true or approximately so. Yet, if the absolute power loss is low in comparison to the potential power gain in the presence of non-additivity, genotypic models should be revisited. Likewise, modeling SNP-sex interaction comes at the cost of increased d.f., but may be crucial for X-chromosome analysis. For autosomes, we show analytically that under an additive model, compared with the classical 1 d.f. additive test, the maximum power loss of the 2 d.f. test is capped at 11.4%, regardless of the sample size, genetic effect size and the significance alpha level. With departure from additivity, the power gain can be more than 40%. Similarly, for analysis of the X-chromosome, with a 3 d.f. test that includes a sex-by-SNP interaction term, power loss is capped at 18.8% while the gain can be as high as 60%. Theoretical insights demonstrate why the proposed 3 d.f. regression-based test addresses all the aforementioned X-chromosome analytic complexities. In a sample of over 7,000 individuals with cystic fibrosis from the International CF gene modifier consortium, application of this approach to a new X-inclusive whole-genome scan of meconium ileus provides compelling empirical corroboration.

3442T

A genome wide association study on next-generation sequencing profiled plasma miRNA levels. M. Roux^{1,2}, F. Thibord^{1,2}, C. Perret^{1,2}, M. Ibrahim^{3,4}, L. Goumidi⁴, P. Suchon^{3,4}, M. Germain^{3,4}, J-F. Deleuze⁵, P-E. Morange^{3,4}, D-A. Trégouët^{1,2}, GENMED consortium. 1) Sorbonne Universités, Université Pierre et Marie Curie (UPMC Univ Paris 06), Institut National pour la Santé et la Recherche Médicale (INSERM), Unité Mixte de Recherche en Santé (UMR_S) 1166, Team Genomics & Pathophysiology of Cardiovascular Diseases, 75013 Paris, France, 75013 Paris, France; 2) Institute for Cardiometabolism and Nutrition (ICAN), 75013 Paris, France; 3) Laboratory of Haematology, La Timone Hospital, Marseille, France; 4) Institut National pour la Santé et la Recherche Médicale (INSERM), Unité Mixte de Recherche en Santé(UMR_S) 1062, Nutrition Obesity and Risk of Thrombosis, Center for CardioVascular and Nutrition research (C2VN), Aix-Marseille University, Marseille, France; 5) Centre National de Recherche en Génomique Humaine (CNRGH), Direction de la Recherche Fondamentale, CEA, Institut de Biologie François Jacob, Evry, France.

There is now established evidence that circulating miRNAs can be used as reliable biomarkers for human diseases. The identification of genetic variants that influence levels of circulating miRNAs in the context of miRNA Quantitative Trait Loci (miQTL) can help to better decipher the miRNAs dependent pathophysiological mechanisms. As part of the MARTHA study, we profiled 334 samples for plasma miRNA levels using a next generation sequencing technology and sought for single nucleotide polymorphisms (SNPs) associated with plasma levels of the detected miRNAs. After the application of a complete bioinformatics workflow on sequenced data to detect and quantify miRNAs, 120 miRNAs were found highly expressed in MARTHA plasma samples. Each of these miRNAs underwent a Genome Wide Association Study scan and were tested for association with 6,395,990 imputed (imputation criterion $r^2 > 0.4$) SNPs with minor allele frequency greater than 0.05. Five SNP x miRNA association signals reached the pre-specified $5 \cdot 10^{-8}$ threshold for declaring statistical significance. Two of them were cis effects involving hsa-miR-625-3p ($p = 7.8 \cdot 10^{-16}$) and hsa-miR-941 ($p = 1.1 \cdot 10^{-14}$) that had been previously reported in a miQTL study on whole blood. The remaining 3 associations were never reported trans miQTL effects. One miQTL effect was observed between a *CCSER1* coding SNP and hsa-miR-30e-3p ($p = 1.0 \cdot 10^{-6}$). The two other miQTL associations involved hsa-miR-182-5p whose plasma levels were influenced by intronic SNPs in *KIRREL3* ($p = 1.7 \cdot 10^{-6}$) and in *DOCK8* ($p = 3.8 \cdot 10^{-6}$). Replication and in silico investigation of these 3 trans miQTL associations are ongoing.

3443F

The influence of mitochondrial DNA variation on complex traits. E.J. Grzeszkowiak¹, D.W. Clark¹, P.R.H.J. Timmers¹, P.K. Joshi¹, N. Pirastu¹, J.F. Wilson^{1,2}. 1) Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Teviot Place, Edinburgh, UK; 2) MRC Human Genetics Unit, MRC Institute of Genetics & Molecular Medicine, University of Edinburgh, Edinburgh, UK.

Mitochondrial DNA (mtDNA) is one of the least studied parts of human genome, yet exhibits abundant genetic variation, some of which is structured among ethnic groups. MtDNA is involved in numerous rare Mendelian disorders and some research points to the influence of mtDNA on the regulation of lipid levels. However, the impact of mtDNA variation on complex traits remains largely unexplored. We have conducted an association study between mtDNA variation and a number of quantitative traits in 5 ethnicities comprising ~480,000 men and women. 265 markers have been used to infer haplogroups for each participant. ANOVAs were performed separately on each ethnic group for over 70 phenotypes spanning a wide range of health and anthropometric traits and then meta-analysed. Among many others, we detected a significant association with lifespan. We have also performed an mtWAS analysing all 265 genotyped markers as well as those we imputed using the haplogroup assignments. Our results suggest that mtDNA may play a significant and previously under-estimated role in complex traits and a further study utilising WGS and additional cohorts and ethnicities may be needed to determine if some of the population-specific associations are due to downstream variants whose frequencies differ among populations.

3444W

The genetics of food and drink consumption and correlation with health-related traits.

N. Pirastu¹, E.J. Grzeszkowiak¹, N. Taba², C. McDonnell¹, M.P. Concas³, L. Repetto⁴, K. Kentistou¹, A. Robino³, X. Shen^{1,4}, T. Esko^{2,5}, K. Fisher², J.F. Wilson^{1,6}. 1) Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Teviot Place, Edinburgh, EH8 9AG, Scotland; 2) Institute of Genomics, University of Tartu, Tartu, Riia 23b, 51010, Estonia; 3) Institute for Maternal and Child Health - IRCCS "Burlo Garofolo", Trieste, Italy; 4) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Nobels väg 12A, SE-171 77 Stockholm, Sweden; 5) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; 6) MRC Human Genetic unit, Institute of Genetic and Molecular Medicine, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, EH4 2XU, Scotland.

Several studies have shown that food and drink consumption have a considerable heritable component. With the exception of coffee and alcohol, identifying the genetic variants which are responsible has proven extremely difficult. Here, we conduct genome-wide association on consumption of coffee, alcohol, three genetically-determined food groups (meat, healthy foods and bread & cheese), adding salt to food and preferred drink temperature using food frequency questionnaire data from up to 334,162 subjects in UK Biobank. We identify 226 significant associations corresponding to 134 independent genomic loci, many of which are pleiotropic and 122 of which are completely novel. Several of the loci have been previously linked to numerous late-onset diseases, e.g. *HLA-DQA1*, *APOE*, *NEGR1*, *MC4R*, *GCKR* and *FTO*. Polygenic scores successfully predict food and drink consumption in the Estonian Biobank, validating our findings. Multivariate genetic regression suggests that dietary genetics shows strong genetic correlations with a spectrum of complex traits including obesity, cognition, cancer, personality and lifespan. Gene-set enrichment analyses reveal food choice genes to be preferentially upregulated in the brain and adipose tissue. Stratified LD-score regression shows significant correlation with brain areas involved in the reward system. Finally, clustering the genes previously associated with BMI based on their effects on food reveals groups of genes with specific dietary influences, giving new insights into their roles and the mechanisms underlying obesity and other diseases.

3445T

Hepatic gene expression and DNA methylation associates with African ancestry: Uncovering the role of the genome in disease disparities in African Americans.

C.S. Park¹, Y. Xu², Y. Zhong¹, C. Alarcon¹, T. De¹, M.A. Perera¹. 1) Northwestern University, Chicago, IL; 2) University of Chicago, Chicago, IL.

African Americans (AAs) suffer disproportionately from many diseases and have major differences in hepatic drug metabolism and drug response compared to other populations. As an admixed population, AAs have varying proportions of African and European ancestry. To uncover whether ancestry may explain variabilities in gene expression and thus inform health disparities in AAs, we investigated whether hepatic gene expression and DNA methylation are associated with the degree of African ancestry. We performed genome-wide genotyping, RNA sequencing, and profiled the DNA methylation states of 60 AA-derived primary hepatocytes. In the 64 PharmGKB "very important pharmacogenes" – genes with extensive literature support for association to drug responses – we identified six genes, VDR ($p = 0.041$), PTGIS ($p = 0.030$), NR112 ($p = 0.043$), ALDH1A1 ($p = 0.0095$), ADH1B ($p = 0.047$) and ADH1C ($p = 0.045$), that were significantly associated with African ancestry. We replicated the association of Vitamin D Receptor (VDR) to African ancestry in an independent GTEx liver dataset, such that for every 1% increment in African ancestry, there is a corresponding 1.2% decrement in VDR gene expression. Genome-wide analysis identified 32 genes significantly associated with African ancestry ($FDR < 0.05$). They include genes in coagulation and cardiovascular pathways that are dysregulated in diseases that disproportionately afflict AAs, including ALDH1A1, APOL1, GGT1, NPR2 and PTGIS. In our DNA methylation data, we identified 27,188 significant differentially methylated (DM) CpG sites (BH-adjusted p -value < 0.05) corresponding to 1459 significant DM regions (BH-adjusted p -value < 0.05) associated with greater African ancestry. Notably, and consistent with their gene expression profiles, VDR was identified as a PharmGKB gene that had higher methylated CpG sites with greater African ancestry while ALDH1A1, GGT1 and NPR2 had reduced methylation. Vitamin D is an exogenous hormone created by sun exposure or through diet, with established deficiencies in both Vitamin D and its bioactive metabolites in AAs. Consequently, VDR polymorphisms have been implicated in asthma, cardiovascular diseases and various cancer susceptibilities in AAs as well as peginterferon α -2b response in hepatitis C. Likewise, APOL1 has been strongly associated with kidney disease exclusively in AAs. These findings suggest that ancestry may be as important a predictor of disease and drug response in AAs as known biomarkers.

3446F

NAT2 genetic variants and toxicity related to anti-tubercular agents:

A systematic review and meta-analysis. M. Richardson¹, J. Kirkham¹, K. Dwan², D. Sloan³, G. Davies⁴, A. Jorgensen¹. 1) Department of Biostatistics, University of Liverpool, United Kingdom; 2) Cochrane Editorial Unit, London, United Kingdom; 3) School of Medicine, University of St Andrews, United Kingdom; 4) Department of Clinical Infection, Microbiology and Immunology, University of Liverpool, United Kingdom.

Background Tuberculosis patients receiving anti-tuberculosis treatment may experience serious adverse drug reactions, such as hepatotoxicity. Genetic variants of the *NAT2* gene may increase the risk of experiencing such toxicity events. This systematic review and meta-analysis provides a comprehensive evaluation of the evidence base for associations between *NAT2* genetic variants and anti-tuberculosis drug-related toxicity. The scope of our review is wider than those previously conducted, as we considered associations between all *NAT2* genetic variants (including individual *NAT2* SNPs as well as acetylator phenotypes) and all anti-tuberculosis drug-related toxicity outcomes. Our review and meta-analysis therefore updates and adds to the existing evidence base on associations between *NAT2* genetic variants and anti-TB drug-related toxicity. **Methods** Our review was conducted in line with methods outlined in our protocol (PROSPERO number: CRD42017068448). We searched for studies in Medline, EMBASE, BIOSIS and Web of Science. We performed a qualitative quality assessment of each included study and calculated pooled effect estimates for each genotype on each outcome using meta-analyses. **Results** We included data from 39 distinct patient cohorts. The quality of included studies was variable, with many areas of concern. *NAT2* slow/intermediate acetylators were statistically significantly more likely to experience hepatotoxicity than rapid acetylators (OR=1.59, 95% CI 1.26–2.01, *P*=0%). We also meta-analysed data for six individual SNPs of the *NAT2* gene. For 590G-A and 857G-A, both heterozygous genotype (OR=1.30, 95% CI 1.06–1.59, *P*=0% and OR=1.30, 95% CI 1.03–1.64, *P*=0.9%, respectively) and homozygous mutant-type (OR=2.05, 95% CI 1.24–3.40, *P*=47.7% and OR=1.99, 95% CI 1.02–3.91, *P*=11.3%, respectively) significantly increased hepatotoxicity risk compared with homozygous wild-type. **Conclusions** *NAT2* acetylator status is significantly associated with the likelihood of experiencing anti-tuberculosis drug-related hepatotoxicity. To the best of our knowledge, no meta-analyses on individual SNPs of the *NAT2* gene have been published, therefore our results add to the existing knowledge base. We encountered many challenges in conducting our meta-analyses, particularly due to variability in reporting between individual studies. We outline recommendations for the future reporting of pharmacogenetics studies to enable high-quality systematic reviews and meta-analyses to be performed.

3447W

The composite kernel association test (CKAT) for genetic biomarker analysis in pharmacogenetics studies. J. Shen¹, H. Zhang¹, N. Zhao². 1)

Merck & Co., Inc., Kenilworth, NJ; 2) John Hopkins University, Baltimore, MD.

In Pharmacogenetics (PGx) studies, it is of substantial interest to discover novel genetic biomarkers that influence drug response so that personalized treatment strategies can be developed that maximize therapeutic efficacy and safety. Set-based approach, in which a group of SNPs (e.g., SNPs in a gene) are combined and tested for association with a phenotype, has gained much popularity in PGx studies. In this project, we develop the Composite Kernel Association Test (CKAT), a flexible and robust kernel machine based test, to test the joint effect between a SNP set and an outcome of interest (i.e., drug safety or efficacy response), considering both the genetic main effect and gene-treatment interaction effect for PGx studies. A score test is developed for hypothesis testing. We further develop an analytic p-value calculation procedure so that the computationally extensive procedures (i.e., permutation or perturbation) via simulation of the null distribution can be avoided. We evaluate the CKAT through extensive simulations and application to the gene-level association test of the reduced risk in *Clostridium difficile* Infection recurrence in patients treated with bezlotoxumab compared with placebo. We show that the proposed CKAT works well for PGx studies with relatively small sample size (i.e., ≤ 1000), controls the type I error at genome-wide significance level, and is more powerful than existing approaches such as rareGE and MGME across multiple scenarios. Further, CKAT readily accommodates any types of kernels, and works for both continuous and binary outcomes.

3448T

HCLC: A novel statistical method for phenome-wide association studies.

X. Liang, Q. Sha, S. Zhang. Department of Mathematical Sciences, Michigan Technological University, Houghton, MI 49931.

The emergence of genetic data coupled to longitudinal electronic medical records (EMRs) offers the possibility of phenome-wide association studies (PheWAS) for disease-gene associations. In PheWAS, the whole phenome can be divided into numerous phenotypic categories according to the first three digits of the International Classification of Disease (ICD) code. Currently, statistical analyses for PheWAS is mainly univariate analysis, which tests the association between one genetic variant and one phenotype at a time. In this article, we derived a novel and powerful multivariate method, which we referred as HCLC, to test the association between a genetic variant with multiple phenotypes for each phenotypic category. The proposed method (HCLC) involves three steps. In the first step, we use the bottom-up hierarchical clustering method (HCM) to partition a large number of phenotypes into disjoint clusters within each category. In the second step, we apply the clustering linear combination method (CLC) to combine test statistics within each category based on the phenotypic clusters and obtain p-values from each phenotypic category. In the third step, we propose a large-scale association testing procedure with theoretical guarantees for false discovery rate (FDR) control under flexible correlation structures (Cai et al., 2015). We show that applying our new FDR-controlling procedure on the entire collection of hypotheses can control the rate of FD of associated genetic variants as well as the expected value of the average proportion of FD of phenotypic categories influenced by such variants. We perform extensive simulation studies to compare the performance of HCLC with that of other existing methods. The results show that our proposed HCLC method controls FDR very well and outperforms other methods we compared with.

3449F

Reducing the blind spots in target validation, examples from UK Bio-

bank. L.M. Yerges-Armstrong¹, K. Song¹, M.R. Nelson¹, T. Johnson², J.C. Whitaker², R.A. Scott². 1) GlaxoSmithKline, Collegeville, PA, USA; 2) GlaxoSmithKline, Stevenage, UK.

Deeply-phenotyped genetic resources like UK Biobank are enabling large-scale drug discovery efforts. While these efforts are accelerating new target identification, they can also be invaluable for target validation by using genetic variants to mimic pharmacological modulation of a drug target. While there are established examples of functionally validated, instrument variants which mimic the pharmacological perturbation of a drug target, such variants are exceedingly rare. Thus, increasing our ability to instrument genes would accelerate genetic target validation efforts. Here, we explore the potential of multi-SNP predictors of gene expression to provide an alternative source of target validation evidence. We evaluated 21 genes with low frequency predicted loss-of-function (pLOF) variants previously reported for association with phenotypes in UK Biobank (PMID: 29691411) to see if multi-SNP predictors of expression would recapitulate reported associations. Specifically, we used the PrediXcan method, applying an omnibus test of association across all tissues. 17 of the 21 genes replicated the reported association signal ($p < 5 \times 10^{-6}$) despite the specific pLoF not being tested directly. We then extended our findings for each of these genes with a PheWAS of 157 quantitative and 2254 case-control traits. The multi-SNP predictor for *ENPEP*, for example, replicated association for both diastolic ($p = 6 \times 10^{-6}$) and systolic blood pressure ($p = 3.7 \times 10^{-6}$) and was also associated with atrial fibrillation ($p = 6.5 \times 10^{-20}$). Further work to determine the utility and specificity of this approach is underway, this proof of concept demonstrates that instrumentation methods based on gene expression will reduce the blind spots in genetic target validation.

3450W

Investigating fine-scale population structure in the UK biobank. *J.P.*

Cook, A.P. Morris. Department of Biostatistics, University of Liverpool, Liverpool, United Kingdom.

The United Kingdom (UK) is a genetically diverse population with strong genetic differences between regions, which may adversely affect genome-wide association studies (GWAS) of complex traits if not accounted for in the analysis. Principal components (PCs), calculated from a genetic relatedness matrix (GRM), are routinely included as covariates in regression models to account for population structure in GWAS. Linear mixed models (LMMs), which account for population structure by including a random effect for the GRM, have also become commonly used in GWAS. UK BioBank provides an opportunity to examine fine-scale UK population structure in unprecedented detail, with the final release of genetic data consisting of ~500,000 individuals recruited from 23 centres. Easting and Northing (E&N) coordinates were collected for every participant at recruitment and birth, and genotypes were imputed up a combined reference panel from the Haplotype Reference Consortium, 1000 Genomes Project and UK10K Project. We performed regression analyses of E&N at birth, with and without adjustment for PCs, and using an LMM approach implemented in BOLT-LMM. Genome-wide inflation factors for N in regression analyses were extremely high without PC adjustment ($\lambda=17.76$) and even when using 10 PCs ($\lambda=4.25$). Residual inflation also remained after accounting for population structure in the LMM analysis ($\lambda=1.76$). To allow for linkage disequilibrium (LD), we applied LD score regression and LD adjusted kinship analyses to the LMM results, but residual inflation remained: intercepts of 1.65 and 1.60 respectively. Variants mapping in *TLR1* showed the strongest signals of association with E&N genome-wide, including two variants previously associated with other asthma (rs4833095) and alcohol tolerance (rs4543123). We analysed rs4833095 with a self-reported asthma phenotype in UK BioBank including 40,020 cases and 336,697 controls. The PC unadjusted association ($p=5.1 \times 10^{-13}$) strengthened when adjusting for population structure using either 10 PCs ($p=5.9 \times 10^{-15}$) or a LMM (1.8×10^{-13}), demonstrating both the robustness of the association and the benefit of removing as much geographical confounding as possible. Our study has important implications for accounting for population structure in large scale GWAS performed in UK BioBank, and highlights the need for caution in interpreting association results from different analytical approaches.

3451T

Genetic clusters and population stratification in the Estonian Biobank and the association with complex traits. *K. Fischer^{1,5}, N. Pervjakova^{1,2,3}, S.E.*

Ojavee⁵, K. Läll^{1,5}, M. Mändul^{1,5}, A.P. Morris^{1,4}, R. Mägi¹. 1) Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu, Estonia; 2) Department of Biotechnology, Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia; 3) Genomics of Common Disease, Division of Diabetes, Endocrinology and Metabolism, Department of Medicine, Imperial College London, UK; 4) Department of Biostatistics, University of Liverpool, Liverpool, UK; 5) Institute of Mathematics and Statistics, University of Tartu, Estonia.

It has been shown that principal components (PCs) derived from Multidimensional Scaling (MDS) analysis of the genetic relatedness matrix primarily reflect the geographic ancestry of an individual. Whilst PCs have been thought to be appropriate for accounting for population structure in genetically heterogeneous populations (e.g. across Europe), the availability of large-scale GWAS data in population biobanks offers the opportunity to investigate the utility of this approach to correct for fine-scale stratification and cryptic relatedness. We analysed 39,769 participants from the EGCUT cohort, genotyped with a range of Illumina arrays. The individuals are sampled from all counties of Estonia (the area of the country is 45000m² and with a population ~1.2 million). Only the data of ethnic Estonians (self-identified) was analysed. We calculated the first 250 PCs using 61,044 independent variants. First, we use K-means cluster analysis to demonstrate that genetically more similar clusters, on the basis of genetic PCs, can be associated with 18 distinct areas of Estonia, despite the small size of the country. As most of the first 250 PCs were significantly associated with the county of birth, we show that it is most efficient to use all of them in the k-means algorithm. Next, we tested the association of cluster membership, actual county of birth, and individual PCs with prevalent coronary heart disease (CAD, 3152 cases) and type 2 diabetes (T2D, 1691 cases). The first and third PC were significantly associated CAD and only the first one with T2D. Also the predicted cluster membership as well as the actual county of birth were associated with both diseases ($p < 5 \times 10^{-5}$). While adjusting for the actual county of birth, the effect of PCs disappeared for both diseases, but the effect of cluster membership was still significant for T2D. Our results indicate that although PCs may help to adjust for population structure, a genetic cluster indicator may be more efficient in capturing the confounding due to geographic variation. However, to also account for cryptic relatedness, a mixed modelling framework with a random effect for the genetic relatedness matrix may still be optimal.

3452F

Extensive fine-scale population structure in the UK Biobank. S. Hu¹, G. Hellenthal², J. Marchini^{1,3}, D. Lawson⁴, S. Myers^{1,3}. 1) Department of Statistics, University of Oxford, Oxford, United Kingdom; 2) UCL Genetics Institute (UGI), University College London, London, United Kingdom; 3) Wellcome Center for Human Genetics, University of Oxford, Oxford, United Kingdom; 4) University of Bristol, Bristol, United Kingdom.

We developed and applied a novel, rapid, scalable statistical ancestry inference pipeline and applied this to the 500,000 UK Biobank samples. The approach breaks down the fine-scale ancestry of a single genome across 21 regions within the UK and Republic of Ireland, and 54 additional regions worldwide. This provides much greater spatial resolution (~70 miles) than existing approaches, distinguishing e.g. northern and southern parts of central England. Among UK-born participants (430,256 self-reported British), inferred UK ancestry was 96.15% vs. 3.85% for participants with non-UK ancestry. For 343,217 UK-born individuals possessing ≥95% UK ancestry, we compared their place of birth to their genetic ancestry breakdown. On average 28.82% of their ancestry derives from their area of birth, varying regionally from 16–46%. Overall, 30.1% of individuals mainly show ancestry (>50%) from a single region. For these people, this region matches their birthplace region fully 71% of the time (odds ratio 48 vs. random chance). Migration is complex: some regions exhibit limited migration. For example, individuals born in South Yorkshire have mean ancestry from South and North Yorkshire of 40.6% and 5.6% respectively. While for individuals born in North Yorkshire the mean ancestries are 14.6% and 16.2% respectively. Other regions, including London, show much higher levels of mixing: only 20.1% of London-born individuals mainly show ancestry from one region, and this is SE England for only 32% of these. Thus, despite recent population movements, substantial fine-scale population structure persists in the UK among the general population and country-wide, and this can be identified based on DNA information alone. Because our ancestry vectors predict (median $r^2=0.80$) but are not predicted by (median $r^2=0.49$) standard principal components, they might offer advantages in correcting for population stratification, particularly for rare variants, in phenotypic association testing. We identified particular mutations showing strong regional differences among UK ancestry regions, and many rare variants show geographic variability, with the maximal estimated frequency difference among regions being 74.74%.

3453W

Existence and implications of ancestry-dependent phenotypic variance. S. Musharoff¹, D. Park², J. Galanter³, X. Liu⁴, S. Forsberg⁵, S. Huntsman⁶, C. Eng¹, E.G. Burchard^{1,6}, J.F. Ayroles⁵, N. Zaitlen¹. 1) Lung Biology, University of California, San Francisco, San Francisco, CA; 2) Encompass Bioscience Inc; 3) Genentech, South San Francisco, CA; 4) Department of Human Genetics, University of Chicago, IL; 5) Ecology and Evolutionary Biology and Lewis-Sigler for Integrative Genomics, Princeton University, Princeton, NJ; 6) Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, CA.

Genetic and environmental differences between worldwide populations produce concomitant variability in the distributions of complex phenotypes and disease prevalence. Studies of worldwide variability to date have focused on the relationship of ancestry and phenotypic mean. Here, we develop a novel framework, the ancestry double generalized linear model (ADGLM), to model phenotypic variance as a function of ancestry. Such ancestry-dependent variance is induced by gene-gene and gene-environment interactions, differential environmental distributions, and, as we show for the first time, differential selection. When applied to ~113K British individuals, ADGLM reveals ancestry (genetic PC1) is associated with the variance of 28 traits including BMI ($p=1.0 \times 10^{-3}$), hemoglobin ($p=7.8 \times 10^{-28}$), red hair color ($p=1.3 \times 10^{-39}$), and skin pigmentation ($p=1.6 \times 10^{-28}$). In ~3K Mexican and Puerto Rican individuals, admixture fraction is associated ($p<6.6 \times 10^{-4}$) with the variance of height, eczema, lung function, and asthma. Importantly, we show that such ancestry-variance associations can produce up to a three-fold difference in disease prevalence between populations while minimally affecting heritability. They may therefore partially account for the 2.8-fold difference in asthma prevalence between Mexicans (8%) and Puerto Ricans (22%) living in the United States. Given this observation of ancestry-dependent variance, we consider its implications for genetic association studies. In GWAS simulations with no genetic effect, linear regression with principal components (LR+PC) and linear mixed model association tests are miscalibrated at population-differentiated SNPs ($\lambda_{OC}=1.56, 0.46$) while ADGLM is calibrated ($\lambda_{OC}=1.06, 0.98$). When there is a genetic effect, ADGLM has up to 17% more power than LR+PC. This gain is observed in data from Mexicans and Puerto Ricans: ADGLM has more significant p-values than LR+PC at 10 of 11 known genetic associations for eczema, rash, rhinitis, skin pigmentation, tanning, and lung function, and ADGLM identifies two novel eczema and rash associations. Furthermore, ADGLM allows direct association testing of genotype and phenotypic variance and finds a novel SNP association with tanning variance ($p=1.2 \times 10^{-8}$) in African-Americans. In conclusion, ancestry-dependent variance is prevalent in human populations and can bias association studies if ignored. Modeling this variance enables discoveries along an understudied axis of phenotypic variation and disease.

3454T

Functional characterization of 3D-protein structures informed by human genetic diversity. *I. Bartha¹, M. Hicks², J. di Iulio³, A. Telenti³, JC. Venter⁴.*

1) Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland; 2) Human Longevity Inc. San Diego CA USA; 3) The Scripps Research Institute, San Diego, CA, USA; 4) J Craig Venter Institute, San Diego, CA, USA.

Sequence variation data of the human proteome can be used to analyze 3-dimensional protein structures to derive functional insights. We use human genetic variation from nearly 140,000 human exomes and genomes and 4,390 x-ray protein structures to model tolerance to amino acid changes in the 3D space of the human proteome. To understand variation in the structural proteome, we first identified 26,593 structures associated with 4,390 Uniprot entries that had x-ray crystal structures with a defined resolution and a minimum chain length greater than 10 amino acids. In total, we mapped 139,535 Uniprot features to the structures, and extracted a 3-dimensional context by defining a 5-Angstrom radius space for each feature. We identified 481,708 missense variants for these proteins from the analysis of 138,632 individuals' exomes. From these contextualized data, we constructed a model that describes functional constraints in three-dimensional protein structures. The strength of constraint (intolerance) was reflected in a 3-dimensional tolerance score (3DTS) that summarized the differences between observation and expectation in genetic variation at the level of 3D-sites. We confirmed the utility of structural intolerance data by comparing it to available deep mutational scanning data. The intolerance data correlated with deep mutational scanning functional read-outs for PPARG, MAPK1/ERK, UBE2I, SUMO1, PTEN, CALM1, and CALM2 ($r^2=0.23$ to 0.81). In conclusion, the increasing detail of the limits of protein diversity that can be gathered through large scale sequencing of the human population provides valuable information on the functional consequences and pathogenicity of genetic variants and offers new data on orthosteric, allosteric and additional functional sites that could be harnessed for drug development.

3455F

A unified method for rare variant analysis of GxE interactions with application to GxSmoking interactions on obesity-related traits. *E. Lim¹, H. Chen^{2,3}, CT. Liu¹.*

1) Department of Biostatistics, Boston University, Boston, MA; 2) Human Genetics Center, Department of Epidemiology, Human Genetics and Environmental Sciences, School of Public Health, University of Texas Health Science Center, Houston, TX; 3) Center for Precision Health, School of Public Health and School of Biomedical Informatics, University of Texas Health Science Center, Houston, TX.

Majority of single nucleotide polymorphisms (SNP) discovered by genome-wide association study only explain a small proportion of heritability, and thus much of the heritability is left unexplained. Gene-environment (GE) interactions, or the presence of rare variants may have substantial contributions in explaining missing heritability. Advanced technology in whole-genome sequencing has offered the opportunity to comprehensively investigate the genetic contribution, particularly rare variants, to complex traits. Many rare variants analysis have focused on jointly modeling the marginal effect, but the method to detect GE interactions, especially for rare variants, is underdeveloped. Identifying the modification effects of environmental factors on genetic risk poses a considerable challenge. To tackle this challenge, we developed a unified method in detecting GE interactions of a set of rare variants using generalized linear mixed effect model. The proposed method can accommodate both binary and continuous traits in related or unrelated sample. Under this model, genetic variants, sample relatedness and the GE interaction were modeled as random effects. We adopted a kernel-based method to leverage the joint information across the rare variants and implemented a variance component score test, which only requires fitting the model under the null hypothesis, and thus, reducing multiple testing and computational burden. Since the integrated quasi-likelihood function for binary traits contains high dimensional integral that is intractable, we use Laplacian method to approximate the integral and iteratively estimate our parameters with Fisher scoring algorithm. Our extensive simulation study shows that the proposed method maintains correct type I error and high power under various scenarios, such as differing the direction of main genotype effects and the number of variants in the model for both continuous and binary traits. We illustrated our method to test gene-based interaction with smoking on body mass index (BMI) or overweight ($BMI \geq 25$) in 7188 Framingham Heart Study subjects and replicated the *CHRNA4* gene reported in previous large consortium meta-analysis of common SNP-smoking interaction. Our proposed set-based GE test is computationally efficient and is applicable to both binary and continuous phenotypes, while appropriately accounting for familial relationships.

3456W

Enabling cost-effective multi-sample single cell RNA-seq studies via in-silico demultiplexing and doublet detection without external genotype data. T.H. Yao, H.M. Kang. Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI.

The advent of massive parallel single-cell RNA sequencing (scRNA-seq) technologies dramatically enhances the resolution to understand how genetic and environmental factors alter transcriptomic profiles of individual cells and affect interactions between them. To accurately understand the effect of genetic and environmental variations on each cell, it is very important to rapidly escalate the scale of scRNA-seq studies, both experimentally and analytically. Recently, we developed a experimental workflow with a computation tool, demuxlet, to sequence single cell transcriptomes across multiple samples in a single library preparation by harnessing natural genetic variations. Our strategy reduce the per-cell and per-sample cost of scRNA-seq experiments by several folds while avoiding batch effects between multiplexed samples. Although our sample multiplexing strategy is rapidly becoming popular for large-scale scRNA-seq studies, the underlying statistical method, demuxlet, requires external genotypes to be collected, thus not applicable to circumstances such as model organisms or samples without separate DNA specimen available. Moreover, steps to collect, process, and impute the genotype data makes the multiplexing procedure less seamless and convenient than the standard scRNA-seq workflow. We develop a new method freemuxlet to demultiplex and to detect doublets from multiplexed scRNA-seq reads without requiring external genotypes. The major challenge in developing freemuxlet is that many pairs of droplets do not have enough overlapping reads, making pairwise distance between them uncertain. We define the distance between two droplets as Bayes Factors (BF) between the two hypotheses of sample origin to account for uncertainty. Multi-class clustering guided by hierarchical t-SNE followed by Gibbs sampling simultaneously estimates the consensus genotypes of each individual while deconvoluting the sample provenance of each droplet. We apply these methods to the 8-sample mixture dataset published in our previous study, and observed that 97% of droplets that were estimated as singlet using the original demuxlet (with external genotypes) have concordant sample deconvolution assignment with the Bayes Factor clustering results. Our results suggest that multiplexed scRNA-seq experiment can be more widely applied across a variety of species, sample conditions, and cell types much more seamlessly through freemuxlet.

3457T

MMP3 gene variant is a promising biomarker for periodontal disease among Hispanics. V. Navarrete¹, I. Ramirez¹, F.J. Estrada², A.C. Perez^{2,3}, E. Lira². 1) Seccion de Estudios de Posgrado e investigacion, Escuela Superior de Medicina, Instituto Politecnico Nacional. Plan de San Luis y Diaz Miron s/n Casco de Santo Tomas 11340, Mexico; 2) Universidad Panamericana. Escuela de Medicina. Donatello 59, Ciudad de México, 03920, México; 3) Yale University School of Public Health, Laboratory of Epidemiology and Public Health. 60 College St, New Haven, CT, United States of America.

Genetic and environmental factors contribute to the development of periodontal disease (PD). In bioinformatics analyses, *MMP3* is a promising candidate gene associated with the development and prognosis of periodontitis. However its remains unclear whether single nucleotide polymorphisms (SNPs) in this gene are associated with PD, especially among Hispanics. Hence, we performed a case-control study of 134 periodontitis cases and 134 controls from a referral center in Mexico City. All our cases were evaluated and selected by experienced DMDs based on current guidelines. Typical patients, with chronic periodontal disease, had evidence of supragingival and subgingival dentobacterial plaque, stone formation, gingival inflammation, pocket formation, loss of periodontal attachment, and loss of alveolar bone. We assessed a previously reported PD associated polymorphism, rs679620 (top scoring SNP in bioinformatics analyses), by TaqMan probes on RT-PCR. Genotyping quality check, descriptive, univariate, and multivariate analyses were performed with PLINK v.1.9 and corroborated in R v.3.3.2. Our preliminary results indicate that allele T- (MAF = 0.35) compared to C-carriers have significantly ~61% lower odds of PD. Furthermore, there is a trend towards decreasing odds of periodontal disease on genotypic and trend test models. Taken all together, we propose that the rs679620 of *MMP3* might be a promising good prognosis biomarker for PD in Mexican population. Our results warrant further replication and validation in other race/ethnicities, especially of its predictive applicability in the DMD office setting.

3458F**Two-phase sampling designs for post-GWA fine-mapping studies:**

Optimality criteria. O. Espin-Garcia^{1,2}, R.V. Craiu³, S.B. Bull^{1,2}. 1) Dalla Lana School of Public Health, University of Toronto, Toronto, Canada; 2) Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, Canada; 3) Department of Statistical Sciences, University of Toronto, Toronto, Canada.

Given typically modest genetic effect sizes, large-scale population studies are generally necessary to achieve genome-wide significance. These studies require careful design and budgeting. Post-GWAS analysis focuses on fine-mapping of targeted genetic regions, but as yet costs of next-generation sequencing remain prohibitively expensive for large studies. Two-phase sampling design and analysis is a cost-reduction technique that utilizes collected data during phase 1 GWAS (e.g. outcome, auxiliary variable) to select an "informative" (biased) subsample in phase 2 fine-mapping. The main goal is to make inference on a genetic variable that is costly and/or onerous to measure (e.g. sequencing data). In two-phase sampling designs less attention has been paid to determination of optimal study designs compared to effect estimation and hypothesis testing. To date, most of the work on optimal designs has focused on case-control studies, where for example, a "balanced" design has been claimed to be "near optimal". Strategies and tools that help researchers to select the (optimal) phase 2 subsample are not readily available beyond binary outcomes. Considering the effect size and haplotype distribution between a GWAS-auxiliary-variable and a sequence variant as inputs, we develop two approaches for optimal two-phase designs within the exponential family under a semiparametric maximum likelihood framework. One is based on analytic expressions for the variance-covariance matrix (VCM), and alternatively, another is based on a computational approach via a genetic algorithm for optimization of the phase 2 sampling space. We evaluate measures of statistical efficiency based on features of the VCM that have been proven useful to define optimality criteria in design of studies.

3459W**Detection of genetic similarities using unsupervised random forest.**

C.J.K. Fouodo, I.R. Koenig. Institut für Medizinische Biometrie und Statistik, Universität zu Lübeck, Universitätsklinikum Schleswig-Holstein, Campus Lübeck, Lübeck, Germany.

Genome-wide association studies (GWAS) have in the past been successful in the identification of associations with well-defined phenotypes as well as in the establishment of supervised classification models based on univariable analyses. To perform multivariable genome wide analyses in GWAS data, classical statistical approaches like linear or logistic regression models are not suitable due to the high-dimensional nature of the data. As an alternative, random forest (RF) based approaches are known to be one of the more appropriate supervised machine learning methods for high dimensional multivariable data analysis. RF based approaches have been shown to be fast and to produce good predictive performances in high dimensional classification problems. Furthermore, the method can be extended to unsupervised learning analysis. Our aim is to apply an unsupervised random forest (URF) strategy to GWAS data to cluster individuals into genetically homogeneous subgroups. Incorporating a novel boosting approach, we further expect to obtain more stable clusters and improve the performance of the method provided each boosting iteration takes into account the result of the previous iterations. Clustering individuals according to their genetic similarities and without knowledge of a phenotype can allow us to discover novel subgroups of patients. The evaluation of our approach is demonstrated in an application to samples from the Pan-Asia SNP Consortium database with the objective to validate the derived clusters against the known ethnical background.

3460T

Genetic risk scores for COPD: Identifying high risk individuals and understanding disease pathways. P. Sakornsakolpat^{1,2}, B.D. Hobbs^{1,3}, M. Parker¹, D. Prokopenko¹, L.P. Hayden¹, N. Shrine⁴, L.V. Wain⁴, M.D. Tobin⁴, E.K. Silverman^{1,3}, M.H. Cho^{1,3}. 1) Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 2) Department of Medicine, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand; 3) Division of Pulmonary and Critical Care Medicine, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 4) Department of Health Sciences, University of Leicester, Leicester, UK.

Chronic obstructive pulmonary disease (COPD) susceptibility is influenced by genetics. Identifying individuals at risk for COPD will impact public health and aid in the understanding of disease. We developed a genetic risk score for COPD by applying PRSice-2 to a case-control analysis in the UK Biobank (N=200,794), and applied this risk score to individuals in an independent cohort, COPDGene (N=6,760). We performed linkage disequilibrium (LD) clumping of the summary statistics from the UK Biobank and chose a P value threshold that best fitted to the training dataset of COPDGene. We included 15,155 genetic variants with $P_{\text{GWAS}} < 0.0467$ in the risk score. We investigated the association of this risk score with various COPD-related phenotypes in the test dataset of COPDGene. The genetic risk score alone had an area under the curve (AUC) of 0.60 in predicting COPD case status ($P=1.6 \times 10^{-12}$). The AUC of a model combining genetic risk score and clinical variables (age, sex, and smoking pack-years) was significantly larger than the AUC of a clinical model alone (0.77 vs 0.75, respectively; $P=2.2 \times 10^{-16}$). This genetic risk score significantly predicted forced expiratory volume in one second (FEV_1), ratio of FEV_1 to forced vital capacity, % emphysema, % gas trapping, and severe COPD exacerbations ($P < 0.05$). We also developed a functional risk score by imputing individuals' genetic data to gene expression using PrediXcan. We computed 186 functional risk scores from KEGG pathways, by aggregating imputed gene expression of genes within the same set. We identified significant associations at false discovery rate (FDR) $< 10\%$ between the pentose phosphate pathway and four phenotypes (COPD, FEV_1 , FEV_1/FVC , airway wall area) with $P < 1.1 \times 10^{-4}$, as well as a gene set related to arrhythmogenic right ventricular cardiomyopathy (ARVC) and COPD exacerbation frequency ($P < 6.2 \times 10^{-5}$). The latter gene set includes genes encoding desmosomal proteins; notably, an eQTL for desmoplakin (*DSP*) is associated with COPD in prior investigations. In conclusion, we demonstrated the utility of a genetic risk score to predict COPD and COPD-related phenotypes. An approach combining risk scores with gene expression implicates molecular pathways that may lead to further understanding of the disease.

3461F

Sequential testing of multivariate phenotypes associated with SNP-sets by mixed model kernel statistics. D.J. Schaid, J.P. Sinnwell, J. Chen. Health Sciences Research, Mayo Clinic, Rochester, MN.

Detecting genetic pleiotropy and understanding its causes can improve the biological understanding of a gene in multiple ways. Current statistical methods to evaluate the association of multiple traits with a set of SNPs (e.g., gene-level tests) are based on kernel statistics that test the null hypothesis that none of the traits are associated with a variant. Rejecting the null hypothesis does not provide information on which traits are associated with a set of SNPs. We developed a formal testing framework to determine which traits are associated with a set of SNPs, while accounting for correlations among the traits. Our methods are based on sequential testing with mixed models, starting with the usual null hypothesis that none of the traits are associated with a set of SNPs, using multivariate kernel statistics. If this hypothesis is rejected by a small p-value, we proceed by evaluating all possible ways of allowing one associated trait while testing the remaining traits, using a special kernel statistic. The maximum p-value from these tests is used as a summary p-value. If the maximum p-value is smaller than the nominal criterion (e.g., p-value < 0.05), we continue onto the next sequential step of evaluating all possible ways of allowing two associated traits and testing the remaining traits, etc. By simulations, we illustrate that the Type-I error rate is well controlled, and illustrate power of our new methods. We illustrate how the sequential method is influenced by sample size, the number of traits, and the trait correlations. We apply the new methods to multivariate phenotypes to illustrate interpretation and benefits of our new approach. This strategy provides new ways to perform gene-level analyses for multiple traits while accounting for correlations among the traits.

3462W

Incorporating genetic networks into case-control association studies with high-dimensional DNA methylation data. *H. Sun, K. Kim.* Pusan National University, Busan, South Korea.

In human genetic association studies with high-dimensional gene expression data, it has been well known that statistical selection methods utilizing prior biological network knowledge such as genetic pathways and signaling pathways can outperform other methods that ignore genetic network structures in terms of true positive selection. In recent epigenetic research on case-control association studies, relatively many statistical methods have been proposed to identify cancer-related CpG sites and their corresponding genes from high-dimensional DNA methylation array data. However, most of existing methods are not designed to utilize genetic networks although methylation levels among linked genes in the networks tend to be highly correlated with each other. In this article, we propose new approach that combines independent component analysis with network-based regularization to identify outcome-related genes for analysis of high-dimensional DNA methylation data. The proposed approach first captures gene-level signals from multiple CpG sites using independent component analysis and then regularizes them to perform gene selection based on prior biological network information. In simulation studies, we demonstrated that the proposed approach overwhelms other statistical methods that do not utilize genetic network information in terms of true positive selection. We also applied it to the 450K DNA methylation array data of the four breast invasive carcinoma cancer subtypes from The Cancer Genome Atlas (TCGA) project.

3463T

Co-localization analyses provide counterintuitive findings: Application to cystic fibrosis lung disease. *F. Wang^{1,7}, G. He^{2,7}, J. Gong², N. Panjwani², B. Xiao², F. Lin², J. Avolio³, J.M. Rommens^{2,4}, L. Sun^{1,5}, L.J. Strug^{2,5,6}.* 1) Department of Statistical Sciences, University of Toronto, Toronto, Canada; 2) Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, ON, Canada; 3) Translational Medicine, Hospital for Sick Children, Toronto, Canada; 4) Respiratory Medicine, Hospital for Sick Children, Toronto, Canada; 5) Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada; 6) Biostatistics Division, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada; 7) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, ON, Canada; 7) These authors contributed equally to this work.

Genome-wide association study (GWAS) by the International Cystic Fibrosis (CF) Gene Modifier Consortium (ICFGMC) identified SNPs between *SLC6A14* and *AGTR2* on chromosome X as contributing to CF lung disease. In the same region, we conducted expression quantitative trait loci (eQTL) analysis using FastQTL with RNA-sequencing of human nasal epithelial cells from 63 CF Canadians enrolled in the ICFGMC. The *SLC6A14* eQTL p-values mirror the lung association pattern in the region, providing strong visual evidence that the *SLC6A14* expression mediates CF lung disease. Complications arise when we formally test the eQTL-GWAS co-localization. Associated SNPs at the *AGTR2/SLC6A14* locus span a 150kb region with high linkage disequilibrium (LD) and providing evidence of allelic heterogeneity. Detecting colocalization in the presence of high LD and/or allelic heterogeneity is challenging for several commonly used Bayesian approaches, such as COLOC and eCAVIAR, that estimate the posterior probability of co-localization under specific causal variation assumptions. COLOC assumes a single causal variant in both eQTL and GWAS, and the presence of allelic heterogeneity reduces the co-localization posterior probability. eCAVIAR can accommodate allelic heterogeneity, however the posterior probability of co-localization will be averaged across several variants in high LD, resulting in ambiguous colocalization conclusions. We recently developed iRegress, a Frequentist approach that assesses the pattern similarity between the GWAS and eQTL evidence through regression and incorporates a robust covariance estimator to account for LD. Although iRegress can accommodate allelic heterogeneity and LD, and shows benefit over the other methods in certain scenarios, high LD at the *SLC6A14/AGTR2* locus reduces the effective sample size limiting the power of testing the co-localization. Application of an intuitive sum of test statistics that compares the difference in averaged GWAS test statistics between eQTL SNPs and non-eQTL SNPs, accounting for LD through a mixture of chi-square distributions demonstrates strong evidence of co-localization ($p\text{-value} = 9.5 \times 10^{-9}$) consistent with the visual pattern. Our findings suggest that no one method can accommodate all possible genetic architectures for co-localization analysis, and simple procedures can complement existing state-of-the-art methods to better understand mechanisms underlying GWAS loci.

3464F

A general statistic to test an optimally weighted combination of common and/or rare variants in association studies. *J. Zhang¹, H. Chen², B. Wu³, X. Wang¹.* 1) Department of Mathematics, University of North Texas, Denton, TX; 2) Texas Academy of Mathematics & Science, University of North Texas, Denton, TX; 3) School of Public Health, University of Minnesota, Minneapolis, MN.

Both traditional genome-wide association study and next generation sequencing data analysis are increasingly used to interrogate the whole genome to comprehensively identify both common and rare disease variants in many large genetic studies. Rare variants generally have large effects though are hard to detect due to their low frequency. Currently, many existing statistical methods for rare variants association studies have employed a weighted combination scheme, which usually puts subjective weights or suboptimal weights based on some ad hoc assumptions (e.g. ignoring dependence between rare variants). In this study, we analytically derive optimal weights for both common and rare variants in a gene or pathway. We propose a General and novel approach to Test association between an Optimally Weighted combination of variants (G-TOW) in a gene or pathway and a continuous or dichotomous trait while easily adjusting for covariates. G-TOW is robust to the directions of effects of causal variants. We conduct extensive simulation studies to evaluate the performance of G-TOW. G-TOW has properly controlled type I error rates. Simulation results demonstrate that it is the most powerful test comparing with existing methods, when testing effects of either both rare and common variants or rare variants only. We also illustrate the effectiveness of G-TOW using the Genetic Analysis Workshop 17(GAW17) data.

3465W

Predictive modeling of gene expression in admixed populations. *A.V. Mikhaylova, T.A. Thornton.* Department of Biostatistics, University of Washington, Seattle, WA.

In recent years, several gene-based association methods, such as PrediXcan, have been developed. These methods provide valuable insight into relationships between genetic data and complex traits. However, PrediXcan models were trained on the Depression and Gene Network (DGN) and the Genotype-Tissue Expression (GTEx) subjects, the majority of which are of European descent. It is thus not clear how accurate the gene expression predictions would be in diverse populations. We use transcriptomic data from the GEUVADIS (Genetic European Variation in Health and Disease) RNA sequencing project and whole genome sequencing data from the 1000 Genomes project to compare a range of models from PrediXcan PredictDB and evaluate their predictive performance on one non-European (Yoruban) and four European populations. We show that predictive performance varies across populations, with the Yoruban (YRI) sample having the lowest correlations between the observed and predicted gene expression values, on average. To understand and quantify the level of transferability of prediction models, we perform a series of simulations using real genotype data and evaluate prediction accuracy in various settings: 1) where a training set and a test set come from populations with similar ancestries, 2) a training and a test set come from population with different ancestries, and 3) a test set consists of admixed subjects. Our findings indicate that predictive accuracy is lower when using a training set with ancestry different from the test set and, thus, demonstrate the importance of including more diverse samples in genetic studies. Finally, we propose a new approach of modeling gene expression that takes ancestry into account in order to improve prediction accuracy for admixed populations.

3466T

The intersection of genetics and longitudinal survey data from the health and retirement study: A preliminary study of glaucoma. *J.N. Cooke Bailey, T.G. Kinzy, N.K. Schiltz.* Population and Quantitative Health Sciences, Case Western Reserve University School of Medicine, Cleveland, OH.

Glaucoma is a leading cause of irreversible blindness worldwide; in the United States alone, over 2.7 million individuals are affected. Various risk factors for glaucoma are known and include age, race/ethnicity, genetics, and ocular measures. Despite numerous studies, molecular and environmental factors that contribute to glaucoma remain elusive. The University of Michigan Health and Retirement Study (HRS) is a longitudinal survey of a representative sample of Americans over the age of 50. Supported by the National Institute on Aging and the Social Security Administration, the HRS is designed to provide reliable data on the decisions, choices, and behaviors of people as they age and respond to changes in public policy, the economy, and health. The study obtains information every two years about income and wealth, health and use of health services, work and retirement, and family connections. Through its unique and in-depth interviews, the HRS provides an invaluable and growing body of multidisciplinary data that researchers can use to address important questions about the challenges and opportunities of aging. Because of its innovation and importance, the HRS has become the model and hub for a growing network of harmonized longitudinal aging studies around the world. Saliva was collected on half of the HRS sample each wave starting in 2006 and respondents were genotyped on the Illumina Human Omni2.5-Quad (Omni2.5) BeadChip at the NIH Center for Inherited Disease Research. We accessed survey results to evaluate prevalence of glaucoma in this dataset and performed a genome-wide association study (GWAS) adjusting for age, sex, and significant Principal Components (PC1-4). Of 8,245 genotyped individuals, 9.4% self-reported glaucoma. Respondents had a mean age of 76.55 (SD 7.50) and median age of 75.41 (IQR 70.75-81.50) and were 57.9% female (N=4775). A majority of respondents had genetic samples collected during the 2006 HRS year (N=4388, 53.22%; 2008, N=532, 6.45%; 2010, N=3325, 40.33%). Median time between genetic sample collection and survey assessment used was 2 years (IQR 2-6) with a majority assessed in 2012 (N=6369, 77.25%). A preliminary GWAS did not replicate known glaucoma loci and no variants attained genome-wide significance. Future analyses will subset respondents by race/ethnicity and evaluate genetic association with combinations of glaucoma and comorbidities.

3467F

Identifying hidden ancestries in publicly available summary genetic data.

A.E. Hendricks¹, K. Koach¹, M.M. Sorenson¹, T. Dinh¹, Y. Wu¹, C.R. Gignoux².

1) Mathematical and Statistical Sciences, University of Colorado Denver, Denver, CO; 2) Colorado Center for Personalized Medicine, University of Colorado, Anschutz Medical Campus, Aurora, CO.

Recent summary level genetic data have enabled advancements in understanding the relationship between genetics, health, and disease. Summary test statistics from GWAS are particularly useful for secondary analyses (e.g. genetic correlation) and data dissemination. Publicly available databases of allele and genotype frequencies, such as the genome Aggregation Database (gnomAD), are rich resources for looking up potentially causal variants in studies of rare diseases and as controls in rare variant tests given new methods. However, these databases rely on allele frequencies, which are known to vary across global and local populations. Ancestries provided in the summary data are often broad (e.g. European, African, Asian) and do not include information on finer scale ancestries (e.g. British, Italian) or admixture patterns (e.g. African Americans within African summary groups). Lack of precise ancestry information can result in confounding in secondary or association analyses, and can contribute to variant misclassification in molecular diagnosis of rare diseases. To address this, we have developed a modified EM algorithm that estimates the proportion of finer scale ancestry within publicly available data (e.g. GWAS, genotype databases). Unlike traditional individual level ancestry estimation methods, our method only requires allele or genotype frequencies. We are able to accurately and precisely detect the underlying ancestry proportions across a wide variety of simulations. Specifically, we have sensitivity to detect ancestries that make up only 1% of the summary data. We further apply our method to the BRAVO allele frequencies from the Trans-Omics for Precision Medicine (TOPMed) and to gnomAD African and Latino groups identifying hidden ancestries not provided in the summary data. With applications to many publicly available data types, including GWAS summary data, our method has the ability to provide precise ancestry information for the increasing number of genetic resources. Increased precision of ancestry information will help to reduce inaccurate results due to unknown or hidden ancestries.

3468W

KIBAG -- KIR genotype imputation with attribute bagging. S. Khor¹, X. Zheng², A. Ishitani³, F. Azuma⁴, C. Pyo⁵, Y. Omae¹, T. Yabe⁴, D. Geraghty⁵, K. Tokunaga¹, *HLA & KIR Imputation Network (HKimp.net)*. 1) Department of Human Genetics, The University of Tokyo, Tokyo, Japan; 2) Department of Biostatistics, University of Washington, Seattle, WA, USA; 3) Nara Medical University, Kashihara, Nara 634, Japan; 4) Japanese Red Cross Kanto-Koshinetsu Block Blood Center, Tokyo, Japan; 5) Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA.

Natural killer (NK) cells are essential in innate immunity and determine the outcome of antigen-specific responses. NK cell function based on the complex interactions mediated by membrane-bound and soluble gene products. NK cell activity is genetically modulated by differential expression of inhibitory and activating killer cell immunoglobulin-like receptors (KIR) that recognize HLA class I molecules. The high level of polymorphism of KIR genes together with their functional relation with HLA class I molecules suggest selection pressure driven by exposure to a wide variety of diseases. Indeed, long established KIR-HLA association has been revealed with HIV infection, hepatitis C infection, human papilloma virus (HPV) induced cervical cancer, psoriatic arthritis and type 1 diabetes. With the recent advancement of next generation sequencing (NGS) technology, KIR can now be characterized to high resolution including KIR copy number (CP), KIR A&B haplotypes and KIR allele. NGS based KIR typing remains a costly and laborious method to determine KIR types from a large cohort of individuals. Thus, we developed a new methodology named KIBAG (KIR Genotype Imputation with Attribute Bagging) which enable a rapid and inexpensive way to determine KIR typing from GWAS dataset. KIR typing of 726 Japanese individuals were performed using Scisco Genetics KIR typing kit and Miseq sequencing system. SNP information of these samples were determined using Affymetrix Axiom Japonica V1 and Japonica V2. Internal validation and external validation of KIR CP imputation were performed and an average accuracy of 97.9% and 95.7% were obtained respectively across 16 KIR genes. KIR with > 3 copies remains the most difficulties to impute due to the insufficiency of 3 copies carrying individual in the reference. Re-examination of GWAS with HLA class I associations or GWAS with SNP peaks in KIR region may reveal novel HLA-KIR associations. This work is part of the project of HKIMP.net (HLA & KIR Imputation Network) with the aim of providing population specific HLA & KIR imputation system to research community. KIR imputation system for a wide range of populations is currently undergoing.

3469T

A novel association test for joint analysis of multiple phenotypes using conditional CUR. Q. Sha, Z. Wang, X. Li, S. Zhang. Michigan Technological University, Houghton, MI.

Genome-wide association studies (GWAS) have successfully detected tens of thousands of SNP-trait associations. Earlier researches have primarily focused on association testing of genetic variants and some well-defined functions or specific phenotypic traits. Emerging evidence has shown that pleiotropy widely exists in complex human diseases. The joint analysis of multiple phenotypes within a cohort presents an attractive alternative to single phenotype analyses and may provide new insights into the etiology of human diseases. CUR matrix decomposition is a low-rank matrix approximation that uses only a small number of actual columns and/or actual rows from the original data matrix. Traditionally, CUR selects columns/rows that exhibit high statistical leverage on the best low-rank fit of the data matrix based on an "importance sampling" probability distribution. In this study, we demonstrate that the traditional CUR will not perform well in certain situations. We develop an innovative conditional CUR procedure that we refer to as cCUR to alleviate the intrinsic drawback of the traditional CUR in genetic association studies. We apply cCUR to jointly analyzing multiple correlated phenotypes. Extensive simulations as well as an analysis of COPDGene real dataset have been used to evaluate the performance of the proposed method. Our results demonstrate that the type I error rates of cCUR are well maintained under different scenarios and cCUR is either the most powerful test or comparable to the most powerful test among the existing methods we compared with. In addition, cCUR is able to detect more causal genetic variants in GWAS than the existing methods and is thus suitable for GWAS.

3470F

Leveraging allele-specific expression to refine fine-mapping for eQTL studies. J. Zou¹, F. Hormozdizadeh², B. Jew³, S. Castel^{4,5}, J. Ernst^{1,6}, J. Sul⁶, E. Eskini^{1,7}. 1) Computer Science Department, University of California Los Angeles, CA; 2) Genetic Epidemiology and Statistical Genetics Program, Broad Institute of MIT and Harvard, Cambridge, Massachusetts; 3) Bioinformatics Interdepartmental Program, University of California, Los Angeles; 4) New York Genome Center, NY, USA; 5) Department of Systems Biology, Columbia University, NY, USA; 6) Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles, California; 7) Department of Human Genetics, University of California, Los Angeles, California; 8) Department of Biological Chemistry, University of California, Los Angeles, California.

Many disease risk loci identified in genome-wide association studies are present in non-coding regions of the genome. These variants can affect complex traits by acting as expression quantitative trait loci (eQTLs) that influence expression of nearby genes. Therefore, identifying causal variants for gene expression is also important for elucidating the genetic basis of complex traits. However, detecting causal variants is challenging due to complex genetic correlation among variants known as linkage disequilibrium (LD) and due to the presence of multiple causal variants within a locus. Although several fine-mapping approaches have been developed to overcome these challenges, they may produce large sets of putative causal variants when true causal variants are in high LD with many non-causal variants. In eQTL studies, there is an additional source of information that can be used to improve fine-mapping called allele-specific expression (ASE) that measures imbalance in gene expression of two haplotypes in a diploid organism. It has also been shown that ASE is enriched in eQTLs, indicating that genetic variants which cause changes in total gene expression are also likely to cause ASE. In this work, we develop a novel statistical method that leverages both ASE and eQTL information to detect a minimal set of variants ("causal set") that contain the true causal variant with high probability. We illustrate through simulations and application to the Genotype-Tissue Expression (GTEx) dataset that our method identifies the true causal variants with higher specificity than an approach that uses only eQTL information. In the GTEx dataset, our method achieves a median reduction of 18% in the number of putative causal variants compared to the naïve approach which only uses eQTL information. Using orthogonal genome annotation, we also show that the SNPs in the causal sets computed are significantly enriched for functional elements compared to the SNPs excluded from the causal sets. While this method is not the first to combine ASE and eQTL data to identify cis-regulatory variants, it is unique in that it uses ASE and eQTL data to perform fine-mapping while taking LD and multiple causal variants into account. The majority of existing methods perform a test for each variant instead of modeling the entire locus. We show that using these statistics in our fine-mapping framework achieves a smaller median reduction rate compared to the statistics developed in our approach.

3471W

An integrated analysis tool for big genomic data from biobanks. H.C. Yang, M.C. Huang, C.W. Chen. Institute of Statistical Science, Academia Sinica, Taipei, Taiwan.

An integrated analysis of big genomic data from a biobank provides an unmet opportunity to study health-related issues but must face a challenge in batch effect. There is an urgent demand for analysis tools that can simultaneously adjust for batch effect and provide an unbiased analysis of allele frequency (AF), allelic imbalance (AI), loss of heterozygosity (LOH), long contiguous stretch of homozygosity (LCSH), and copy number variation or alteration (CNV/CNA). In this study, we extended our software ALICE (AF/LOH/LCSH/AI/CNV/CNA Enterprise) [1] by incorporating a new plate-based tiling array method to reduce batch effect. Local references for different types of chromosomal aberrations were generated based on local samples in the nearest neighbor plates to a sample to be examined, and then used to detect relative chromosomal aberrations. A parallel computing algorithm was implemented to increase computational efficiency. In addition, the results of our qPCR experiments for previous copy number validation were used to examine the performance of ALICE compared to PennCNV. We applied our batch effect reduction method to a large genomic dataset of 10,245 samples, genotyped with the Axiom TWB arrays, from the Taiwan Biobank Project. A dramatic reduction, 68.6% – 100.0% of the between-batch variation, was observed. The first draft of genomic maps of AF, AI, LOH/LCSH, and CNV/CNA in a Taiwan population and its three subpopulations (Minnan, Hakka, and Mainlanders) were established. The qPCR experiments showed ALICE outperformed PennCNV with regard to sensitivity in detecting copy number aberrations. In conclusion, ALICE with a user-friendly interface provides a powerful tool for an integrated analysis of big genomic data from biobanks. Reference: [1] Huang, M.-C., Chuang, T.-P., Chen, C.-H., Wu, J.-Y., Chen, Y.-T., Li, L.-H. and Yang, H.-C. (2016). An integrated analysis tool for hybridization intensities and genotypes using new-generation population-optimized human arrays. *BMC Genomics* 17, 266.

3472T

Hi-C deconvolution via joint modeling of bulk and single-cell Hi-C data.

R. Huh¹, Y. Yang², Z. Hou², J. Szatkiewicz², Y. Li^{1,2,3}. 1) University of North Carolina at Chapel Hill, Department of Biostatistics; 2) University of North Carolina at Chapel Hill, Department of Genetics; 3) University of North Carolina at Chapel Hill, Department of Computer Science.

Hi-C is a genome-wide ('all-against-all') variant of chromosome conformation capture technique, which measures spatial interactions important in providing information for gene regulation and 3D structure of the genome. Standard Hi-C data are generated from millions of cells, thus providing a population average measure of heterogeneous cells. When Hi-C samples are collected from different physiological states, observed differences in contact information are confounded by relative proportions of cell types among samples. Many deconvolution methods have been developed to address cellular heterogeneity in bulk RNA-seq data. However, to date, there are no deconvolution methods applied to Hi-C data. Here, we adopt a matrix decomposition based framework to estimate the proportions of each cell type in bulk Hi-C samples, which can be crucial in downstream analysis where, for example, differences in cellular composition across samples need to be adjusted. Recent developments in measuring single-cell Hi-C enable us to simulate bulk mixtures of Hi-C data to test the validity of our deconvolution method in estimating cell type proportions. In contrast to RNA-seq data, where gene expression levels are used to describe the cells, we leverage different levels of Hi-C measurements such as metrics for topologically associating domains, interchromosomal and intrachromosomal contacts, for profiling. We have performed extensive real data based simulations to demonstrate the power of our method. For one example, we mix HAP1 and HeLa (two different human cancer cell lines) single cells, at varying (10-90%) mixture proportions. For another example, we mix two mouse cell lines: Patski and MEF. Using interchromosomal contacts as our Hi-C metric, our method is able to closely replicate the simulated mixture weights. Absolute deviations of estimated proportions from the truth range from 0.001 to 0.09 for the human dataset and 0.024 to 0.092 for the mouse dataset. We aim to assess our deconvolution method using different levels of Hi-C metrics to provide the optimal solution for cell type proportions.

3473F

Why cross-study common controls are hard---and what can we do about it?

S. Buyske¹, C. DeBoever², A. Felsenfeld³, C.R. Gignoux⁴, X. Lin⁵, K.E. North⁶, M.A. Rivas⁷, T. Stephan³, *NHGRI Genome Sequencing Program*. 1) Statistics & Genetics Depts, Rutgers University, Piscataway, NJ; 2) Data4Cure, La Jolla, CA; 3) National Human Genome Research Institute, Bethesda, MD; 4) Colorado Center for Personalized Medicine, University of Colorado, Anschutz Medical Campus, Aurora, CO; 5) Dept of Biostatistics, T.H. Chan School of Public Health, Harvard University, Boston, MA; 6) Dept of Epidemiology & Carolina Center for Genome Sciences, University of North Carolina, Chapel Hill, NC; 7) Department of Biomedical Data Science, Stanford University, Stanford, CA.

Cross-study common controls for genetic association studies seem like a great idea---why not take advantage of the thousands of already sequenced subjects? However reality is more complicated than first appears. Issues related to common controls fall into two categories: subject selection and data. Within subject selection are many sub-issues including: How similar in genetic space are the cases and controls? Can standard tools (e.g., principal components) measure structure across the frequency spectrum? How about environmental design characteristics and secular trends? Studies also typically differ in sampling methods, which can have an impact on subject selection. For example, medical system biobanks tend to recruit patients who more frequently utilize healthcare systems compared to individuals in cohort studies. Even something as straightforward as sex ratio can affect study power. More fundamentally, what is a common control? Traditionally, controls are disease-specific, screened for the disease in question. On the other hand, common controls are much more likely to have to be considered as population-based controls, with unknown disease status (the effect on power depends on both disease and allele frequency). Furthermore, common controls may lack covariates generally used in the analysis of any particular trait, leaving researchers to consider the effect of disease prevalence in a putative pool of common controls. On the genotyping side, the use of common controls raises the specter of batch effects confounding the genetic-phenotypic association signal. Sequencing is not yet standardized to a point that such effects can be disregarded. How important are they and can statistical approaches reduce or eliminate them? For a researcher with study-specific cases and controls, as well as additional controls from a common controls resource, statistical methods fall into two approaches: 1) using just genotype calls, which allows one to consider if the study-specific controls differ from the common controls and weight accordingly, or 2) using read counts, which can be tested without calling genotypes. Many groups, including GSP and TOPMed, are making sequenced samples available for potential use as common controls, but their use requires careful consideration. Here, we provide recommendations based on initial discoveries from the GSP and other large data repositories to aid researchers in understanding the advantages and the caveats of leveraging common controls.

3474W

High dimensional stratified LD score regression. *R. Cui¹, T. Jones¹, F-E. Eid¹, Y. Reshef^{1,2}, A. Bloemendal¹, A. Price^{1,2}, H. Finucane¹.* 1) Broad Institute, Cambridge, MA; 2) Department of Epidemiology, Harvard T.H. Chan School of Public Health, Cambridge, MA; 3) Department of Computer Science, Harvard University, Cambridge, MA.

Stratified LD score regression (S-LDSC, Finucane et. al. 2015) models the effects of functional annotations on per-SNP heritability, enabling us to gain insight into genetic architecture and disease biology. However, S-LDSC has so far only been applied to model fewer than 100 functional annotations simultaneously due to poor performance of ordinary least squares (OLS) when large number of features are involved. Here, we extend stratified LD score regression to accommodate a large number of functional annotations. We use the lasso shrinkage method and modified stochastic gradient descent to fit the data. We show in simulations using real genotypes and real annotations a dramatic increase in accuracy for estimating categorical contributions to SNP heritability (i.e., regression coefficients) under sparse prior. In a model with 690 annotations including cell type specific annotations from Roadmap Epigenomics Consortium and the GTEx project as well as annotations from the baselineLD model, we ran 1000 simulations in which the regression coefficients are assumed to be sparse. To compare lasso performance with OLS, we took the median over all simulations of the mean squared error of the regression coefficients for both methods. lasso showed a near 5 order of magnitude increase in accuracy when estimating categorical contributions than the traditional OLS method. We applied our method to GWAS summary statistics of Schizophrenia, LDL, and Rheumatoid Arthritis, with the same 690 annotations. To evaluate our method without knowledge of the true regression coefficients, we computed a weighted mean squared error in predicted chi-square statistics on a left-out chromosome. We compared five different methods: ordinary least squares (OLS), lasso, lasso+OLS, Ridge, Elastic Net, and a "mean estimator" in which we predict every chi-square statistic to be the mean of the observed chi-square statistics. We see an improved prediction accuracy for summary statistics with regularization, with OLS performing similarly to the mean estimator, and Elastic Net and lasso performing the best.

3475T

A correction for sample overlap in genome-wide association studies in a polygenic pleiotropy-informed framework. *M. LeBlanc¹, V. Zuber², WK. Thompson³, OA. Andreassen⁴, THE. Schizophrenia/Bipolar⁵, A. Frigessi⁶, BK. Andreassen⁶.* 1) Oslo Centre for Biostatistics and Epidemiology, Oslo University Hospital, Oslo, Norway; 2) MRC Biostatistics Unit, University of Cambridge; 3) Department of Psychiatry, University of California, San Diego; 4) Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway; 5) Oslo Centre for Biostatistics and Epidemiology, University of Oslo and Oslo University Hospital, Oslo, Norway; 6) Department of Research, Cancer Registry of Norway, Oslo Norway; 7) Psychiatric Genomics Consortium.

Background There is considerable evidence that many complex traits have a partially shared genetic basis, termed pleiotropy. It is therefore useful to consider integrating genome-wide association study (GWAS) data across several traits, usually at the summary statistic level. A major practical challenge arises when these GWAS have overlapping subjects. This is particularly an issue when estimating pleiotropy using methods that condition the significance of one trait on the significance of a second, such as the covariate-modulated false discovery rate (cmfdr). **Results** We propose a method for correcting for sample overlap at the summary statistic level. We quantify the expected amount of spurious correlation between the summary statistics from two GWAS due to sample overlap, and use this estimated correlation in a simple linear correction that adjusts the joint distribution of test statistics from the two GWAS. The correction is appropriate for GWAS with case-control or quantitative outcomes. Our simulations and data example show that without correcting for sample overlap, the cmfdr is not properly controlled, leading to an excessive number of false discoveries and an excessive false discovery proportion. Our correction for sample overlap is effective in that it restores proper control of the false discovery rate, at very little loss in power. **Conclusions** With our proposed correction, it is possible to integrate GWAS summary statistics with overlapping samples in a statistical framework that is dependent on the joint distribution of the two GWAS.

3476F

Epigenetics-prioritized imputation of gene expression identifies more tissue-specific disease genes. W. Liu¹, M. Li², W. Zhang², G. Zhou¹, X. Wu³, H. Zhao^{1,2,3}. 1) Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT; 2) Department of Biostatistics, School of Public Health, Yale University, New Haven, CT; 3) Molecular Cell Biology, Genetics and Developmental Biology Program, Yale University, New Haven, CT.

With ever increasing data available from Genome-Wide Association Studies (GWAS) and transcriptome studies, many methods have been developed to identify disease-related genes by considering gene expression-mediated genetic effects on traits. However, most existing methods have not fully captured tissue-specific gene expression regulations, leading to loss of statistical power in identifications of disease-associated genes in relevant tissues. In this presentation, we propose a powerful and robust statistical method incorporating epigenetics signals to better prioritize biologically functional Single Nucleotide Polymorphism (SNPs) in gene expression level imputation and to better identify tissue-specific disease genes. Built on the assumption that SNPs with epigenetic signals are more likely to regulate the expression levels of their nearby genes, our method can more accurately predict gene expression levels by better identifying expression quantitative trait loci (eQTLs) and improve power to identify disease-related genes. To assess the performance of our method, we first evaluated the gene expression imputation accuracy via a 5-fold cross-validation in the whole blood tissue of the GTEx data. Compared with the commonly adopted elastic net model, our method achieved higher accuracies (mean value: 0.079 versus 0.019) across all genes ($p < 2.2 \times 10^{-16}$, with a sample size of $n=338$). Furthermore, when our method was applied to identify genes associated with schizophrenia (SCZ), type 2 diabetes, Alzheimer's disease and other complex traits, our method identified more (22.22% ~ 37.31%) disease-associated genes across all tissues than existing methods. More importantly, our method identified more genes (32.94% ~ 66.67%) in disease-associated tissues. For example, for SCZ, our method identified not only risk genes inferred in previous studies, but also novel genes, such as *PCDHA7*, which encodes a cell adhesion protein and plays an important role in establishing neuronal connections in brain, implying its potential role in schizophrenia. Besides, compared with PrediXcan, our method recalled more (138 versus 114) published risk genes. Overall, through integrating epigenetic information, our new method not only can better impute gene expression levels, but can also increase the statistical power to identify gene-trait associations both tissue-specifically and ubiquitously, providing novel insights in understanding complex traits mechanisms using multi-omics data.

3477W

Estimating genetic effect size from top-ranking genome-wide association statistics to remedy the winner's curse. D. Zaykin¹, O. Vsevolozhskaya². 1) National Institute of Environmental Health Sciences, NC; 2) University of Kentucky, KY.

Effect size estimates of genetic associations with diseases, such as odds ratios (ORs), tend to overstate their true values in whole genome studies due to selection of top-ranking loci. Various methods have been proposed for reducing this selection bias, known as the winner's curse phenomenon, but in the absence of *a priori* assumptions about the effect size distribution across loci in the genome, there is no remedy for the winner's curse: the bias cannot be eliminated, only reduced. To correct for selection bias during statistical analysis, it is therefore important to incorporate biological information about aspects of the effect size distribution, such as plausible frequency of truly associated variants in the genome and possible range of the effect size magnitude. The effect size distribution can be modeled naturally as a prior distribution in the Bayesian framework, but accurate specification of parameters for the prior distribution remains a challenge owing to sparsity of truly associated variants. Focusing on the $\log(\text{OR})$ as a measure of effect size, we derive a posterior distribution for $\log(\text{OR})$ utilizing only association statistics as summary of data. We model the prior as a mixture of two distributions, one for the bulk of non-associated loci, carrying effects of negligible magnitude, and the second one describing effect size distribution for associated loci, governed by the variance parameter to capture the spread of $\log(\text{OR})$. We develop a simple and powerful method for estimating the prior mixture distribution from the top ranking association statistics. In our approach, the fitted prior parameters are utilized to obtain a bias-corrected posterior estimator of $\log(\text{OR})$. We demonstrate efficiency of the proposed estimator, as well as its resistance to the winner's curse.

3478T

VikNGS C++ Tool for sequence-based association testing enabling consortium research: Updates and applications. Z. Baskurt¹, S. Mastro-matteo^{1,3}, J. Gong¹, S.W. Scherer^{3,4}, L.J. Strug^{1,2,3}. 1) Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, ON, Canada, Toronto, ON, Canada; 2) Biostatistics Division, Dalla Lana School of Public Health, University of Toronto, ON, Canada; 3) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, ON, Canada; 4) McLaughlin Centre and Department of Molecular Genetics, University of Toronto, ON, Canada.

The Variant Integrating Kit for NGS (VikNGS) is a C++ package designed for common and rare variant association analysis using sequence data generated across different study cohorts. Integrating sequencing data from different sources to increase sample size can improve power but can also introduce spurious association findings due to differences in uncertainty in genotype calls and unbalanced case-control sample sizes. A robust variance score statistic (RVS) that uses expected genotypes has been shown to control Type I error in case-control association studies when the cases and controls were sequenced using different read depths (rd). We generalized the RVS method to include multiple sequenced cohorts for quantitative trait analysis with covariate adjustment, as well as for the integration of an arbitrary number of sequenced cohorts. VikNGS enables power analysis to estimate sample size for study planning through a simple user interface, handles rare variant testing with linear or quadratic RVS classes of tests and any user-defined weights, e.g. SKAT, and provides a regular score test with genotype calls. To demonstrate the utility of VikNGS we consider a highly unbalanced case-control association study integrating three independently sequenced samples: 1) 101 samples with cystic fibrosis (CF) sequenced at 30x by Complete Genomics Inc, 2) 1927 samples without CF from the UK10K Consortium sequenced at an average rd of 6.5x, and 3) 379 European samples without CF from the 1000 Genomes Project Phase 1 sequenced at an average rd of 4x. We generated a multi-sample VCF file for chr 22 (no expected association with CF) and conducted association tests using 1) the generalized 3-group RVS method, 2) a method that uses expected genotypes but does not implement the RVS method and 3) the conventional score test using genotype calls. The generalized RVS approach is robust to unbalanced case-control designs and differential read-depth across the multiple cohorts. The other two tests produced many spurious associations. Quantitative trait analysis of lung function using the UK10K Consortium data and the generalized RVS demonstrate association at a CF lung disease modifier locus at chr 1, suggesting the CF modifier locus is relevant in the general population. VikNGS is a computationally efficient package that provides functionality for association analysis of WGS from multiple study groups through a user-friendly graphical interface.

3479F

GASPOR: Detecting gene-level allele-specific expression in a population by RNA sequencing. J. Fan¹, J. Hu¹, C. Xue², M. Reilly², R. Xiao¹, M. Li¹. 1) Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, Philadelphia, PA; 2) Cardiology Division, Department of Medicine, Columbia University Medical Center, New York, NY.

Allele-specific expression (ASE) analysis, which quantifies the relative expression of two alleles in a diploid individual, is a powerful tool for identifying *cis*-regulated gene expression variations that underlie phenotypic differences among individuals. To understand the role of ASE in human diseases, we exploit allelic imbalance by RNA-seq, which provides allele-specific read counts at exonic SNPs distinguished by heterozygous sites. Existing methods for ASE detection report evidence of ASE in single individuals, in which the ASE is quantified for each SNP and a gene-level ASE is obtained by integrating effects across SNPs in the same gene for an individual. However, evidence of ASE is often shared across individuals. Failure to accommodate such shared information not only loses power, but also makes it difficult to interpret results across individuals. It is desirable to have a method that simultaneously models both multi-SNP and multi-individual information. However, ASE detection across individuals is challenging because the data often include some individuals that are heterozygous for the unobserved *cis*-regulatory SNP and others that are homozygous for the *cis*-regulatory SNP, leading to heterogeneity in ASE as only those heterozygous for the *cis*-regulatory SNP will show evidence of ASE, whereas those homozygous individuals will have balanced expression. To simultaneously model gene-level ASE across individuals and to account for such heterogeneity, we develop GASPOR, a mixture model of two binomial distributions with subject-specific random effect to account for correlation of multiple SNPs within the same gene. GASPOR is able to detect gene-level ASE under one condition and differential ASE between two conditions (e.g., pre- vs post-treatment). Extensive simulations show that GASPOR performed consistently well under a variety of scenarios, and was substantially more powerful than individual-based ASE detection methods. We further applied GASPOR to RNA-seq data in the Genetics of Evoked Response to Niacin and Endotoxemia Study, and identified genes showing strong evidence of differential ASE pre- vs post-LPS stimulation, which were missed by existing methods. To the best of our knowledge, GASPOR is the first method for gene-level ASE detection in a population. With the growing popularity of RNA-seq, we believe GASPOR will be well-suited for various ASE studies for human diseases. An R package of GASPOR will be available.

3480W**Fast and robust ancestry inference using principal component analysis.**

D. Zhang, R. Dey, S. Lee. Department of Biostatistics, University of Michigan, Ann Arbor, MI.

Population stratification (PS) is a major confounder in genome-wide association studies (GWAS) and can lead to false positive associations. To adjust for PS, principal component analysis (PCA)-based ancestry prediction has been widely used. This approach first performs PCA on the reference samples with detailed ancestry information and then uses the reference sample PCA results to predict the PC scores of the study samples. Simple Projection (SP) and Augmentation-Decomposition-Transformation (ADP), such as LASER and TRACE, are popular methods for predicting PC scores. However, they are either biased or computationally expensive. For example, the predicted PC scores from SP can be biased toward NULL. On the other hand, since ADP requires running PCA separately for each study sample on the augmented dataset, its computational cost is high. To address these problems, we propose two alternative approaches, Adjusted Projection (AP) and Online ADP (OADP), and have implemented them in our genoPCA software. Using random matrix theory results, AP asymptotically estimates and adjusts for the bias of SP. OADP uses online singular value decomposition, which can greatly reduce the computation cost of ADP. We carried out extensive simulation studies to show that these alternative approaches are unbiased and that the computation times are 10-100 times faster than ADP when 2000 reference samples are used. We also applied our approaches to 500,000 study samples from the UK biobank data with 2,500 samples from the 1000 Genomes data as the reference. AP and OADP required 16 and 92 CPU hours, respectively, while the projected computation time of ADP is 1,933 CPU hours. Furthermore, when we only used the European reference samples in the 1000 Genomes to infer sub-European ancestry, SP clearly showed bias, unlike the proposed approaches. By using AP and OADP, we can infer ancestry and adjust for PS in GWAS robustly and efficiently.

3481T**Kernel association test for rare copy number variants using CNV intensity profiles.** *A. Brucker¹, W. Lu¹, R.M. West¹, J. Szatkiewicz², J.-Y. Tzeng¹.* 1) NC State University, Raleigh, NC; 2) UNC Chapel Hill, Chapel Hill, NC.

Copy number variants (CNVs) are the gain or loss of DNA segments in the genome that can vary in dosage and length. Rare CNVs have been shown to be associated with phenotypes including cancers and psychiatric disorders. Recent studies have shown that kernel-based tests are a useful tool for assessing the aggregate effect of rare CNVs on phenotypes, due to their ability to capture between- and within-locus etiological heterogeneity. To perform a kernel association test, a CNV locus needs to be defined so that similar variants in a single locus correspond to similarity in profiles. Then collapsing methods are applied to summarize similarity across all loci and a kernel is constructed. There is nuance in the choice of locus with which the "locus"-specific effects are retained during aggregation, and current work has suggested that different locus definitions can lead to different relative performance depending on the underlying effect patterns. In this work, we propose a new kernel-based test that does not require a definition of locus but makes use of the alignment of CNV profiles performed in standard CNV data processing. This test has power to detect associations driven by CNV dosage, length, and dosage-length interactions. In a variety of simulation settings the proposed test shows comparable and improved power over existing approaches. A real data analysis will examine the performance of the test applied to known gene pathways for data from the Psychiatric Genomics Consortium.

3482F

Enrichment of survey responses by genetic community within a large genetic genealogy database. E.P. Sorokin¹, K.A. Rand¹, J.M. Granka¹, A. Girshick¹, L. Nguyen¹, J. Byrnes¹, K. Noto¹, E.L. Hong¹, K.G. Chahine², C.A. Ball¹. 1) Ancestry.com DNA, LLC, San Francisco, CA; 2) Ancestry.com DNA, LLC, Lehi, UT.

The increased scale of human genetic cohorts has empowered the elucidation of cryptic population structure, often through the detection of regions of identity-by-descent. Our analysis, coupled with community detection algorithms across a network of AncestryDNA customers who consented to research, has yielded a portrait of the fine-scale population structure within post-colonial America, including genetic communities of European, African, East Asian, South Asian, Middle Eastern and Oceanian ancestry. In addition to rich metadata stemming from historical records and other genealogical sources, we sought to determine whether additional sources of data could elucidate cultural or other factors underlying the formation or maintenance of this population structure. To further understand patterns within and across genetic communities, we collected survey responses from a subset of AncestryDNA customers. Here, we examine the distributions of these survey responses by genetic communities and characterize their community-specific prevalence. We also perform association analyses with certain traits and directly genotyped SNPs. We show that some traits have distinct patterns across fine-scale population structure and we also contribute to the replication and identification of loci associated with certain traits. .

3483W

Imputed gene associations identify replicable *trans*-acting/target gene pairs and suggest a larger *trans* than *cis* contribution to complex trait architecture. H.E. Wheeler^{2,3}, S. Ploch¹, A.N. Barbeira⁴, H.K. Im⁴. 1) Dept of Biology, Loyola University Chicago, Chicago, IL; 2) Dept of Computer Science, Loyola University Chicago, Chicago, IL; 3) Dept of Public Health Sciences, Loyola University Chicago, Maywood, IL; 4) Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, IL.

Genetic variation contributes to complex trait variation through regulation of gene expression. Much progress has been made in uncovering local *cis*-acting expression quantitative trait loci (eQTLs), but distal *trans*-acting eQTLs have been more difficult to identify and replicate. Here, we apply PrediXcan to map *trans*-acting genes, rather than SNPs. We compared results between the discovery Framingham Heart Study (FHS, n=4838) and the replication Depression Genes and Networks (DGN, n=922) whole blood cohorts. We used *cis*-acting models built in the Genotype Tissue Expression (GTEx) Project whole blood cohort to predict gene expression in FHS and DGN. Examining the correlations between predicted and observed gene expression of gene pairs on different chromosomes, 124 pairs were significantly correlated in FHS, with an expected true positive rate (π_1) of 0.70 in DGN. To determine if combining predicted expression across tissue models could improve our ability to detect *trans*-acting/target gene pairs, we used the multivariable regression approach called MultiXcan, which accounts for correlation among predicted expression levels across 44 GTEx tissues. When we applied MultiXcan to the FHS data, we discovered 3657 *trans*-acting/target gene pairs, a dramatic increase compared to the single tissue whole blood model. Although the expected true positive rate of the multi-tissue model was lower ($\pi_1=0.43$) than the single tissue model, the absolute number of replicate gene pairs was much higher. Thus, our *trans*-MultiXcan model increases power to detect *trans*-acting gene pairs >10-fold, indicating shared *cis*-eQTL allow us to leverage multiple tissues to detect more *trans*-acting effects. Pathway analysis of replicated MultiXcan gene pairs revealed the *trans*-acting genes are enriched in transcription and nucleic acid binding pathways and target genes are enriched in transcription factor binding sites and highly conserved motifs, indicating that our method identifies genes of expected function. We found that MultiXcan *trans*-acting and target genes are more heritable than other expressed (background) genes. Using height as a representative complex trait, *trans*-acting gene associations with height are more significant than target and background gene associations in both UK Biobank (n=500,131) and the GIANT Consortium (n=253,288), consistent with a larger contribution of *trans*-acting effects than *cis*-acting effects to complex traits.

3484T

Discovery of novel hepatic eQTLs in African Americans: Disparities in precision medicine. Y. Zhong¹, Y. Xu², T. De¹, A. Alarcon¹, CS. Park¹, BM. Lec¹, MA. Perera¹. 1) Northwestern University, Chicago, IL; 2) University of Chicago, Chicago, IL.

Genetic regulation of gene expression through eQTLs has been studied in a variety of treatment conditions and tissue types. However, the catalog of eQTLs needs to be expanded to be inclusive of other minority populations. Importantly, population-specific eQTLs promise to provide mechanistic insight of transcriptome regulation within and across populations. Liver is a critical organ for the understanding of drug metabolism, as well as diseases such as coagulopathies and metabolic disorders. Both of these broad disorders show disparities in morbidity, mortality and response to drug therapy between African Americans (AAs) and non-African populations. In the GTEx project, only 15 individuals of African ancestry comprised the liver eQTL mapping of 449 individuals, making extrapolation of this more diverse population under-powered and not representative of the larger AA population. Here, we performed the first eQTL mapping in AA hepatocytes and identified 28,855 significant gene-SNP pairs corresponding to 387 eGenes and overlapped identified eQTLs with GTEx liver eQTLs (v.7) results. While there was a significant correlation of effect sizes of common eQTLs, we observed a large number of eQTLs (9,655) and eGenes (145) were unique to AA hepatocytes. We found larger allele frequency differences between African and European populations for AA-specific eQTLs compared to common eQTLs, suggesting differences in allele frequency underlie population-specific eQTLs. Given the scarcity of exon QTLs data available, we identified 3,190 exon QTLs corresponding to 150 eExons. Surprisingly, 101 eExons do not belong to identified eGenes, suggesting mapping exon QTLs may reveal additional regulatory events related to differential transcript expression. We next characterized the functional properties of AA hepatic eQTLs. AA eQTLs cluster primarily near the transcriptional start site as compared to other cis regions. Using CaVEMaN, we identified 5% of lead SNPs with larger than 0.6 probability of being a causal variant. Finally, we built a PrediXcan model that predicted hepatic transcriptome (852 reliable genes) with AA genotype and found a moderate correlation (spearman $\rho = 0.13$) of prediction accuracy when compared with GTEx European liver model. Together, this is the first comprehensive study that characterizes the genetic regulation of gene expression in AA hepatocytes, revealing population-specific regulation of eQTLs with implications for disease genomics and pharmacogenomics.

3485F

Robust and powerful analysis of gene-environment interactions using polygenic risk scores in case-control studies. N. Chatterjee^{1,2}, A. Meisner¹.

1) Department of Biostatistics, Johns Hopkins School of Public Health, Baltimore, MD; 2) Department of Oncology, Johns Hopkins School of Medicine, Baltimore, MD.

Background Assessment of gene-environment interactions is critical for understanding disease etiology, developing models for risk prediction, and evaluating the impact of lifestyle interventions by genetic background. Despite the availability of powerful methods and large datasets, investigations of gene-environment interactions have led to very limited findings to date, possibly due to weak effects of individual SNPs. Use of polygenic risk scores (PRS), overall or possibly grouped by pathways, can be a powerful strategy for detecting global patterns of interactions across many genetic variants. **Methods** Motivated by the power of the case-only design for the efficient analysis of single SNP gene-environment interactions, we consider retrospective approaches to the analysis of PRS in case-control studies. We show that under the assumptions of normality of the PRS and independence between the PRS and the environmental risk factor (E), relative risk interaction parameters can be estimated by performing a linear regression analysis of PRS on E in a sample of cases. Furthermore, if genotype data are available on a sample of subjects representative of the population, information on the mean of the PRS can be used to estimate the main effect of the PRS in the underlying disease risk model. Further extensions are considered to account for population stratification and gene-environment dependence, the latter of which could be modeled using a PRS associated with E. **Results** Simulation studies indicate the efficiency of the case-only method is similar to that of a full cohort study with a comparable number of cases and is about twice that of a standard logistic regression analysis of a case-control study with an equal number of cases and controls. We observe similar efficiency of the method in the analysis of gene-environment interactions and breast cancer risk in the UK Biobank cohort study and simulated case-control studies within this cohort. Results from additional simulation studies and analyses investigating the effects of population stratification and gene-environment dependence will be presented. **Conclusion** Case-only and other retrospective methods have the potential to greatly enhance the power of gene-environment interaction analysis based on PRS in case-control studies. However, careful consideration is needed to avoid bias due to population stratification and gene-environment associations.

3486W

Reconstruction of patterns of cryptic relatedness and identification of the resulting shared genomic segments in a large electronic health record-linked DNA biobank. *H-H. Chen, L.E. Petty, J.E. Below.* The Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN.

In regionally ascertained samples it is common to enroll individuals that are cryptically related to one another; when genetic data are available for the sample, these cryptic relationships can be discovered and utilized in epidemiologic and genetic analyses. Specifically, the number and length of shared genomic segments can be used to estimate the relatedness between individuals, reconstruct pedigrees, trace population history, and even map disease genes. The Vanderbilt DNA biobank (BioVU) has collected DNA samples from over 280,000 participants with linked electronic health records from Vanderbilt University Medical Center. Characterization of full pedigree reconstruction, distant relatedness, and shared genomic segment detection in all 280,000 participants is on-going, however, in preliminary analyses, we estimated relatedness, reconstructed pedigrees, and identified shared segments in a subset of 6,551 individuals from BioVU genotyped on the Illumina Multi-Ethnic Global Array (MEGA), using 815,371 variants. Shapeit2 was applied to estimate each individual's haplotype, and then GERMLINE was performed to identify the shared segments between all pairs of participants who shared segments identical-by-descent (IBD). We then reconstructed pedigrees using PRIMUS and determined pair-wise relatedness using ERSa and PADRE. In our preliminary investigation, the average shared segment length is 4.538 cM, and among individuals who shared at least one segment, the mean number of shared segments is 2.054. The maximum number of shared segments observed was 62. We identified 42 first-degree relatives (parent-offspring or siblings), 28 second-degree relatives, and 83 third-degree relatives. 941 pairs were related at 4th or 5th degree, 1,284 pairs were related at 6th or 7th, and over 48% of pairs were estimated to be related at 8th or 9th degree. These results highlight the complex and significant patterns of unreported relatedness captured in regionally ascertained biobanks, a finding that has important consequences for future epidemiologic and genetic analyses, as well as providing an untapped avenue for IBD-based disease-gene discovery.

3487T

Exome-chip association study of refractive error in U.S. Caucasians. *D.D. Lewis¹, J. Ishika¹, T. Alexander^{1,2}, A.M. Musolf¹, D. Stambolian³, J.E. Bailey-Wilson¹.* 1) Computational and Statistical Genomics Branch, National Human Genome Research Institute, Baltimore, MD; 2) National Institute of Arthritis and Musculoskeletal and Skin Diseases, Bethesda, MD; 3) Department of Ophthalmology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

Refractive error occurs when light fails to properly focus on the retina of the eye. In the case of myopia, the light is focused in front of the retina, causing blurry distance vision. Myopia is defined in terms of refractive error (RE), a quantitative trait measured in Diopters (D), such that individuals with RE of -0.5D or more negative generally require glasses to correct their vision. Individuals with high myopia (RE < -6D) are vulnerable to ocular complications later in life. Consequently, great efforts have been undertaken to identify and understand the mechanisms underlying the development and progression of myopia. Some environmental exposures contribute to risk and genome-wide association studies (GWAS) and linkage studies have identified loci underlying the distribution of refractive error and influencing the risk of myopia. However, few causal variants have been definitely identified and a large proportion of the variation in refractive error remains unexplained. Therefore, this study aims to determine whether rare variants may account for some of the variation of refractive error. Here we present results of a single-variant association test for a quantitative trait, refractive error, in 1562 control individuals from the Age-Related Eye Disease Study (AREDS). After quality control, close to 100,000 variants were analyzed and EMMAX single-marker analysis identified 18 rare variants that were genome wide significant ($p < 5 \times 10^{-8}$). Interesting candidate variants were found in genes *NIPSNAP2*, *TSSK3* and *IDH1*. One study identified a role for *IDH1* in bovine eyes, where the enzyme protein fulfills the criteria for a corneal epithelial crystallin, which may be involved in maintaining corneal epithelial transparency. We are extending the analysis to gene-based tests and will present these results.

3488F**Robust region-based test for unbalanced case-control phenotypes.** Z.Zhao^{1,2}, L. Fritsche^{1,2}, M. Zawistowski^{1,2}, C. Brummett³, G. Abecasis^{1,2}, S. Lee^{1,2}.

1) Biostatistics, University of Michigan, Ann Arbor, MI; 2) Center for Statistical Genetics, University of Michigan School of Public Health, Ann Arbor, MI; 3) Department of Anesthesiology, University of Michigan Medical School, Ann Arbor, MI.

In Biobank data analysis, most binary phenotypes from Electronic Health Records (EHR) have unbalanced and often extremely unbalanced case-control ratios (1:10 or 1:100, respectively), which can cause inflation of type I error rates. Recently, a saddlepoint approximation (SPA) based single variant test has been developed to provide an accurate and scalable method to test for associations of such phenotypes. For gene or region-based multiple variant tests, a few methods exist which adjust for unbalanced case-control ratios; however, these methods are either less accurate when case-control ratios are extremely unbalanced or not scalable for large data analyses. Here we propose SKAT/SKAT-O type region-based tests, where the single-variant score statistic is calibrated based on SPA and Efficient Resampling (ER). Our method also utilizes a small number of resampling to estimate higher moments of overall test statistics, which can further calibrate association p-values. Through simulation studies, we show that the proposed method can both adjust for covariates and provide well-calibrated p-values. In contrast, the unadjusted approach has greatly inflated type I error rates (200 times of exome-wide $\alpha=2.5 \times 10^{-6}$) when case: control ratio was 1:99. In the application to the Michigan Genomics Initiative (MGI), we show that the proposed method can control type I error rates even for traits with a very small number of cases and a large number of controls (e.g. Breast cancer: 1160 cases and 15473 controls; Gout: 449 cases and 17644 controls; and Primary hypercoagulable state: 168 cases and 16401 controls).

3489W**Haplotype-based GRMs complement SNP-based GRMs in a regional heritability analysis of complex traits.** R.F. Oppong¹, P. Navarro², C.S. Haley^{2,3}, S. Knott¹.

1) Institute of Evolutionary Biology, The University of Edinburgh, Edinburgh, United Kingdom; 2) MRC IGMM, The University of Edinburgh, Edinburgh, United Kingdom; 3) The Roslin Institute and Royal Dick School of Veterinary Studies, The University of Edinburgh, Edinburgh, United Kingdom.

We have proposed and implemented a genome-wide analytical method that analyses blocks of genomic regions using a regional GREML model. The novelty in our method is that genomic regions in our data are defined naturally by recombination hotspots drawn from a reference human genome and we use a regional GREML model that fits a haplotype-based GRM. We compared this model to a regional GREML model that fits a SNP-based GRM. We hypothesise that the haplotype-based model will complement conventional SNP-based GREML methods by capturing regions in the genome contributing to the phenotypic variation that the SNP-based methods fail to capture. This can arise where there are causative variants associated with individual haplotypes but not in linkage disequilibrium with individual SNPs. This scenario is particularly likely for relatively newly-arisen and rarer variants. We tested this hypothesis in a simulation study involving 20,000 individuals in which we simulated 20 replicates each of two types of phenotypes with SNPs and three types of phenotypes with haplotypes as the effect variants. The results show that the two models are very specific to the type of marker effect they can capture; – the haplotype-based model can specifically target haplotype effects which are mostly missed by SNP-based analysis and vice versa. These results, therefore, support the hypothesis that haplotype-based GREML models will complement SNP-based GREML models. In conclusion, the use of haplotype-based GREML regional heritability models in the analysis of real data is of interest as these models can identify causal effects when the variance is explained by haplotype-associated variants that cannot be detected using other analytical models.

3490T

Kids First Data Resource Center: Advancing genetic analyses of structural birth defects and childhood cancer. D.M. Taylor^{1,2}, S. Volchenboum^{3,4}, A. Heath¹, Y. Zhu¹, L. Graglia³, B. Fumer³, J. Lilly¹, P. Raman¹, B. Farrow¹, T. Patton¹, A. Resnick¹, Gabriella Miller Kids First Data Resource Center. 1) The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Department of Pediatrics, Perelman School of Medicine, The University of Pennsylvania, Philadelphia PA; 3) Center for Research Informatics, The University of Chicago, Chicago, Illinois; 4) Department of Pediatrics, Biological Sciences Division, The University of Chicago, Chicago, Illinois.

Structural birth defects and childhood cancers share a common context of altered developmental biology. Numerous studies have provided clear evidence that children with certain structural birth defects or developmental disorders have an increased risk of developing cancer during childhood. Moreover, it is also known that various structural and developmental defects often co-occur independent of known syndromes, providing complex phenotypes of developmental defects that appear to share a common genetic basis. A better understanding of genetic variation within and across a disparate set of developmental and cancer phenotypes, genes, and pathways could spur advancements in prevention, detection, and therapeutics to improve outcomes for affected children and families. The NIH Common Fund's Gabriella Miller Kids First Pediatric Research Program (GMKF) is a collaborative initiative focused on providing large-scale clinically-annotated genomic data for pediatric cancer and structural birth defect cohorts, including trio germline whole genome sequencing (WGS) and tumor WGS and RNA-seq. The GMKF Kids First Data Resource Center (DRC, <https://kidsfirstdrc.org/>) is charged with empowering collaborative research and discovery through integration of GMKF cohorts and external data. The Kids First Data Resource Portal (DRP) is a first-of-its-kind international repository for GMKF cohorts and data discovery environment, which will facilitate data release to the community through the dbGAP application and data use agreement process. As of 2017, there are 23 GMKF cohorts funded for sequencing (15 structural birth defects and 8 pediatric cancers). Approximately 8,000 WGS patient samples and phenotypes will be available at the launch of the Kids First DRC portal in June 2018, and 25,000 WGS will be available by early 2019, making the Kids First DRP one of the largest pediatric data resources of its kind. We present preliminary summarization analyses, including of phenotype and genotype data across available cohorts. Germline genomic summaries include *de novo* events, rare variant frequencies, affected genes and any relationships to cancer mutation data. Phenotype summaries include information content and cross-cohort relationships through ontological analysis. We show how the Kids First DRP can help investigators at all levels access and analyze the Kids First cohort data.

3491F

Refining fine-mapping: Effect sizes and regional heritability. C. Benner^{1,2}, A. Havulinna^{1,3}, V. Salomaa², S. Ripatti^{1,2,4}, M. Pirinen^{1,2,5}. 1) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; 2) Department of Public Health, University of Helsinki, Helsinki, Finland; 3) National Institute for Health and Welfare, Helsinki, Finland; 4) Broad Institute of MIT and Harvard, Cambridge, USA; 5) Helsinki Institute for Information Technology HIIT and Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland.

Recent statistical approaches have shown that the set of all available genetic variants explains considerably more phenotypic variance of complex traits and diseases than the individual variants that are robustly associated with these phenotypes. However, an ultimate goal in genetics research is to narrow down the polygenic regional heritability into individual causal variants contributing to heritability. Because rapidly increasing sample sizes constantly improve fine-mapping of causal variants, it is useful to routinely estimate how much phenotypic variance the detected causal variants explain for each region. We have therefore extended the FINEMAP software to estimate the effect sizes and regional heritability under the probabilistic model that assumes a handful of causal variants per each region. Using the UK Biobank data to simulate genomic regions with only a few causal variants, we demonstrate that FINEMAP provides higher precision and enables more detailed decomposition of regional heritability into causal variants than the variance component model implemented in BOLT or the fixed-effect model implemented in HESS. Using data from 51 serum biomarkers and four lipid traits from the FINRISK study, with a broad spectrum of genome-wide heritability levels, we estimate that FINEMAP captures on average 24% more regional heritability than the variant with the lowest P value alone and 20% less than BOLT. As an example where both FINEMAP and BOLT give a similar estimate of about 70% regional heritability, we study in detail the association between the *PON1* gene and Para-oxonase 1 levels. Our simulations suggest how an upward bias of BOLT and a downward bias of FINEMAP could together explain the observed difference between the methods. We conclude that FINEMAP provides a computationally efficient framework to deduce a variant-level picture of the regional genetic architecture and expect that these new features prove useful as the accuracy of fine-mapping constantly increases with rapidly growing sample size in the era of biobank-scale data.

3492W

A regression-based framework for large-scale integration of pathway knowledge in GWAS. S. Bhattacharjee, S. Biswas, S. Pal. National Institute of Biomedical Genomics, Kalyani, West Bengal, India.

Traditional unbiased Genome-wide Association Studies (GWAS) have successfully identified thousands of genetic polymorphisms (SNPs) to be associated with various complex traits. However, GWAS are under-powered to detect all associated variants due to a huge multiple-testing burden; particularly true for variants with lower frequency or weaker effects. There is recent interest in conducting prioritized analysis of GWAS using SNP-level functional annotations and some rigorous statistical methods have been developed to enable incorporation of SNP-level knowledge. Another, very rich source of biological knowledge exists in the form of gene-level functional annotations (e.g. pathways). Pathway enrichment analysis, occasionally used after a GWAS, can help in the interpretation of findings; but does not by itself improve the power of detecting additional novel loci. It is imperative to use knowledge of pathways/networks a-priori to 'inform' the genome-wide search. Here, we develop a statistical method to enable a prioritized genome-wide scan using knowledge from a very large number of gene-level annotations (e.g. pathways from a pathway database). Our method is computationally efficient and scalable for routine use with large number of annotations. It takes as input p-values of SNPs from a GWAS and re-weights them optimally with pathway information utilizing a flexible framework based on regularized logistic regression. We demonstrate using simulations and real GWAS data on Type-2 Diabetes, Coronary Artery Disease etc., that such a 'pathway-guided GWAS' can indeed gain substantial power over unbiased GWAS, specifically when the associated SNPs map to genes that cluster into a few biological pathways. Unlike most prioritization tools using FDR, our method maintains global false-positive rates at the same stringency as in a usual GWAS (i.e. $p < 5 \times 10^{-8}$). We have created an R package 'GKnowMTest' (Genomic Knowledge-guided Multiple Testing) for flexible incorporation of gene-level knowledge into GWAS. Further, we have extended our method to allow large number of SNP-level functional annotations retaining the computationally simple regression-based framework. In conjunction with knowledge from pathways, we are using SNP-level knowledge including histone marks, transcription factor binding and regulation of gene expression in tissues to prioritize GWAS and assess the gains in power of detecting associated loci.

3493T

Revisiting the genome-wide significance threshold. Z. Chen, M. Boehnke, B. Mukherjee. Department of Biostatistics, University of Michigan, Ann Arbor, MI.

Genome-wide association studies (GWAS) have successfully yielded reproducible results by using a P-value threshold of 5×10^{-8} to classify association results as genome-wide significant. However, this threshold was proposed a decade ago, before the advent of newer experimental and imputation strategies that now allows us to study many more common and low-frequency variants. In this study, we revisit the P-value threshold of 5×10^{-8} with empirical and simulation analysis and compare that decision rule to two other metrics for controlling multiple hypotheses testing: the Benjamini-Hochberg-Yekutieli (BHY) procedure and Bayesian false-discovery probability (BFDP). We applied various threshold and decision rules to a series of GWAS with sequentially growing datasets from the Global Lipids Genetics Consortium (GLGC) and evaluated the performance of these metrics in detecting noteworthy signals from the most recent GWAS results in 2013 assuming the latest GWAS represents "true signals". In addition, we re-estimate a genome-wide significance threshold using a permutation procedure suggested by Dudbridge and Gusnanto (2008) but with a simulated dataset at the size and scale of more recent GWAS with much higher marker density. For the fixed P-value metric, compared to a threshold of 5×10^{-8} a less stringent threshold of 5×10^{-7} gains on average 6 additional true discoveries and controls the empirical false-discovery rate (FDR) between 6 and 17%. BHY and BFDP, which seeks to directly control the FDR, did not produce consistently better empirical FDRs than the fixed P-value metric and are more complex to use. Applying these methods to testing multiple traits simultaneously (e.g. three lipid traits together) lowered the empirical FDR by an average of 5.6% when compared to the combined univariate tests. To estimate a genome-wide significance threshold, we will permute phenotypes for 6,000 samples and ~9.5 million variants from the most recent UK Biobank study on body mass index and repeat the procedure 1,000 times. We then calculate the effective number of independent tests at multiple marker densities using the minimum P-values and extrapolate to full saturation of the whole-genome by fitting a Monod function.

3494F

Efficient and accurate estimation of genotype odds ratios in biobank-based unbalanced case-control studies. *R. Dey, S. Lee.* Biostatistics, University of Michigan, Ann Arbor, MI.

In genome-wide association studies (GWASs), genotype odds ratios (ORs) quantify the effects of the variants on the binary phenotypes, and calculating the genotype ORs for all of the markers is required for several downstream analyses. Calculating genotype ORs in a genome-wide scale is computationally challenging, especially when analysing large-scale biobank data, which involves performing thousands of GWASs phenome-wide. Since most of the binary phenotypes in biobank-based studies have unbalanced (case:control = 1:10) or often extremely unbalanced (case:control = 1:500) case-control ratios, the existing methods cannot provide a scalable and accurate way to estimate the genotype ORs. The traditional logistic regression provides biased OR estimates in such situations. Although the Firth bias correction method can provide unbiased OR estimates, it is not scalable for genome-wide or phenome-wide scale association analyses typically used in biobank-based studies, especially when the number of non-genetic covariates is large. On the other hand, the saddlepoint approximation-based test (fastSPA), which can provide accurate p values and is scalable to analyse large-scale biobank data, does not provide the genotype OR estimates as it is a score-based test. Here, we propose a scalable method using the fastSPA approach, to accurately estimate the genotype ORs, adjusting for non-genetic covariates. Comparing to the Firth method, our proposed method reduces the computational complexity by $O(k^3)$, where k is the number of non-genetic covariates. Our method is ~10x faster than the Firth method when sample size is 20000, and 10 non-genetic covariates are being adjusted for. Through extensive numerical simulations and real data examples, we show that the proposed method is both scalable and accurate in estimating the genotype ORs in genome-wide or phenome-wide scale.

3495W

Harmonizing genetic ancestry and self-identified race/ethnicity (HARE) for genome-wide association study in the Million Veteran Program. *H. Fang¹, Q. Hu², J. Huang^{3,4}, S. Damrauer^{5,6}, J. Honerlaw^{4,7}, T. Assimes^{1,8}, J. Lynch⁹, P. Wilson², C. O'Donnell^{3,4}, S. Pyarajan^{3,4}, M. Gaziano^{4,7}, S. DuVal¹⁰, K. Cho^{3,4,7,10}, P. Tsao^{1,8}, K. Chang^{5,6}, Y. Sun^{2,11}, H. Tang¹ on behalf of the VA Million Veteran Program.* 1) Stanford University, Stanford, CA; 2) Emory University, Atlanta, GA; 3) Harvard Medical School, Boston, MA; 4) Veterans Affairs Boston Healthcare System, Boston, MA; 5) Corporal Michael Crescenz VAMC, Philadelphia, PA; 6) Perlmutter School of Medicine, University of Pennsylvania, Philadelphia, PA; 7) Massachusetts Veterans Epidemiology and Research Information Center (MAVERIC), Boston, MA; 8) Palo Alto VA Health Care System, Palo Alto, CA; 9) VA Informatics and Computing Infrastructure, Salt Lake City VA, Salt Lake City, UT 84148; 10) Brigham and Women's Hospital, Boston, MA; 11) Atlanta VA Health Care System, Atlanta, GA.

Background: Minorities are underrepresented in many epidemiologic studies and biorepository programs to date. Large-scale, multi-ethnic cohorts, such as the Million Veterans Program (MVP) provide unprecedented opportunities to elucidate genetic factors influencing complex traits and disease risks in minority populations. Among 350,000 MVP participants, 25% represent ethnic minorities. Genetic diversity of these cohorts present new challenges to analysis and interpretation. We considered the definition and utility of race/ethnicity categories in the context of multi-ethnic GWAS studies. **Method:** We have developed a semi-supervised machine learning algorithm, which integrates race/ethnicity information from survey data, electronic health records (EHR), and ancestry inferred from genetic variation. The output variable is termed Harmonized Ancestry and Race/Ethnicity (HARE). Using simulated and real phenotypes, we investigate the statistical power of identifying ethnicity-specific trait loci. **Results:** Among participants whose self-reported race/ethnicity information was ambiguous or incomplete, over 95% were assigned to one of four HARE groups (Hispanics, non-Hispanics White, non-Hispanics Black, and non-Hispanics Asian). On the other hand, of participants with self-reported race/ethnicity information, less than 0.2% had conflicting HARE information. Through simulations, we demonstrated that stratifying by HARE achieved greater power in identifying population-specific trait loci than an un-stratified analysis of the entire MVP cohort. For height, numerous ethnicity-specific loci were identified in HARE-stratified analysis, which cannot be identified by simply analyzing all individuals together and adjusting for principal components. **Conclusion:** HARE enhanced population-specific genetic association analyses by increasing sample sizes and improving interpretability. It offers a practical approach to leverage the genetic diversity in a multi-ethnic cohort, and enables meaningful joint analyses across such cohorts.

3496T

Genotype-covariate correlation and interaction disentangled by a whole-genome multivariate reaction norm model. S.Hong Lee¹, G. Ni¹, J. Van der Werf², X. Zhou¹, E. Hyppönen¹, N. Wray³. 1) University of South Australia, Adelaide, SA, Australia; 2) University of New England, Armidale, NSW, Australia; 3) University of Queensland, Brisbane, Queensland, Australia.

The genomics era has brought useful tools to dissect the genetic architecture of complex traits. We propose a reaction norm model (RNM) to tackle genotype-environment correlation and interaction problems in the context of genome-wide association analyses of complex traits. In our approach, an environmental risk factor affecting a trait of interest (e.g. BMI) can be a phenotype modeled as a continuous covariate (e.g. smoking quantity) that is itself regulated by genetic as well as environmental factors. Hence, a correct model for the genetic component of the trait of interest would need to account for its interaction with the covariate as well as its correlation with the genetic component underlying the covariate. A multivariate RNM allows the modelling of the relation between the genotype (G) and the covariate (C), so that their correlation (association) and interaction (effect modification) can be jointly modelled (genotype-covariate correlation and interaction, GCCI). We demonstrate using simulation that the proposed multivariate RNM better performs than the current state-of-the-art methods that ignore G-C correlation and generate an inflated signal for interactions. Deflated p-values are also observed in a real data example on detecting BMI-smoking interaction when BMI-smoking correlations are not properly modelled. We also show using simulation that the estimated residual variances from genomic restricted maximum likelihood (GREML) or linkage disequilibrium score regression (LDSC) can be inflated in the presence of interaction, implying that the currently reported SNP-heritability estimates from these methods should be interpreted with caution. Correspondingly, real data analyses demonstrate that the estimated residual variance of BMI from GREML is significantly higher than a RNM fitting multiple covariates (smoking and neuroticism) jointly (p-value = 1.57E-04). We conclude that it is essential to account both for interaction and correlation when dissecting the genetic architecture of complex traits, and that the failure to do so may lead to substantial bias in estimated variance components of interaction, as well as the estimated SNP-heritability.

3497F

A meta-analysis method to detect pleiotropic loci of correlated traits.

C.H. Lee^{1,2}, H. Shi³, P. Bogdan^{4,5}, H. Lee⁶, E. Eskin^{4,7}, B. Han^{1,2}. 1) Department of Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea; 2) Department of Convergence Medicine, University of Ulsan College of Medicine, Seoul, Republic of Korea; 3) Bioinformatics Interdepartmental Program, University of California, Los Angeles, California, USA; 4) Department of Human genetics, University of California, Los Angeles, Los Angeles, California, USA; 5) Department of Pathology and Laboratory Medicine, University of California, Los Angeles, Los Angeles, California, USA; 6) Department of Biochemistry and Molecular Biology, university of Ulsan College of Medicine, Seoul, Republic of Korea; 7) Department of Computer Science, University of California, Los Angeles, Los Angeles, California, USA.

Genome-wide association studies have identified many genetic risk loci, some of which have been found to be associated with multiple complex traits. Identifying these pleiotropic loci can be valuable because they can be potential drug targets to control the cause of comorbid conditions. Recent studies have conducted cross-phenotype meta-analyses using summary statistics of readily accessible large-scale consortium data. In this study design, it is reasonable to assume that the pleiotropic loci may have differing effect sizes to multiple traits. Therefore, the use of random effect meta-analysis model assuming heterogeneity can be more suitable for identifying novel loci than the use of ordinary fixed effect meta-analysis model. The difficulty of this cross-phenotype meta-analysis is that the model should optimally account for both the genetic correlations for each pair of the traits as well as the environmental correlations due to the cryptic relatedness or the overlapping samples. Here, we developed a novel random effects model meta-analysis framework which is optimized for detecting pleiotropic loci. Using the linkage disequilibrium information from the reference data and the summary statistics of the diseases, our framework can distinguish the genetic and environmental correlations. Then, we use this information to maximize the statistical power for identifying pleiotropic loci. We applied our framework to analyze seven autoimmune disease data obtained from published results. In this analysis we confirmed associations at 210 known genes to autoimmune conditions and identified novel pleiotropic associations at five loci: *BCL2L11*(2q13, p=4.18E-8), *RPL13AP21*(12q13.11, p=2.45E-8), *NEURL4*(17p13.1, p=1.95E-8) and 2 intergenic variants including: rs9852014(3q21, p=2.06E-28) and rs1464446(3q24, p= 7.42E-10). We anticipate that our framework can help identifying pleiotropic loci with heterogeneous effects in cross-phenotype meta-analysis.

3498W

Prioritizing variants and reducing false positive associations in GWAS.

K. Lin, A. Auton, 23andMe Research Team. 23andMe, Mountain View, CA.

With the deeper coverage of sequencing technology, a significant number of new rare variants are identified and tested for associations with phenotypes in genome-wide association studies (GWAS). The conventional genome-wide significance threshold 5×10^{-8} is based on the estimated number of independent common variants. However, rare variants are in low linkage-disequilibrium with common variants, thus applying the same significance threshold to rare variants is guaranteed to increase false positives as there are much more rare than common variants. Studies have suggested using more stringent thresholds for rare variants; however, there is no consensus on the appropriate threshold as it depends on the number of rare variants included in the studies. Without estimating the number of independent variants, the naive solution of assuming all rare and common variants are independent and utilizing Bonferroni correction leads to an extremely conservative test. Alternatively, controlling for false discovery rate (FDR) is known to be not as conservative as Bonferroni correction, and therefore it could be well-suited to reduce false positives while maintaining desirable power. Independent Hypothesis Weighting (IHW; Ignatiadis et.al. 2016) is an approach to control FDR and can incorporate covariates to prioritize variants to improve power. Here, we employed IHW to re-analyze GWAS results. With controlling for FDR, we supplied functional categories and allele frequency as covariates to IHW. We aimed to answer two questions: (1) if applying IHW can remove false positive signals due to inappropriate threshold and retain true signals, and (2) if applying IHW can identify additional new signals. Applying IHW to various disease phenotypes, we observed that IHW suggested to down-weight rare variants and up-weight common variants. This is intuitive as in general GWAS have more power to detect associations for common variants; down-weighting rare variants leads to reduced false positive signals and retaining true signals from common variants. In addition to up-weighting common variants, IHW also suggested to up-weight coding variants; for example, protein-truncating variant rs34358 was sub-significant in the original type 2 diabetes GWAS (p-value: 1.5×10^{-6}) and became significant (q-value: 0.00071) after applying IHW. Together, these results suggest that controlling for FDR and prioritizing functional variants can reduce false positives and further improve power for GWAS.

3499T

Probabilistic fine-mapping of transcriptome-wide association studies.

N. Mancuso¹, G. Kichaev², H. Shi², M. Freund³, A. Gusev⁴, B. Pasaniuc^{1,2,3}. 1) Dept of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California, Los Angeles, CA 90024; 2) Bioinformatics Interdepartmental Program, University of California, Los Angeles, CA 90024; 3) Dept of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA 90024; 4) Dana-Farber Cancer Institute, Boston, MA 02215.

Transcriptome-wide association studies (TWAS) using predicted expression have identified thousands of genes whose locally-regulated expression is associated to complex traits and diseases. In this work, we show that linkage disequilibrium (LD) among SNPs induces significant gene-trait associations at non-causal genes as a function eQTL weights used in expression prediction. Here, we introduce FOCUS, a probabilistic framework that models the induced correlation among TWAS signals to assign a probability for each gene to be causal while controlling for pleiotropic SNP effects and unmeasured causal mechanisms. Our approach yields credible-sets of genes containing the causal gene at a nominal confidence level (e.g. 90%) that can be used to prioritize genes for functional assays. In detail, we model the distribution of TWAS statistics using a multivariate Gaussian distribution parametrized by LD, eQTL weights, and causal gene effect sizes. We control for bias from pleiotropic SNP effects by including an intercept term that models the average SNP effects tagged by predicted expression. Importantly, to allow for genes without prediction models in the relevant tissue (either due to QC or low power in eQTL studies), we leverage recent work demonstrating that eQTLs are shared across tissues and include prediction models from proxy tissues for such genes. To account for missing causal molecular mechanisms, we include the null model as a possible outcome in the credible set. We validate our approach through extensive simulations starting from real genotype data and find that FOCUS is accurate under a variety of trait architectures, with only a slight decrease in performance when proxy tissues are used. When causal mechanisms are absent, we find FOCUS prioritizes the null model in most simulations, despite nearby significant genes. To illustrate FOCUS's performance in real data, we re-analyzed TWAS results for lipids. Using 90%-credible sets, FOCUS prioritizes a single gene at 61/75 risk regions for HDL, LDL, triglycerides, and total cholesterol. FOCUS finds several genes with well-studied functional impact on HDL and LDL are highly ranked using our method. For example, *GALNT2* and *SORT1* genes were ranked with marginal posterior probabilities >0.95 . Overall, FOCUS prioritizes genes with well-characterized functional impact and while accounting for confounding when predicted expression is unlikely to explain observed genetic risk.

3500F

Power and sample size calculations for genetic association studies in the presence of genetic model misspecification. C.M. Moore^{1,2}, S. Jacobson², T.E. Fingerlin². 1) Division of Biostatistics and Bioinformatics, National Jewish Health, Denver, CO; 2) Center for Genes, Environment and Health National Jewish Health, Denver, CO.

When analyzing data from large-scale genetic association studies, such as data generated from targeted or genome-wide resequencing studies, it is common to assume a single genetic model, such as dominant or additive, for all tests of association between a given genetic variant and the phenotype. However, for many variants, the chosen model will result in poor model fit and may lack statistical power due to model misspecification. This potential genetic model misspecification is particularly important when genetic variants identified in one race or ethnic background are further investigated and tested in other races or ethnic groups. Multiple degree of freedom tests, sometimes called 'genotypic' tests, do not assume a particular genetic model and therefore offer more flexibility in modeling the relationship between genotype and phenotype. While the flexibility of these tests is appealing, they are often not used in practice for multiple reasons, including ease of application using popular analysis software. In addition, it is often assumed that the test lacks power due to small numbers of subjects homozygous for the minor allele and the extra degrees of freedom used in testing. Multiple degree of freedom tests are less powerful than tests from a correctly specified 1-degree of freedom genetic model, but can be more powerful than an incorrectly specified model, depending on the sample size, minor allele frequency, and the true underlying genetic model. We have developed an open-source software program ("genpwr") that performs power and sample size calculations under mis-specification of the genetic model (for example using a dominant test, when the true genetic model is recessive). The package allows for dichotomous or continuous phenotypes, adjustment for covariates, and gene x environment interactions. Understanding the impact of model misspecification on power and detectable effect sizes can aid in study design and in the development of appropriate analysis plans that maximize power to detect a range of true underlying genetic effects. In particular, these calculations can help identify when a multiple degree of freedom test may be advantageous given 1) the wide range of sample sizes now available and 2) the varying level of certainty in the appropriate genetic model, especially across race and ethnic groups. We illustrate how we used these methods in the development of an analysis plan for a resequencing study in idiopathic pulmonary fibrosis.

3501W

Drugs for specific diseases tend to target genes whose expression is causally linked to those diseases. S. Rüeger^{1,2}, E. Porcu^{2,3}, Z. Kutalik^{1,2}.

1) Institute of Social and Preventive Medicine, Lausanne University Hospital, Lausanne, Switzerland; 2) Swiss Institute of Bioinformatics, Lausanne, Switzerland; 3) Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland.

While GWASs alone are limited to SNP-trait associations, in combination with gene expression QTLs via Mendelian randomization (MR) studies they have the potential to pinpoint key causal genes for diseases. Our MR analysis provided 603'404 causal effect estimates (of which 5'009 are significantly non-zero) between whole blood expression of 15'985 genes and 43 traits/diseases), which were combined with known drug target genes using the Drugbank.ca database. For this, we developed a score to quantify for each drug-trait pair how well the drug targets correspond to the set of genes causally implicated for the given trait. For 15 diseases (among the 43 traits) we compiled a curated list of known treatments, 309 *disease-specific drugs* in total. First, for a given disease, we compared the score of drugs known to treat that disease to that of all other drugs, and found that disease-specific drugs score significantly ($P < 0.05$) higher for 6 out of 15 diseases, with depressive symptoms and schizophrenia showing the clearest separation ($P = 8 \times 10^{-6}$ and $P = 6 \times 10^{-4}$, respectively). Second, for each drug we calculated the rank of its score with the disease for which it is clinically used relative to the scores with the other 42 traits. We found that for 54 out of 309 drugs the score for the relevant disease ranked in the top five (out of 43). Notably, certain drugs for fasting insulin, HDL, LDL (Mifepristone, Saxagliptin, Niacin, Ezetimibe) ranked first. Given these promising findings, our drug-disease score could provide guidance for repositioning existing drugs to off-target diseases.

3502T

Genetic and genomic stability across lymphoblastoid cell line expansions. L. Scheinfeldt¹, K. Hodges¹, J. Pevsner², D. Berlin¹, N. Turan¹, N. Gerry^{1,2}. 1) Coriell Institute for Medical Research, Camden, NJ; 2) Kennedy Krieger Institute, Baltimore, MD; 3) Advanced BioMedical Laboratories, Cinnaminson, NJ.

Lymphoblastoid cell lines (LCLs) are a valuable resource to the genetics and genomics communities primarily due to their ability to provide a renewable source of DNA. Other common applications include studies of gene expression and gene regulation, drug metabolism, immunology, cell biology, and more recently in the development of iPSCs. Previous work has characterized variant stability in transformed cultures and across culture passages. Our objective was to extend this work to evaluate single nucleotide polymorphism and structural variation across cell line expansions. Our study included DNA extracted from whole blood, transformed cell lines, and several generations of expansions sampled from six research participants. In total, 29 LCLs were successfully established and sequentially expanded three times, and 118 DNA samples and 11 technical replicates were processed for analysis on the Affymetrix Human SNP Array 6.0. PennCNV was used to call CNVs from the B allele frequencies, log R ratios, and signal files for each sample set of autosomal chromosomes. Single nucleotide variation was stable between DNA and expanded cell lines (ranging from 99.90% to 99.98% concordance). Structural variation was less consistent across expansions (median 33% concordance) with a noticeable decrease in later expansions. These results are most applicable to genome-wide association studies (GWAS). By design, GWAS gain statistical power to detect modest disease risk effect sizes with increased sample sizes. One way in which GWAS can increase sample sizes is to leverage existing clinical samples and biorepositories that offer DNA extracted from LCLs. Our results support this strategy and suggest that SNPs can be reliably measured from LCL DNA. More caution should be taken in studies that are focused on the accurate measurement of germline structural variation in culture given the documented changes in CNV loci across LCL transformation, passaging, and expansion.

3503F

Gene hunting with hidden Markov model knockoffs. M. Sesia¹, C. Sabatti^{1,2}, E. Candès¹. 1) Department of Statistics, Stanford University, Stanford, CA; 2) Department of Biomedical Data Science, Stanford University, Stanford, CA.

Modern scientific studies often require the identification of a subset of explanatory variables. Several statistical methods have been developed to automate this task, and the framework of knockoffs has been proposed as a general solution for variable selection under rigorous type-I error control, without relying on strong modeling assumptions. In this paper, we extend the methodology of knockoffs to problems where the distribution of the covariates can be described by a hidden Markov model. We develop an exact and efficient algorithm to sample knockoff variables in this setting and then argue that, combined with the existing selective framework, this provides a natural and powerful tool for inference in genome-wide association studies with guaranteed false discovery rate control. We apply our methodology to datasets on Crohn's disease and some continuous phenotypes.

3504W

LabWAS: Study designs and novel findings from a meta-analysis of quantitative lab values from independent health systems. *J.S. Weinstock¹, J. Goldstein², M. Zawistowski¹, L. Bastarache³, D.B. Larach⁴, L.G. Fritsche¹, C.M. Brummett⁴, S. Kheterpal⁴, G.R. Abecasis¹, J.C. Denny^{3,5}.* 1) Department of Biostatistics and Center for Statistical Genetics, University of Michigan School of Public Health, Ann Arbor, MI; 2) Department of Pathology, Northwestern Memorial Hospital, Chicago, IL; 3) Department of Biomedical Informatics, Vanderbilt University, Nashville, TN; 4) Department of Anesthesiology, University of Michigan Medical School, Ann Arbor, MI; 5) Department of Medicine, Vanderbilt University Medical Center, Nashville, TN.

Electronic Health Record (EHR) systems have enabled discovery of novel associations across a rich collection of phenotypes. In contrast to phenotypes measured exclusively for research, EHR derived phenotypes are influenced by unique properties of the host health system. Recruitment mechanisms, which vary across biobanks, are likely to affect which participants are included and the nature of the measurements. The diversity of disparate health systems makes analyses across multiple EHR systems challenging, which is especially true for laboratory measurements, which may differ in measurement protocols. We performed a meta-analysis of GWASes of 66 quantitative laboratory measurements across two independent health systems: the Michigan Genomics Initiative (MGI) from Michigan Medicine and BioVU from Vanderbilt University. Participants in MGI were recruited through surgical encounters at Michigan Medicine, and 37,412 blood samples were genotyped. The BioVU cohort was recruited on patient check-in during outpatient encounters, and 20,515 genotyped blood samples were used for this analysis. The size of individual GWASes varied by the frequency of those laboratory measurements. We restricted analysis to unrelated individuals of European ancestry. Meta-analyses used the intersection of 798,338 SNPs genotyped in both studies. After aligning GWAS analysis protocols, we systematically examined different study design choices in parallel to study the impact of analysis decisions on GWAS results. We examined laboratory measurement context and the consequences of different health state covariates in analysis. We also performed GWASes based on a subset of measurements that were recorded in outpatient contexts, to assess the tradeoff of sample size and phenotype fidelity. For example, for glucose, we observed the health state covariates to reduce power in known loci in both systems. We observed heterogeneity in these effects across laboratory phenotypes – for red blood cell counts, using only outpatient values improved power in BioVU, but reduced power in MGI. We produced a set of best analysis practices for performing GWASes on most common laboratory traits in EHR systems. We found that taking the mean of lab measurements, applying an inverse normal transformation, excluding disease covariates, and using all measurements in all contexts to be a useful default practice. The results of our meta-analyses are browsable at a public web portal, pheweb.sph.umich.edu:5004.

3505T

Bivariate logistic Bayesian LASSO for finding genetic association with two correlated phenotypes. *X. Yuan.* University of Texas at Dallas, Richardson, TX.

Often multiple correlated outcomes are collected in health-related studies. However, typically each outcome is analyzed separately. In genetic association studies, joint modeling of multiple traits (phenotypes) can utilize the correlation amongst them and thereby be more powerful and insightful. Also, rare variants are currently a major focus of genetic research as a key to missing heritability. A recent method named logistic Bayesian LASSO (LBL) can detect rare haplotype variants associated with a single binary phenotype. Here we extend LBL to jointly model two correlated binary traits. Our proposed bivariate LBL model uses a latent variable to induce correlation between the two outcomes. To compare the bivariate LBL with the univariate LBL, we simulate two binary traits using a probit model with varying correlation between the two traits. We consider a number of settings, in which a rare haplotype is associated with one or both traits with same or different directions of effects. Our results show that the bivariate LBL performs better or similar to univariate LBL in most settings. It has most gain in power when there is a high positive correlation between the traits and the effects of the associated haplotype on the two traits are of opposite directions.

3506F

Implications of post-colonial demographic structure on interpretations of association analyses. *M. Zhang¹, J.M. Granka¹, E.P. Sorokin¹, K.A. Rand¹, S. Hateley¹, K. Noto¹, E.L. Hong¹, K.G. Chahine¹, C.A. Ball².* 1) AncestryDNA, San Francisco, CA; 2) AncestryDNA, Lehi, UT.

Recent work has highlighted the need to evaluate the impact of fine-scale population structure on association analyses. Despite thousands of high-powered genome-wide association studies (GWAS), only about one-fifth of all studies include individuals of non-European descent. It is understood that GWAS conducted in European populations are not always generalizable to non-European populations, and that continental population structure is an important factor to consider. However, in the age of regionally-run genomic studies, it remains unclear whether fine-scale population structure, particularly more recent within-continental structure, contributes to bias in the interpretation of GWAS results. We identify regional fine-scale structure, within continental ethnicity groups, in a subset of over one million AncestryDNA customers who have consented to research. Such regional United States population structure, identified to relate to post-colonial demography during the last several hundred years, is generally undetectable in a principal component analysis (PCA), although regional allele frequency variation at the single-variant level still exists. To explore whether this variation plays a role in association analyses, we examine the regional distribution of polygenic risk scores (PRS) calculated from previously-identified variants in the GWAS catalog for various well-studied phenotypes. To account for continental population structure, we examine PRS distributions using variants that have been identified in the same continental population. While observed differences occur primarily across continental groups and only some regional groups, in some cases these differences could suggest residual confounding due to within-continental population structure. It is unclear whether the observed differences can be attributed to true differences in disease risk, or whether they are solely an artifact of generalizing GWAS results to other, unstudied populations. Even when considering very recent time scales, small frequency differences may impact GWAS conclusions and interpretations -- particularly in the transferability of results to similar, but unstudied, populations. This work highlights the fundamental need to understand the generalizability of GWAS results through the lens of both continental, and within-continental, population structure, particularly in large-scale analyses.

3507W

On set-based association tests: Insights gained from a regression, based on summary statistics. *Y. Zhao¹, L. Sun^{1,2}.* 1) Department of Statistical Sciences, University of Toronto, Toronto, Canada; 2) Dalla Lana School of Public Health, University of Toronto, Toronto, Canada.

Set-based tests that jointly analyze a group of variants have been applied in various settings, including rare variants analysis. In each case, myriad tests have been proposed with competing and sometimes conflicting claims in terms of method performance. We developed a unifying regression framework based on summary statistics. We show that existing tests such as burden, SKAT and SKAT-O can be derived as a special case from this new architecture. Moreover, the regression framework provides a straightforward way to account for any inherent correlations between the variables in a set. Subsequently, we propose new test statistics that are SKAT-O-like, but without the apparent 'optimal' weighting. Under mild conditions, we show that the asymptotic distribution of the new test statistics are Chi-square distributed, and we provide accurate p-value calculation for finite samples with respect to the number of variables analyzed in a set. Focusing on rare variants analysis and through extensive simulation studies, we show that the proposed method is often, but not always, more powerful than SKAT-O across a number of alternative scenarios. In addition, the summary statistics-based regression framework is easy to implement, and it provides a principled way for additional improvements such as incorporating available genomic information for data integration. Finally, previous tests designed for PhewWAS, Polygenic Risk Score (PRS), PrediXcan or TWAS can also be re-examined through this new lens.

3508T

Statistical power in GWAS revisited: Sample size, genetic relatedness, and gene-by-environment interactions. A. Ziyatdinov¹, J. Kim¹, P. Kraft¹, H. Aschard^{1,2}. 1) Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA; 2) Centre de Bioinformatique, Biostatistique et Biologie Intégrative (C3BI), Institut Pasteur, Paris, France.

Genome-wide association studies (GWAS) have identified thousands of genetic associations for complex diseases and heavily relies on increasing the sample size. Recent analyses of biobank-scale genetic data suggest: (i) inclusion of genetically related individuals empowers GWAS (Loh *et al.*, 2017); (ii) the wealth of collected environmental exposures has potential to uncover gene-by-environment interactions (Young *et al.*, 2016). However, quantification of GWAS power - the non-centrality parameter (NCP) of an association test, which is proportional to the sample size (n) and the variance explained by genetic variant (q^2) - holds only for unrelated individuals. Here, we first expand it by incorporating individual relationships by linear mixed model. We next consider gene-by-environment interactions, where interaction effect on trait is tested in the presence of marginal genetic effect. Consider a set of related individuals, which relatedness is expressed by the kinship matrix (K), and association between outcome (y) and genotype (g) is evaluated by linear mixed model. For a given phenotypic variance-covariance matrix (V), we show that the non-centrality parameter for marginal association test can be approximated as $NCP_{rel} = \text{tr}(V^{-1}\Sigma_g)$, where tr is the trace operator and Σ_g is the covariance matrix of genotypes. For families we have $V = [h^2K + I]$ and $\Sigma_g = q^2K$, that results in $NCP_{rel} = \text{tr}([h^2K + I]^{-1}K)q^2$. That means $NCP_{rel} \leq NCP$, i.e. for the same sample size, power is expected to decrease as relatedness between individuals increases – which was previously shown analytically only for sib-pairs in (Visscher *et al.*, 2008). Further, we consider testing for fixed effect gene-by-environment interaction in related individuals. We have derived $NCP_{rel+int} = \text{tr}(V^{-1}\Sigma_{int})$, where Σ_{int} is the covariance matrix in distribution of the interaction variable. For families we have $V = [h^2K + I]$, $\Sigma_{int} = q^2w^2K_{int}$, where $w^2 = f(1-f)$ and f is the frequency of binary exposure. K_{int} is a matrix derived from K and depends on realization of exposure in a given study. We argue that K_{int} is different from the matrix recently proposed in (Sul *et al.*, 2016). We validated all results by simulations. Our formulas have a range of implications. For testing marginal genetic effect, one can quickly assess the power in studies involving related individuals. Because of the potential gain in power for testing gene-by-environment interaction, the formula will allow optimization of the study design of related individuals.

3509F

Joint analysis of multiple phenotypes using a clustering linear combination method based on hierarchical clustering. X. Li, Q. Sha, S. Zhang. Department of Mathematical Sciences, Michigan Technological University, Houghton, MI.

Emerging evidence suggests that a genetic variant can affect multiple distinct, but possibly correlated phenotypes, especially in complex human diseases. Individual phenotype analyses are generally less informative and underpowered to uncover the genetic variants underlying complex traits and diseases. The joint analysis of multiple phenotypes may offer new insights into the disease etiology. In this paper, we develop an innovative method for joint analysis of multiple phenotypes using a hierarchical clustering approach followed by a clustering linear combination method. We have named this method the HCCLC method. The proposed method consists of two consecutive steps: the Hierarchical Clustering (HC) and the Clustering Linear Combination (CLC). The HC step partitions the phenotypes into a small number of clusters. The phenotypes within each cluster strongly correlate with each other, while phenotypes between clusters are significantly less likely to be correlated. The CLC method is then adopted to test the association between a genetic variant of interest and multiple phenotypes. This is done by combining individual test statistics, while taking advantage of the natural grouping information of phenotypes. Extensive simulation studies together with a COPDGene real data analysis have been used to assess the applicability of our proposed method. Our simulation results demonstrate that the type I error rates of HCCLC are well maintained in different realistic settings. HCCLC either outperforms all other methods or has statistical power that is very close to the most powerful method we have compared it with. In addition, our real data analysis shows that HCCLC is an appropriate test for genome-wide association studies (GWAS).

3510W

Heritability estimates in admixed populations from Greenland. *G. Athanasiadis, A. Albrechtsen.* Section for Computational and RNA Biology, Department of Biology, Copenhagen University, Copenhagen, Denmark.

Standard restricted maximum likelihood (REML) methods for complex trait heritability estimates in human populations require that samples are unrelated and un-admixed. As a result, smaller non-European cohorts cannot benefit entirely from the methodological advancements that produce reliable heritability estimates in large European cohorts. In light of the above, we used three Greenlandic cohorts ($N = 4,824$) and an extensive simulation setup to investigate how REML-based heritability estimates are affected by a high degree of relatedness in the sample, as well as by recent admixture. We also propose specific heritability estimate guidelines for setups similar to ours. Finally, we also applied our model of choice to a number of quantitative traits from our Greenlandic cohorts and report considerable genetic components for many of them.

3511T

RSSp: An effective method for estimating the heritability of polygenic traits using GWAS summary statistics. *N.W. Knoblauch², X. He¹.* 1) Department of Human Genetics, University of Chicago, Chicago, IL; 2) Committee on Genetics, Genomics and Systems Biology, University of Chicago, Chicago, IL.

Genome-wide association study (GWAS) summary statistics are the most universal statistical tool for describing the relationship between genetic variants and a trait of interest in a population. As dozens of large-scale GWAS have demonstrated, the variants most significantly associated with a particular phenotype often explain only a small amount of the heritability of that trait, a reality that has pushed researchers to look towards the full distribution of marginal phenotype/genotype associations to better understand genetic architecture. Unfortunately, linkage disequilibrium (LD) distorts the observed relationship between genotype and phenotype, and must be accounted for when analyzing GWAS summary statistics. With an estimate of LD, often obtained from an out-of-sample reference panel, it is possible to account for the distortionary effect of LD on GWAS summary statistics when estimating heritability. Regression with Summary Statistics (RSS) (Zhou and Stephens 2017) is a Bayesian statistical framework for obtaining multivariate regression estimates using univariate estimates and a correlation (LD) matrix. We have developed an implementation of RSS tailored for heritability estimation for highly polygenic traits, and we call our method RSSp. To assess RSSp's effectiveness, we simulated effect sizes for hundreds of traits under a range of narrow-sense heritability values using real genotype data from the Genetics of Recurrent Early-Onset Depression (GENRED) cohort ($N=1902$), as well as the Wellcome Trust Case-Control Consortium cardiovascular disease cohort ($N=4686$). We then performed GWAS on these simulated traits, and estimated heritability using a held out set of samples as a reference LD panel. LD was estimated using either the sample correlation or using a regularization technique called LDshrink, which uses a genetic map to regularize LD estimates. Comparing our heritability estimates obtained using RSSp to heritability estimates using LD score regression, the most popular method for estimating heritability from summary statistics, two things are especially noteworthy. First, RSSp heritability estimates compare favorably to estimates from LD score regression. RSSp had a relative RMSE of 0.34, a 44 percent improvement over LD score regression (RMSE=0.60), and a 52 percent improvement over LD score regression fit without an intercept term (RMSE=0.70). Second, RSSp heritability estimates are improved by using LDshrink (RMSE=0.34 vs RMSE=0.44).

3512F

Reliable heritability estimation in admixed populations using covariate-adjusted LD score regression. X. Li^{1,2}, Y. Luo^{1,2}, J.M. Mercader^{2,3}, J. Florez², H.K. Finucane², S. Raychaudhuri^{1,2}, *The SIGMA Type 2 Diabetes Consortium*. 1) Medicine-Brigham and Women's Hospital, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 2) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA; 3) Diabetes Unit and Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA.

To date, the majority of genome-wide association studies (GWAS) have been performed in Europeans, but recent studies are now including additional populations, including admixed populations. Genetic heritability (h_g^2) is an essential measurement to understand the polygenic traits. However, current methods used to measure h_g^2 based on GWAS are designed for homogeneous populations, and may not work for admixed populations. Here, we introduce covariate-adjusted LD score regression (cov-LDSC) that can be applied to admixed GWAS when estimating h_g^2 . LD score regression (LDSC) is a family of methods that takes advantage of the linear relationship between χ^2 statistics and the linkage disequilibrium (LD) to measure h_g^2 (Bulik-Sullivan, et al., 2015). This linear relationship still holds in admixed populations; however, the LD measured by LDSC is often inflated by counting the admixture-LD, especially for population specific SNPs. We first demonstrate that the LDSC leads to deflated h_g^2 in admixed population. We simulated genotypes of 10,000 admixed samples with msprime (J. Kelleher, et al., 2016) using demographic history inferred from 1000 Genome. We then developed three simulation scenarios for evaluation: (1) different total h_g^2 ; (2) different degrees of polygenicity in simulated traits and (3) different percentages of ancestral proportions. We found that LDSC systematically underestimated h_g^2 due to inflations in LD score estimates; for example, for a 40% heritable trait in a 50% admixed population, LDSC yields an estimate of 32%, whereas cov-LDSC provides an unbiased point estimate. Next we implemented cov-LDSC where we correct inflation in LD scores by regressing out genome-wide covariates such as principal components at each SNP. Applying cov-LDSC to these same simulations, we observed that it yields robust h_g^2 estimates in all scenarios. We also simulated different phenotypes based on real genotypes of 8,214 Mexican individuals from the SIGMA cohorts (The SIGMA Type 2 Diabetes Consortium, 2014), who have in average 60% Native American and 40% European ancestry. Again, cov-LDSC yields unbiased h_g^2 estimates, compared to LDSC. We thus conclude, through extensive studies based on both simulated and real genotypes, that cov-LDSC can yield reliable h_g^2 estimates in admixed GWAS. Cov-LDSC can be applied on GWAS summary statistics without raw genotype access and has the potential to be extended to measure partitioned heritability efficiently in admixed populations.

3513W

Extending SNP-based heritability analysis: How many variants show strong effect in a GWAS. F. Takeuchi, N. Kato. Research Institute, National Center for Global Health and Medicine, Shinjuku-ku, Tokyo, Japan.

In SNP-based heritability analyses, the allele substitution effects of causal variants are modeled to follow a zero-mean normal distribution, whose variance represents the heritability. One topic unstudied so far is the proportion of SNPs showing strong associations in a GWAS, which is large for lipids and small for BMI and height. In terms of probability distribution, the proportion corresponds to the tail-heaviness, which characterizes a distribution independently of the variance. Without understanding that proportion, we cannot calculate the power of a GWAS, which is the expected number of loci to be discovered. To address those topics, we modeled the allele substitution effects of single SNPs by the t -distribution, which can adjust tail-heaviness by the degrees-of-freedom parameter. We extended the Popcorn program, which analyzes GWAS summary statistics and estimates heritability in a population and genetic correlation between two populations of different ancestries. Using the observed parameters, we simulated possible distribution of SNP effect-size. GWAS datasets for blood pressure, lipids, type 2 diabetes, BMI and height in East Asian or European-ancestry populations were analyzed. The estimated heritability ranged 0.07–0.31, and the genetic correlation ranged 0.77–0.99. The degrees of freedom were ~3 for lipids, ~5 for BMI, and ~4 for other traits, indeed indicating heavy tail for lipids. Using simulation, we computed the power of GWAS under various sample sizes. The extension of SNP-based heritability analysis by adopting t -distribution enabled the modeling of the spectrum of associations in a GWAS and power calculation. The source code is available from <https://github.com/fumi-github/Popcorn-t>.

3514T

Heritability and genetic correlation analysis of biobank data. Y. Wu, S. Sankararaman. UCLA, Los Angeles, CA.

SNP heritability, i.e. the proportion of phenotypic variance explained by SNPs, and genetic correlation are important parameters in efforts to understand the genetic architecture of complex phenotypes as well as in the design and interpretation of genome-wide association studies. Attempts to estimate SNP heritability as well genetic correlations among complex phenotypes attributable to genome-wide SNP variation data have motivated the analysis of large datasets as well as the development of sophisticated tools to estimate heritability and genetic correlation in these datasets. Linear Mixed Models (LMMs) have emerged as a key tool for heritability and genetic correlation estimation where the parameters of the LMMs, i.e., the variance components, are related to the heritability attributable to the SNPs analyzed and genetic correlations across phenotypes. Likelihood-based inference in LMMs, however, poses serious computational burdens. We propose a scalable randomized Method-of-Moments (MoM) estimator of SNP heritability and genetic correlations in LMMs. Our method, RHE-reg, leverages the structure of genotype data to obtain runtimes that are sub-linear in the number of individuals (assuming the number of SNPs is held constant). We perform extensive simulations to validate the accuracy and scalability of our method. RHE-reg can compute the heritability and genetic correlations on the UK biobank dataset consisting of 430,000 individuals and 460,000 SNPs in 3 hours on a stand-alone computing machine. We used RHE-reg to estimate heritability and genetic correlations for all 60 traits in the UK Biobank.

3515F

Relatedness disequilibrium regression estimates heritability without environmental bias. A.I. Young^{1,2,3}, M.L. Frigge¹, D.F. Gudbjartsson^{1,4}, G.

Thorleifsson¹, G. Bjornsdottir¹, P. Sulem¹, G. Masson¹, U. Thorsteinsdottir^{1,5}, K. Stefansson^{1,5}, A. Kong^{1,3,4}. 1) deCODE genetics/Amgen Inc., Reykjavik 101, Iceland; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, U.K; 3) Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, UK; 4) School of Engineering and Natural Sciences, University of Iceland, Reykjavik 101, Iceland; 5) Faculty of Medicine, University of Iceland, Reykjavik 101, Iceland.

Heritability measures the relative contribution of genetic inheritance (nature) and environment (nurture) to trait variation. Estimation of heritability is especially challenging when genetic and environmental effects are correlated, such as when indirect genetic effects from relatives are present. An indirect genetic effect on a proband (phenotyped individual) is the effect of a genetic variant on the proband through the proband's environment. Examples of indirect genetic effects include effects from parents to offspring, which occur when parental nurturing behaviours are influenced by parents' genes. Indirect genetic effects from parents to offspring and between siblings have been shown to be substantial for educational attainment. We show that existing methods for estimating heritability, including GREML methods based on SNP sharing between distant relatives, can be severely biased by indirect genetic effects. We introduce a novel method for estimating heritability that does not exhibit such bias: relatedness disequilibrium regression (RDR). RDR removes environmental bias by exploiting variation in relatedness due to random Mendelian segregations in the probands' parents. We show mathematically and in simulations that RDR estimates heritability with negligible bias due to environment in almost all scenarios. We use a sample of 54,888 Icelanders with both parents genotyped to estimate the heritability of 14 traits, including height (55.4%, S.E. 4.4%), body mass index (BMI) (28.9%, S.E. 6.3%), and educational attainment (17.0%, S.E. 9.4%). We find evidence that existing identity-by-descent based methods severely overestimate heritability for educational attainment, height, BMI, and age at first child in women. We find evidence that GREML methods overestimate heritability for height, age at first child in women, and educational attainment. We find that GREML estimates of the heritability of educational attainment are inflated by a factor of around 1.7, consistent with the estimated magnitude of indirect genetic effects from parents and siblings. Furthermore, without genotype data on close relatives of the proband - such as used by RDR - our results show that it is impossible to remove the bias due to indirect genetic effects. Unfortunately, large datasets including many close relative pairs are currently rare. We therefore argue that greater efforts to collect such data will be essential for understanding genetic causation and the nature of nurture.

3516W

SOS: Are genotype-based association tests robust to departure from Hardy-Weinberg equilibrium? L. Zhang¹, L. Sun^{1,2}. 1) Department of Statistical Sciences, University of Toronto, Toronto, Canada, Canada; 2) Dalla Lana School of Public Health, University of Toronto, Toronto, Canada.

Though the exact origin of the argument is debatable, it is agreed that genotype-based association tests are robust to the departure from Hardy-Weinberg equilibrium (HWE), or not sensitive to Hardy-Weinberg disequilibrium (HWD). Indeed, empirical evidence appears to support this statement, particularly in studies using unrelated individuals. But how about related individuals? Linear mixed model (LMM) has become the most popular method for family-based association studies. The covariance structure of the phenotype is partitioned into a weighted sum of correlation structure due to genetic relatedness (or kinship coefficient) and shared environmental effect, where the weight is usually referred as 'heritability'. We show theoretically that if the causal SNP(s) are in HWD in the founders, the kinship coefficient matrix no longer captures the true relatedness structure between the genotypes in the related individuals. Thus, LMM may not be robust to the departure from HWE. Consider the simplest sib-pair design and assume the genotypes of the (unrelated) parents are in HWD but not available, the genotypes of the siblings can be shown to be in HWE; a textbook example demonstrating that HWE can be achieved after one single generation of random mating. However, we can also show theoretically that when it comes to association analysis, HWD in the founder generation affects the result through the covariance structure between the sibling genotypes. When heritability is known, based on literature, our simulation studies show that the empirical type 1 error rate can get as high as 10% at a 5% significance level. When heritability is estimated internally by LMM, type 1 error is well controlled but the estimated heritability is then biased. In practice, the problem occurs only if the true causal variants depart from HWE considerably and the sample consists mostly of related individuals. However, as we move beyond low-hanging fruit and continue to struggle with the missing heritability issue, our findings can be impactful. Furthermore, we propose an alternative approach to family-based association studies that is robust to the departure from HWE and remains powerful in the absence of HWD.

3517T

A general framework for estimating the effects of molecular mechanisms on complex traits. K.S. Burch¹, N. Mancuso², R. Johnson³, H. Shi¹, G. Kichaev¹, C. Kalita⁴, R. Pique-Regi^{4,5}, S. Sankararaman^{3,6}, B. Pasaniuc^{2,6,7}.

1) Bioinformatics Interdepartmental Program, University of California Los Angeles, Los Angeles, CA, USA; 2) 1. Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA; 3) Department of Computer Science, University of California Los Angeles, Los Angeles, CA, USA; 4) Center for Molecular Medicine and Genetics, Wayne State University, Detroit, MI, USA; 5) Department of Obstetrics and Gynecology, Wayne State University, Detroit, MI, USA; 6) Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA; 7) Department of Biomathematics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA.

Identifying the genes and pathways that drive GWAS signal in noncoding regions remains a major challenge in deriving clinically relevant meaning from GWAS. Methods that partition SNP-heritability by genomic annotation (e.g., stratified LD score regression) can identify regulatory regions enriched with SNP effects of large magnitude; these methods assume that all allelic effects have mean 0 in the population and variances that depend on their annotation profiles. While useful, functional enrichment is limited in interpretability as it cannot elucidate specific biological mechanisms – e.g., the effect of increased binding of a given transcription factor (TF) on trait. In contrast, methods for estimating the effects of molecular mechanisms on trait (i.e., signed LD profile regression) model the expectation of allelic effects on trait as functions of their effects on the mechanism of interest while assuming the variances of the allelic effects are equal. Here, we propose MV-reg (Mean Variance REGression): a generalized framework for integrating QTLs or any allele-specific annotation with GWAS to estimate both the magnitude and direction of effect of a molecular mechanism (e.g., TF binding or gene expression) on a complex trait. Continuing with the TF binding example, MV-reg models both the expectation and variance of allelic effects on trait as functions of their effects on TF binding, and estimates the genome-wide effects of altered binding on both the mean and variance to identify trait-relevant TFs. We show via simulations that MV-reg produces unbiased estimates of the effects of TF binding on trait, and can use the estimated effects of TF binding on variance to prioritize TFs with equal mean effects. Lastly, we apply MV-reg to estimate the effects of 1370 TFs on BMI and LDL, and find a total of 8 TFs with significant nonzero effects (FDR < 0.1), 1 of which affects both traits. Our results are consistent with known biology; e.g., we find that increased binding of RORA1 (involved in lipid metabolism) and IPF1 (a transcriptional activator of insulin) are associated with higher BMI, and increased binding of AGL2 (involved in glycogenolysis) is associated with both higher BMI and lower LDL. We find an additional 9 TFs with nonzero effects on BMI or LDL (1 of which has effects on both) that do not pass the threshold for significance but show increased evidence for association due to their statistically significant effects on decreasing the allelic variances.

3518F

Finding rare variants of large effect using identity-by-descent mapping.

S. Shringarpure, D. Hinds, V. Vacic, S. Pitts, R. Gentleman, A. Auton, 23andMe Research Team. 23andMe Inc, Mountain View, CA., USA.

With the availability of large cohorts, there is now increased potential for discovering rare variant associations in various phenotypes. However, due to limited imputation accuracy for rare variants, GWAS studies are not well-powered to discover rare variant associations. We demonstrate the utility of identity-by-descent (IBD) mapping methods to discover rare associations using simulations and application on data from 23andMe research participants. We performed extensive simulations examining the power of IBD mapping to detect rare associations as a function of rare variant frequency, effect size, disease prevalence and sample size. We find that IBD mapping can discover associations rarer than 0.5% in frequency, which cannot be detected using GWAS on imputed data. We applied our method on consented research participants from the 23andMe cohort and analyzed 200 phenotypes. Typically, we observe far fewer associations than GWAS, however, we see that discovered associations are rarer and have much larger effect sizes than typical GWAS associations. For instance, in an association analysis for morningness in individuals of Ashkenazi ancestry (11,625 cases, 11,509 controls), we find 2 large-effect rare haplotype clusters in regions previously reported as having common variant associations for chronotype [1], sleep disturbances [2], and insomnia [3], along with 2 novel haplotype clusters. In an association analysis for gout in Ashkenazi Jews, we find an association overlapping *ALDH16A1* (OR=5.7, P=6e-12, Freq=0.55%). WGS data from an individual in the associated cluster shows a rare missense mutation (rs150414818) in *ALDH16A1*. This variant has been previously reported for association with gout in Icelandic populations, where it is more common in frequency (1.9%) and has a weaker reported effect (OR=3.12). In addition to these associations, we find a number of rare variant associations in gene-phenotype combinations where common variant associations have previously been reported, as well other novel associations. These results suggest that IBD mapping can be used to find rare associations that may not be discovered through GWAS. References: [1] Jones et al. bioRxiv 2018 <https://www.biorxiv.org/content/early/2018/04/19/303941> [2] Jansen et al. bioRxiv 2018 <https://www.biorxiv.org/content/early/2018/01/30/214973> [3] Lane et al. Nature Genetics 2017 <https://www.nature.com/articles/ng.3749>.

3519W

Prenatal diagnostics of congenital anomalies in assisted reproduction pregnancies: Population-based data from the Czech Republic.

A. Šípek Jr^{1,2}, V. Gregor^{1,3}, A. Šípek Sr.^{1,3,4}, J. Klaschka^{5,6}, M. Malý^{6,7}, J. Jírova⁸. 1) Department of Medical Genetics, Thomayer Hospital, Prague, Czech Republic; 2) Institute of Biology and Medical Genetics, First Faculty of Medicine, Charles University, Prague, Czech Republic; 3) Department of Medical Genetics, Pronatal Sanatory, Prague, Czech Republic; 4) Institute of Biology and Medical Genetics, Third Faculty of Medicine, Charles University, Prague, Czech Republic; 5) Institute of Computer Science of the Czech Academy of Sciences, Prague, Czech Republic; 6) Institute of Biophysics and Informatics, First Faculty of Medicine, Charles University, Prague, Czech Republic; 7) National Institute of Public Health, Prague, Czech Republic; 8) Institute for Health Information and Statistics, Prague, Czech Republic.

Introduction: The proportion of children born after the use of assisted reproduction techniques is increasing worldwide. In the Czech Republic there were 3.65% children born after assisted reproduction in 2015. Various studies have confirmed higher incidence of selected groups of congenital anomalies among the assisted reproduction children. We are presenting first results of Czech population-wide study focusing the spectrum of congenital anomalies diagnosed during prenatal diagnostics among assisted reproduction pregnancies. **Methods:** Data were obtained from the official medical registries run by the Institute of Health Information and Statistics of the Czech Republic, namely National Registry of Congenital Anomalies and the National Registry of Assisted Reproduction. Those registries are run by the government, have population-wide coverage and the registration process is compulsory by the law. All cases of congenital anomalies (Q00-Q99 codes of ICD-10 classification) that were diagnosed prenatally among assisted reproduction pregnancies were further analyzed. Time period: 2007-2015. **Results:** During the selected time period, there were 538 cases of assisted reproduction pregnancies diagnosed with different types of anomalies. During the same time period, there were 1.012.375 children born in the Czech Republic (among those there were 29.899 = 2.95% assisted reproduction children). Most of those anomalies were part of chromosomal aberration syndromes (numerical and structural): 332 cases (61.7%). Among the non-syndromic cases – the most common were the anomalies of the central nervous system: 71 cases (13.2%), followed by the anomalies of the urinary tract (54 cases = 10.0 %), heart defects (50 cases = 9.3 %) and gastrointestinal/abdominal wall anomalies (39 cases = 7.3%). **Discussion and conclusion:** The distribution of anomalies diagnosed in assisted reproduction pregnancies resembles the distribution of anomalies prenatally diagnosed in spontaneous conception pregnancies. The current limitations in different registries databases connection (incomplete data in healthy controls) do not allow us to provide exact statistical evaluation of this difference right now, but this issue will be overcome in the next stage of this project. **Acknowledgements:** Supported by Ministry of Health of the Czech Republic, grant nr. AZV 17-29622A and RVO project: "Thomayerova nemocnice – TN, 00064190".

3520T

Machine learning of causal biological networks with the principle of Mendelian randomization. *M.B. Badsha, E.A. Martin, A.Q. Fu.* Department of Statistical Science, Institute of Bioinformatics and Evolutionary Studies, Center for Modeling Complex Interactions, University of Idaho, Moscow, ID 83844-1104.

Although large amounts of genomic data are available, it remains a challenge to reliably infer causal relationships among molecular phenotypes (such as gene expression). It is even harder to learn a causal network of many phenotypes. Correlation (or association) is often used as a proxy of a potential causal relationship, but similar levels of correlation can arise from different causal mechanisms. However, a genetic variant (e.g., an expression quantitative trait loci; or eQTL) associated with one of the two genes can help break the symmetry and determine the direction of the edge between genes. This is the rationale behind the Principle of Mendelian Randomization (PMR), which assumes that the alleles of a genetic variant are randomly assigned to individuals in a population, analogous to a natural perturbation experiment and therefore achieving the goal of randomization. We present MRPC, which learns a causal biological network efficiently and robustly from integrating genotype and molecular phenotype data, in which directed edges indicate causal directions. MRPC is the first machine learning algorithm that incorporates the PMR in classical algorithms for learning causal graphs in computer science (bioRxiv, 171348; arXiv, 1806.01899). MRPC performs a large number of statistical independence tests to determine whether an edge should be retained. It implements an online false discovery rate (FDR) control method that controls the overall FDR when the total number of tests is unknown beforehand. We demonstrate through simulation that MRPC achieves higher recall, precision and stability than existing general-purpose network inference methods (such as those implemented in R packages pcalg and bnlearn) and methods using the PMR (e.g., CIT, one of the most popular PMR-based method, and findr, a more efficient version of CIT). We applied MRPC to the GEUVADIS data to distinguish direct and indirect targets among multiple genes associated with expression quantitative trait loci (eQTLs). We also constructed a causal network for frequently altered cancer genes, using the copy number variation (CNV) and gene expression data from ER+ breast cancer patients in the TCGA consortium. From the cancer gene network, we further identified six modules of CNVs and genes that are enriched for distinct biological processes or pathways.

3521F

Heterogeneity of causal estimates inferred from single genetic variants as instrumental variables in Mendelian randomization across complex traits and diseases. *A. Dobbyn^{1,2,3}, M. Verbanck^{1,3,4}, C.Y. Chen^{5,6,7}, D.M. Jordan^{1,3,4}, B.M. Neale^{5,6,7}, R. Do^{1,3,4}.*

1) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY; 2) The Pamela Sklar Division for Psychiatric Genomics, Icahn School of Medicine at Mount Sinai, New York, NY; 3) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 4) The Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY; 5) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 6) Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, MA; 7) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA.

Multi-instrumental variable (IV) MR approaches utilize multiple single nucleotide variants (SNVs) as IVs to infer causality of an exposure with an outcome. Recently, we detected widespread horizontal pleiotropy in causal relationships inferred from MR among complex traits and diseases. As a result, there may be significant heterogeneity in the effects of IVs in MR, suggesting that single-IV MR approaches that account for heterogeneity of causal effects may be more appropriate than multi-IV MR in the context of horizontal pleiotropy. We conducted single IV MR for 82 GWAS traits (39 exposures and 82 outcomes) totaling 140,887 tests. We compared causal relationships inferred through single-IV MR with analogous results from multi-IV MR tests (implemented in MR-PRESSO), which included detection of horizontally pleiotropic outliers. We compared outlier and non-outlier SNVs in the context of their contributions to pleiotropy, as measured by a recently developed score that assesses pleiotropy for a given SNV through the total magnitude of its effect across traits (P_m) and through its number of associations (P_n). Across all exposure-outcome pairs, a total of 1,096 SNVs were significant for the single-IV MR test at a globally-corrected threshold ($P < 0.05/\text{total tests}$, 3.55×10^{-7}), 1,892 were identified as outliers, and 530 were both MR-significant and outliers. Outlier SNVs had a significantly greater mean pleiotropy score ($P_m + 6.0$, $P = 2.65 \times 10^{-27}$; $P_n + 13.47$, $P = 3.88 \times 10^{-32}$) than non-outliers. Of 3,157 exposure-outcome pairs, 325 had at least one significant single-IV MR variant and therefore evidence for causality. In comparison, out of 129 relationships that were significant for the multi-IV MR test, 84 had a SNV that was also significant for the single-IV MR test. In conclusion, single- and multi-IV MR are complementary methods; though 84 causal relationships were identified by both methods, 241 and 45 relationships were identified by only single- or multi-IV MR, respectively. Single-IV MR has added utility when pleiotropy and heterogeneity are present. For example, lack of significance in single-IV MR for the known causal relationships between both triglycerides (TG) and LDL-C, and coronary artery disease (CAD) due to opposite directions of effect for a given SNV (e.g. rs174546) reveals it to be a likely ineffective drug target for CAD. Therefore, single-IV MR has the potential to inform drug discovery in ways in which multi-IV MR cannot.

3522W

Bayesian variable selection for Mendelian randomization. *A. Gkatzionis, P.J. Newcombe.* MRC Biostatistics Unit, University of Cambridge, Cambridge, United Kingdom.

Mendelian randomization is the use of genetic information to assess the existence of a causal relationship between a risk factor and an outcome of interest. A Mendelian randomization analysis relies on a set of genetic variants (SNPs) that are strongly associated with the risk factor, and only associated with the outcome through their effect on the risk factor. Nowadays, SNPs associated with a risk factor can be obtained through large consortia genome-wide association studies (GWAS). The results reported by such studies, typically a list of univariate associations between SNPs and the risk factor, can be used to perform variable selection among a set of candidate SNPs. The JAM algorithm (Joint Analysis of Marginal summary statistics, Newcombe, Conti and Richardson, 2016) is a Bayesian model search algorithm recently proposed for this task. JAM uses a reversible-jump MCMC procedure to perform SNP selection based on GWAS summarized data. It accounts for genetic correlations and can be parallelized to analyse large numbers of SNPs simultaneously. We discuss the use of JAM for Mendelian randomization. In this context, we propose an extension that augments the JAM posterior with a loss function in order to penalize SNPs having a pleiotropic effect on the outcome. The performance of the new algorithm is illustrated with simulated and real data.

3523T

Accounting for confounding in Mendelian randomization using genome wide summary statistics. *J. Morrison, N. Knoblauch, J. Marcus, X. He, M. Stephens.* Human Genetics, University of Chicago, Chicago, IL.

Genome wide association studies (GWAS) have discovered trait associated variants for a wide array of phenotypes including clinically relevant traits such as disease status and intermediate molecular traits such as gene expression and metabolite levels. This wealth of data provides an opportunity to test hypotheses about the causal relationships between traits using Mendelian randomization (MR) and related methods. Current methods towards this goal rely on the assumption that pleiotropy is rare, that is, it is unusual for a single variant to effect multiple biological functions. However, recent research suggests this assumption is false. Many variants discovered via GWAS have strong signals of association with multiple, sometimes seemingly unrelated, traits while studies of gene regulation have uncovered variants that regulate the expression of many genes. A natural consequence of wide-spread pleiotropy is confounder driven genetic correlation. For example, if gene A is causally associated with disease susceptibility but gene A and gene B share some of their regulatory variants, gene B will appear to be "genetically correlated" with the disease. We propose a new method (Causal Analysis Using Summary Effect Estimates; CAUSE) that can accommodate multiple forms of pleiotropy including genetic variants acting on a shared confounder. We first apply our method to identifying relationships between pairs of complex traits. We demonstrate that CAUSE can recover known causal relationships (for example a causal effect of LDL cholesterol on coronary artery disease) and is able to identify several instances where correlation between traits is induced by substantial but incomplete sharing of genetic factors (for example between height and LDL cholesterol) leading to spurious signals using other methods. We also present an application to identifying genes with expression levels causally linked to Crohns disease. This analysis identifies both known Crohns genes and makes several new but plausible discoveries.

3524F

Leveraging genome-wide significant loci to increase the power of S-LDSC to detect enrichment. K. Tashman^{1,2}, A. Ganna^{1,2}, B.M. Neale^{1,2}, H.K. Finucane^{1,2}. 1) Stanley Center, Broad Institute, Cambridge, MA; 2) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Richard B. Simches Research Center Boston, MA.

Stratified LD Score Regression (S-LDSC; Finucane et al. 2015 Nat Genet) is a powerful tool to detect whether disease heritability is enriched in a functional annotation. S-LDSC is currently well-powered to detect enrichment for annotations with signal spread throughout the genome. However, it removes outlying single nucleotide polymorphisms (SNPs) such as those found in genome-wide significant loci, leaving potentially informative signal on the table. Here, we show that by explicitly modeling the signal found in top loci, we can increase the power of S-LDSC to detect enrichment. Specifically, we introduce a method for pathway analysis that combines S-LDSC with INRICH (Lee et al. 2011 Bioinformatics), a gene set enrichment tool that utilizes signal from top loci to test for enrichment of genes from a given pathway in a disease. To leverage signal from both S-LDSC and INRICH, we run both methods, excluding the loci used by INRICH when running S-LDSC, and we meta-analyze the p-values from the INRICH and S-LDSC results using Fisher's method. We applied our method in a pathway analysis of asthma using publicly available summary statistics from the UK Biobank and 6,166 pathways from KEGG and MSigDB provided with the INRICH software. While zero pathways passed FDR<0.05 with S-LDSC alone or with INRICH alone, our combined method identified 59 pathways with FDR<0.05, including gene sets as small as 5 genes; the identified pathways included many immune-related pathways and the top three most enriched pathways were ST-Interleukin-13, JAK-STAT signaling and Biocarta-TOB1. For this analysis, we created an open-source cloud-based pipeline that allows parallelized S-LDSC analyses. Utilizing this pipeline we were able to compute LD scores for 6,166 gene sets in less than two hours; we have released these LD scores publicly. Our work shows that incorporating signal from top loci can increase the power of S-LDSC. We plan to continue to evaluate and apply our method, and to extend it to enrichment of arbitrary functional annotations.

3525W

Robust method for inferring genetic relationship across studies without compromising privacy. X. Zhao¹, L. Franciolli^{2,3}, D. MacArthur^{2,3}, G. Abecasis¹. 1) Department of Biostatistics and Center for Statistical Genetics, University of Michigan School of Public Health, Ann Arbor, MI, USA; 2) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; 3) Broad Institute of MIT and Harvard, Cambridge, MA, USA.

Reference datasets of human genetic variation are critical resources for functional interpretation of putative disease-causing variants by, for example, helping separate genomic positions and regions that are mutation intolerant from others where variation is more common. Although variant summaries like allele frequencies and quality metrics are often shared, through resources like dbSNP and ClinVar, or custom browsers like gnomAD and BRAVO, the underlying individual level data is typically more restricted. Without this information, it can be difficult for researchers to evaluate overlap between different data sources when attempting to aggregate information across them. Here, we propose a privacy-preserving protocol for identifying duplicates and close genetic relatives across variation datasets. Our method does not require disclosure of individual identities when checking for overlap between two datasets. The method works by combining a cryptographic technique, homomorphic encryption, with the robust relationship inference method implemented in KING (Manichaikul et al., 2010). Homomorphic encryption allows addition and multiplication on encrypted data to get encrypted results, which, when decrypted, match the results of calculations performed on the original unencrypted data. By developing a simple algorithm that relies on these operations, we construct a framework under which kinship coefficients can be obtained without disclose any genetic information between studies. Through simulations, we show that our method successfully encrypts genetic data and decrypts kinship coefficient results. We were able to correctly infer all duplicates and almost all (99%) 1st degree relatives with 2% false positive pairs. Our performance is identical to using KING with the original unencrypted genotypes. Our protocol also scales well in computing time and is protected from possible attacks. Then we applied our technique to federate genetic variants from the Trans-Omics for Precision Medicine (TOPMed) web browser with those in the Genome Aggregation Dataset (gnomAD) browser. Between 62,784 TOPMed Freeze65k samples and 138,632 gnomAD samples, we identified 5,568 pairs of duplicates and 2,269 pairs of 1st degree relatives. Securely identifying related samples will control for relatedness in federation of reference datasets when they have data sharing barriers. It will help with providing a reliable large-scale resource for clinical interpretation of genetic variants.

3526T

Identifying associations between genetic variants and microbiome community diversity using kernel methods. *M.C. Wu.* Fred Hutchinson Cancer Research Center, Seattle, WA.

Understanding genetic influences on microbiome composition is important both because the microbiome can serve as an intermediate trait and because such knowledge enhances our understanding of how inter-individual differences in composition may be innate. Community level analysis, wherein overall microbiome profiles across all taxa is assessed for association with outcomes (beta diversity analysis), is a standard mode of analysis in microbiome studies due to its ability to detect subtle, concerted shifts in microbiome composition. However, little has been done on community level analysis within the context of genetic studies. Standard distance-permutation based approaches do not work well in genetic studies due to high dimensionality, computational expense of permutation, and poor capture of features of genetic data. Therefore, we propose a novel framework for community level analysis in relation to groups of genetic variants, e.g. in a region, gene, or pathway within the kernel machine testing framework. The kernel machine is commonly used already for analyzing genetic data (through the SKAT method) and has also been applied to community level analysis in microbiome studies (through the MiRKAT method). Exploiting existing work, we embed microbiome community composition in a kernel matrix which captures ecologically relevant structure. Similarly, genetic variants (or possibly a single variant) are embedded within a kernel matrix. Then association between the kernelized data is measured through a generalized measure of multivariate correlation called the kernel RV (KRV) coefficient. Using finite sample or asymptotic null distributions for the KRV, we can quickly and analytically calculate p-values for the association between composition and genetic variation. Novel statistical strategies are proposed to enable covariate adjustment and to accommodate related individuals. Simulations and data applications demonstrate that the proposed strategy correctly controls type I error while maintaining excellent power, even in the presence of confounders or genetic relatedness.

3527F

FinnGen: Towards a comprehensive catalogue of genomes and major health events for 500,000 Finnish residents. *J. Karjalainen^{1,2,3} on behalf of the FinnGen Consortium.* 1) Broad Institute, Cambridge, MA; 2) Massachusetts General Hospital, Boston, MA; 3) Institute for Molecular Medicine Finland, Helsinki, Finland.

FinnGen is a public-private partnership project that was initiated in August 2017 and aims to combine genome and national health register data from up to 500,000 Finnish individuals. It uses the unique resource of national health care registers that cover health information of every resident for their entire life span. Data freezes are produced in increments of 50,000 individuals every 6 months. At present, we have completed the first data freeze with 52,425 genotyped individuals. By the time of ASHG we will have doubled the sample to 100,000. Compared to other large scale biobank projects like the UK Biobank, FinnGen is enriched for several diseases including late onset diseases like Alzheimer's disease (n=1,459). We continue to prioritize sample collection from specialty clinics to increase the number of late onset diseases. We designed a custom genotyping array containing 837,899 markers. The common variant imputation backbone consist of 548,494 markers optimized for imputation. Additionally, to enable studying variants not captured well by imputation, we identified rare (< 0.5%) missense and protein truncating variants (PTV) from 19,778 exome-sequenced Finns and added all 60,093 PTVs and 137,996 non-singleton missense variants on the array. Chip genotyped individuals were further imputed with a population-specific high-coverage (>30x) WGS-based imputation reference panel, consisting of 3,775 individuals and 17M variants. The Finnish population isolate provides benefits to identify rare and low-frequency variants relevant to disease susceptibility. We performed phenome-wide association scans for 16.8 million imputed variants across 807 clinical endpoints (n cases>100) using 39,000 non-related Finnish participants. We used the logistic saddlepoint approximation method to control for false positives caused by case-control imbalance and low allele frequencies when using conventional regression methods. We identified 651 independent loci with genome-wide significant associations in 302 clinical endpoints. In PTV burden analysis we confirmed known associations with MYOC and glaucoma, and CHEK2 and all cancers combined (strongest associations to breast and prostate cancers). A meta-analysis with UKBB using the same phenotypes is ongoing. In addition, we developed web-based reporting and visualization tools to explore the association results together with previous genetic findings that include data from UKBB and the GWAS catalog, and pharmacological data.

3528W

Polygenicity varies with allele frequency and functional category due to negative selection. L.J. O'Connor^{1,2}, A. Schoech¹, F. Hormozdizadeh¹, S. Gazal¹, N. Patterson³, A.L. Price^{1,3}.

1) Department of Epidemiology, Harvard TH Chan School of Public Health, Boston, MA; 2) Program in Bioinformatics and Integrative Genomics, Harvard Graduate School of Arts and Sciences, Cambridge, MA; 3) Program in Medical and Population Genetics, Broad Institute, Cambridge.

Common diseases and complex traits are highly polygenic. This polygenicity could reflect negative selection, if large-effect SNPs are prevented from becoming common in the population; we refer to this phenomenon as *flattening*. Here, we quantify the effects of flattening by comparing polygenicity across allele frequencies and functional annotations. We introduce a new definition of polygenicity, the *effective number of associated SNPs* (M_e); in the special case of no LD, M_e is inversely proportional to the kurtosis of causal effects, and under a point-normal model it equals the number of causal SNPs. We introduce a method, stratified LD fourth moment regression (S-LD4M), to estimate M_e for subsets of SNPs by regressing squared chi-square statistics on LD fourth moments (sums of r_i^4). We confirmed that S-LD4M produces robust estimates of M_e in realistic simulations. We applied S-LD4M to 32 diseases and complex traits (avg N=403k). Polygenicity varies widely across traits (mean $\log_{10} M_e = 3.36$, s.d. 0.64); brain-related traits are especially polygenic, e.g. $\log_{10} M_e = 4.28$ (s.e. 0.06) for schizophrenia. Approximately 25% of sparsity (excess kurtosis) was explained by functional annotations in the baseline-LD model (Gazal et al. 2017 Nat Genet). Comparing the polygenicity (per-SNP M_e) of common (MAF>5%) vs. low-frequency (MAF>0.5%) SNPs, we determined that common SNPs are 4.2 (s.e. 0.7) times more polygenic than low-frequency SNPs. We compared polygenicity across functional categories, determining that polygenicity enrichment is very nearly equal ($r^2=0.95$) to heritability enrichment; for example, SNPs in conserved regions are approximately 13x enriched in units of both h^2 and M_e . This concordance implies that heritability enrichment predominantly reflects differences in the number of associations rather than their magnitude, probably due to a selection threshold. It suggests that significant GWAS loci represent a mixture of SNPs with deleterious effects on weakly selected genes and SNPs with subtle regulatory effects on strongly selected genes; though the former are more tractable, the latter may provide more insight. Indeed, among fine-mapped IBD GWAS loci (Huang et al. 2017 Nature), 13/30 putative causal genes implicated by fine-mapped regulatory variants were LoF intolerant, vs. 0/6 genes implicated by coding variants ($p=0.003$). Our findings underscore the profound effect of negative selection on the genetic architecture of complex traits.

3529T

A robust unified test for Hardy-Weinberg equilibrium in arbitrarily structured populations. A. Kwong, H.M. Kang, *The Trans-Omics for Precision Medicine Program*. University of Michigan, Ann Arbor, MI.

Hardy-Weinberg Equilibrium (HWE) has long been an important metric to evaluate the quality of variant calls. Deviations from HWE can be caused by many factors, including population structure and genotyping error. For samples arising from similar ancestries, in the absence of population structure, the chi-squared and exact tests for HWE are routinely used in variant quality control. To account for population structure for samples from discrete, non-overlapping populations in HWE tests, subpopulations can first be tested separately, then test statistics can be combined across subpopulations using Stouffer's method. When a set of samples arise from a more continuous spectrum of ancestries, then these traditional tests are unable to control Type I errors. As genetic studies contain samples of increasing diversity and admixture, this will become an increasingly prevalent problem for variant quality control. To address these issues, we propose the Robust Unified Test for Hardy-Weinberg Equilibrium, which uses both a score test and a likelihood ratio test with individual-level adjustments for ancestry. Our method allows the joint computation of ancestry-adjusted HWE statistics, without the need to split samples into stratified subpopulations, by using individual-specific allele frequencies for each sample at each variant, based on ancestry summary statistics. We tested our method against the chi-square test, exact test, and Stouffer's method in both simulated data and variants from the Trans-Omics Precision Medicine (TOPMed) Program, and show that when samples contain population structure, our method was better at controlling Type I errors than the chi-square and exact tests, with more power to detect deviation from HWE after adjusting for population structure when compared to Stouffer's method. For example, when testing 338,722 low-quality variants with high Mendelian discordance in chromosome 20 in TOPMed, our method found 196,962 variants to significantly deviate from HWE at $p < 10^{-6}$, compared to 114,381 variants using the exact test. When testing 18,380 high-quality variants present in both the Omni2.5 array and HapMap3 site lists, our method found only 325 significant variants, compared to 6,754 using the exact test. In summary, we have developed a powerful and flexible method well-suited for variant quality control in modern genetic studies with samples arising from diverse ancestries.

3530F

Spatially-dense 3D facial heritability and genetic correlations in a heterogeneous sample.

J. Li^{1,2}, J.D. White³, K. Indencleef⁴, H. Hoskens^{1,5}, A. Ortega Castrillon^{1,2}, A. Zaidi⁶, R.J. Eller⁶, S. Walsh⁶, M.D. Shriver³, P. Claes^{1,2,7}.

1) Medical Imaging Research Center, MIRC, University Hospitals Leuven, Leuven, Belgium; 2) Department of Electrical Engineering, ESAT/PSI, KU Leuven, Leuven, Belgium; 3) Department of Anthropology, Penn State University, University Park, Pennsylvania, USA; 4) Department of Neurosciences, Experimental Otorhinolaryngology, KU Leuven, Leuven, Belgium; 5) Department of Human Genetics, KU Leuven, Leuven, Belgium; 6) Department of Biology, Indiana University-Purdue University Indianapolis, Indianapolis, IN, USA; 7) Murdoch Childrens Research Institute, Melbourne, Victoria, Australia.

The human face is a complex trait with a strong genetic component. Though facial heritability and the genetic correlation between different facial features have been estimated in numerous studies, most of the methods used rely on a strong assumption of population homogeneity in either a family-based or a population-based designs. We propose a mixed linear model (MLM) to estimate the genetic contributions to facial variation in a sample composed of a heterogeneous combination of populations. The additive genetic variation is decomposed into fixed and random effect components, consisting of the multivariate genomic background (PCA) and the pairwise genomic relationships between individuals adjusted for population heterogeneity and admixture, respectively. Rather than using population allele frequencies, the adjusted genomic relationships are obtained using individual-specific allele frequencies that are estimated from a logistic regression between genotypes and the multivariate genomic background. Finally, the narrow-sense heritability is defined in terms of the proportion of variance explained by the two components. We present the narrow-sense heritability of facial shape calculated from 3D facial images and genotype data of 2,882 unrelated individuals of varying ancestries. First, we estimated the narrow-sense heritability of univariate pairwise distances between 7,160 quasi-landmarks representing facial shape in a spatially-dense arrangement. The highest heritability was observed for the distance between the nasion and the zygion, 65.07%. In addition, we estimated the narrow-sense heritability and the pairwise genetic correlations of all quasi-landmarks as multivariate three-dimensional variables. The narrow-sense heritability of multivariate quasi-landmarks ranged from 13.74% (around the chelion) to 54.05% (the chin). Finally, we performed a hierarchical segmentation of facial shape based on the genetic correlation between quasi-landmarks, grouping them into modules that are internally well-integrated. The resulting genetic correlation-driven facial segmentation shows remarkable overlap with the facial prominences in early embryonic development: frontonasal prominence, maxillary prominence, and mandibular prominence. Our results suggest a promising future for genetic studies of facial shape phenotypes learned from genetic correlations in a heterogeneous sample.

3531W

Accounting for sub-phenotyping in rare variant association tests.

E. Genin¹, O. Bocher¹, A. Saint Pierre¹, H. Perdry², G. Marenne¹. 1) UMR1078, Inserm, Université de Brest, EFS, CHU Brest, Brest, France; 2) UMR669, Inserm, Université Paris Sud, Paris, France.

Technological advances have made genome sequencing a reality and opened up new possibilities to explore genetic disease architecture and test the role of rare variants in common diseases. Investigators have access to the integrality of the genetic information and the challenge now is to understand the phenotypic consequences of these different genetic variants. For that purpose, it could be of interest to go beyond the classical binary disease model of cases and controls and refine the trait under study by sub-phenotyping to take into account for example disease severity, age of onset or any other relevant information that could be important in the disease etiology. Indeed, most common human diseases show clinical and phenotypic heterogeneity that could underlie genetic heterogeneity and it might thus improve power of association tests to identify relevant patient subgroups. Many tests have been proposed for rare variant association testing, either with quantitative phenotypes or with binary phenotypes but there is no test for rare variant association that allows a sub-grouping of patients. In this work, we propose extensions of rare variant burden tests that address this issue of sub-phenotyping. We have developed a multinomial logistic extension of two popular rare variant association tests, CAST and WSS and implemented them in an R package. Through extensive simulations, we show that accounting for phenotype heterogeneity of patients could largely improve the power of rare variant association tests, and, as expected, this is especially true under scenarios where genetic heterogeneity between subgroups is important. Interestingly, in the absence of genetic heterogeneity, we find that the power loss from separating cases into subgroups is rather limited and the cost/benefit balance is strongly in favor of using information on sub-phenotypes. We also illustrate this advantage on some real data examples where we show the benefit of sub-phenotyping cases. The tests developed and the simulation framework are implemented in an R package available on Github. This package makes it possible for investigators to run the tests on their data but also to perform simulations under different scenarios using information from GnomAD or their own information on rare variants found in different genes.

3532T

A survey of integrating age-at-onset genetics for predicting the age-specific disease risk. C. Tian, D. Hinds, 23andMe Research Team. 23andMe, Mountain View, CA.

Human disease is characterized by marked genetic heterogeneity, which has important implications for gene discovery. In particular, evidence suggests distinct genetic susceptibility between early-onset and late-onset diseases. In this study, we demonstrated that stratifying on age-at-onset information collected in the case-control studies could lead to the discovery of novel genetic risk factor so-called modifiers of age-at-onset, despite reduced sample size. For example, two distinct missense variants in *SERPINA1* were associated with high cholesterol in young adults and elderly people, respectively, while variants in *HDAC5* and *CTNND2* were significantly associated with high cholesterol in young adults only. This degree of heterogeneity on age-at-onset also has important implications for development of personalized genetic assessment. As a result, we should include modifiers of age-at-onset into the predictive modeling to increase predicting accuracy, especially for earlier onset diseases. Currently, genetic risk scores (GRS), derived mainly from case-control GWAS, have been used for predictive modeling of many complex disease risks. This approach is considered clinically suboptimal for assessing an age-dependent disease where a subset of "controls" will become "case" over time. Scientists are trying to extend the work by integrating genetics within a survival-analysis-epidemiology framework to derive a polygenic hazard score (PHS) and hope to use it to inform both whether and when to order screening tests. Some preliminary work has been done for Alzheimer's disease and prostate cancer. Here we carefully investigated the Cox survival model for age-specific risk prediction and provided some practical guidance on incorporating age-at-onset genetics into disease risk prediction. When using the survival model, one thing to consider is that the baseline hazard estimates derived from GWAS samples cannot be used directly, especially for senior individuals. We evaluated methods for deriving population-based incidence rate estimates. Previous study showed that simple inclusion or exclusion of future cases in each risk set induced an under- or over-estimation bias in the regression parameters, respectively. A weighted Cox model that weights subjects according to age-conditional probabilities of developing the disease in the source population may provide less biased estimation.

3533F

Importance of protected statistical testing and replicate number in RNA-sequencing studies. R.Z. Blumhagen¹, C.M. Moore¹, B.E. Vestal¹, C.D. Langefeld², T.E. Fingerlin¹. 1) Center for Genes, Environment and Health, National Jewish Health, Denver, CO; 2) Wake Forest School of Medicine, Medical Center Boulevard, Winston-Salem, N.C.

When more than two groups are of interest in RNA-Seq studies, it is common practice to perform all pairwise comparisons between groups to identify differentially expressed genes. Using this approach without proper adjustment for multiple comparisons increases false positive associations. As RNA-Seq studies become more complex, including multiple groups, conditions and time points, the number of pairwise tests grows. To reduce false positive associations, we examine the use of a likelihood ratio test (LRT) as the first step before evaluation of the pairwise comparisons, consistent with other protected testing approaches. We first give results from a simulation study comparing the statistical testing properties of the LRT gate-keeping approach to the pairwise comparison (PC) approach, where a gene is called differentially expressed if at least one pairwise test is significant. We then apply the two methods to real datasets. We simulated RNA-Seq data using a negative binomial model with mean and variance distributions derived from a clinical dataset with a range of parameters for group number, number of replicates per group and proportions of differentially expressed genes. As expected, PC had higher type 1 error and false discovery rates (FDR) than the LRT in all scenarios. For example, in a simulation including 7 groups with 5 replicates each and $\alpha=0.001$, PC had 50.1 times (95% CI: 41.5-61.7) the nominal type 1 error compared to 10.0 times for the LRT (95% CI: 5.2-15.3). Under the alternative of 10% differentially expressed genes with a 2-fold change, PC had 4.6 times (95% CI: 4.2-5.0) the expected FDR compared to 2.4 times for the LRT (95% CI: 1.9-2.8). Despite having lower FDR than the PC approach, the LRT also exhibited inflated FDR for small numbers of replicates. The inflation worsened as the number of study groups increased, illustrating the importance of replicate number within each experimental group. In a simulation including 5 groups, increasing the replicates from 3 to 5 in each group reduced the FDR inflation from 3.3 times to 2 times. In summary, when comparing RNA-Seq expression between more than two groups, the LRT reduces false positives and better controls FDR compared to evaluating all pairwise comparisons. This approach is also more powerful than using a correction for the number of pairwise tests. Further, we recommend careful consideration of the number of replicates within each group to reduce false discoveries.

3534W

Classification of cell type from single-cell RNA sequencing data by deep neural networks. C. Lou, G. Hu, W. Chen, D. Stambolian, M. Li. University of Pennsylvania, Philadelphia, PA.

Recent advances in single-cell RNA-seq (scRNA-seq) have made it possible to systematically characterize cell types by gene expression. Knowledge of cell types is important in the pursuit of precision medicine. Current approaches for cell type classification often perform poorly when the differences of cell types are subtle. As scRNA-seq becomes increasingly popular, more and more scRNA-seq data have been generated. It is appealing to incorporate these data to inform cell type classification when new data become available. Supervised learning using neural networks is suited perfectly for this purpose. To evaluate the performance of this approach, we focus on 9,515 bipolar cells generated from mice retina using 10x Genomics platform. We choose bipolar cells due to their large number of subtypes, and the availability of well characterized subtypes from a previously published study (Shekhar et al. 2016), which includes ~24,000 bipolar cells generated using DropSeq. Using the Shekhar data as a gold standard, we classify our 9,515 bipolar cells using two different methods. The first approach is a semi-supervised clustering method, which involves aligning data from our bipolar cells with the Shekhar data and then clustering using the Louvain-Jaccard algorithm to find clusters that correspond to those identified by Shekhar. The second approach is supervised learning using neural networks in which the model parameters are estimated using the Shekhar data as the training dataset. The trained neural network is then applied to our bipolar cells to assign cell types. With the semi-supervised clustering, we identify 15 different bipolar cell subtypes whose gene expression profiles correspond to those in Shekhar. Using neural networks with one hidden layer and 100 nodes, we also identify 15 different bipolar cell subtypes, of which 77% have concordant cell type assignments with the semi-supervised clustering method. When increasing the number of hidden layers and the number of nodes per layer, the concordance rate increases further. Since the two scRNA-seq datasets were generated using different platforms, our results suggest that deep neural networks are a robust approach for supervised learning of cell type classification. We are currently applying this approach to a larger scRNA-seq dataset that involves ~100K single cells from mouse retina. We anticipate that this analysis will yield a comprehensive cellular map for bipolar cells in the mouse retina.

3535T

Robust adjusted profile likelihood ratios for the analysis of differential gene expression. L. Zhong^{1,2}, G. He^{1,2}, L.J. Strug^{1,2}. 1) Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada; 2) Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada.

RNA Sequencing (RNA-Seq) experiments generate read counts that provide a measure of gene expression. RNA-Seq can then be used, for example, to determine which genes show differences in expression in Cystic Fibrosis patients with low versus high lung function. Popular approaches for differential gene expression analysis include edgeR and DESeq2. Although count data is routinely modeled using a Poisson distribution, the RNA-Seq data does not in general satisfy the Poisson distributional assumption of equal mean-variance. EdgeR and DESeq2 therefore assume a negative binomial distribution for the read counts to model the overdispersed mean-variance relationship for biological replicates and estimate the gene-specific dispersion level for the downstream testing procedure. The negative binomial distribution assumes a quadratic variance function, yet more general patterns for over/under-dispersion could be observed, potentially missing important genes. We propose a general solution to measure gene expression differences for any variance pattern using robust adjusted profile likelihood inference. Poisson regression is used to model the relative abundance, with the estimated effective library size used as an offset term for normalization. Simulation under a variety of scenarios demonstrates that the profile likelihood ratio for the differential gene expression parameter, after a data-driven robust adjustment for departures from the Poisson mean-variance assumption, can reliably detect which genes display differential expression. Findings are compared to analyses using conventional approaches. In RNA-Seq of nasal epithelia from 64 patients with Cystic Fibrosis, we identify 15 differentially expressed genes associated with lung function, adjusted for study site, gender and immune cell contamination. We provide a flexible approach to gene expression analysis that can accommodate over or under dispersion, any mean-variance relationship, is applicable to categorical or quantitative conditions of interest and complex experimental designs. Support intervals and novel graphics provide accessible genome-wide interpretations.

3536F

Deconvolution of bulk tissue cell type proportions from single-cell RNA-seq. *X. Wang, J. Park, K. Susztak, M. Li, N. Zhang.* University of Pennsylvania, Philadelphia, PA.

The characterization of cell types within intact bulk tissue and the estimation of their proportions play an important role in human disease studies. Many clinical and population-based transcriptomic studies use bulk RNA sequencing (RNA-seq) from intact tissues, where the average gene expression across cells in the tissue is measured. Since many tissues are heterogeneous at the cellular level, failure to control for cell type differences may mask the underlying cell-type specific contributions to disease, and sometimes, may even produce false positive results due to confounding from cell type proportions. Deconvolution of cell type proportions from bulk tissue gene expression data, is difficult, especially when the underlying cell types are unknown. Recent advances in single-cell RNA sequencing (scRNA-seq) has paved the way for exploring cellular heterogeneity and allows for the identification of novel cell types and better characterization of known cell types. However, scRNA-seq is still costly, and furthermore, is not well suited for the quantification of cell type proportions in a tissue, because the cell dissociation and isolation steps of the scRNA-seq procedure are biased, leading to over-representation or under-representation of certain cell types. Although progress has been made on the bulk RNA-seq deconvolution problem, all existing methods ignore the cross-subject heterogeneity in gene expression and rely heavily on pre-selected marker genes. Our analysis of multiple scRNA-seq datasets suggests that cross-subject heterogeneity is not ignorable. To overcome these limitations, here we present MuSiC (Multi-Subject Single Cell deconvolution), to estimate cell type proportions with a data-driven scheme to automatically weigh genes rather than pre-selecting markers with arbitrary cutoff. The weight is determined by the cross-subject cell-type specific gene expression variance as well as the technical noise of the bulk RNA-seq data. MuSiC clearly outperforms known methods in several real datasets. Comparing cell type compositions between normal and type 2 diabetes (T2D) individuals, MuSiC detected significant negative correlation between T2D and beta cell proportions, which is missed by other existing methods. MuSiC can successfully detect rare cell types in mouse kidney by hierarchical structure, which grouped similar cell types together and find signals of chronic kidney disease missed by other methods.

3537W

Impact of genotype imputation on genome-wide meta-analysis: To aggregate or to eliminate? *E.F. Acar.* The Centre for Phenogenomics, The Hospital for Sick Children, Toronto, Ontario, Canada.

Genotype imputation is a technique extensively used in genome-wide association studies (GWAS), as well as in their meta-analysis. The impact of imputation on testing genetic associations in a single GWAS has been addressed in the literature and a number of imputation-based test procedures have been proposed. However, the consequences of including imputation-based GWAS results in a meta-analysis remain largely unexplored. In this work, we consider both fixed and random effects models to evaluate the accuracy and efficiency of imputation-based meta-analysis results under different levels of genotype imputation accuracy. Simulation results reaffirm that meta-analysis boosts the power of detecting genetic associations compared to individual study results. However, the power deteriorates with increasing uncertainty in imputed genotypes. To control the contribution of imputation-based studies, we propose a reweighing scheme based on genotype imputation accuracy. Our approach achieves a better detection power relative to the traditional approaches, and improve the validity and reliability of imputation-based meta-analysis results.

3538T

A method to estimate the sampling variance of genotype principal components and residual confounding due to incomplete capture of population structure. C. Chen^{1,2}, D. Palmer^{1,2}, J. Bloom^{1,2}, D. Chen^{1,2}, K. Tashman^{1,2}, C. Churchhouse^{1,2}, B. Neale^{1,2}, A. Bloemendal^{1,2}. 1) Medical and Population Genetics Program and Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA; 2) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA.

Principal component analysis (PCA) is widely applied to genotype data to identify population structure and control for its confounding effect in genome-wide association studies (GWAS). Patterson, Price & Reich (PLoS Genet 2006) applied results from random matrix theory to show that in a structured sample, for population differentiation below a threshold depending on the sample size and number of markers, PCA fails to detect the signal. Real data with linkage disequilibrium require an estimate of the effective number of independent markers, for which this work includes a moments-based estimator. We observe the latter to be highly unstable and often poorly calibrated, and propose an alternative robust estimator based on the eigenvalues of the genetic relationship matrix. The threshold phenomenon for each principal component (PC) depends on its corresponding eigenvalue. We extend previous work to fully characterize the eigenvalues for population structures including clines and admixture, generalizing the models of Balding & Nichols (Genetica 1995) and Pritchard, Stephens & Donnelly (Genetics 2000). We establish that the eigenvalues depend only on the population differentiation parameters and the covariance of the admixture distribution. Further, for each eigenvalue above the threshold, we combine our robust estimator with existing theory to estimate the magnitude of differentiation the corresponding PC attempts to capture. Below the signal-detection threshold, a PC is essentially uncorrelated with the true direction of differentiation; even above the threshold, however, the correlation may be only partial. We estimate this correlation, also in terms of the corresponding eigenvalue. Finally, we predict the correlation for other samples from the same population with the same markers, yielding a critical sample size below which the correlation drops to zero. We validate all these estimators through extensive simulations and downsampling experiments on multiple datasets. Our results imply that using sample PCs as covariates in GWAS may incompletely control for population stratification. We investigate the degree to which study results may be affected, considering both binary and continuous traits under various stratification scenarios. We provide a straightforward software implementation of our method together with some practical recommendations.

3539F

General retrospective mega-analysis framework for rare variant association tests. Y. Chiu¹, L. Chien². 1) National Health Research Institutes, Zhunan, Miaoli 35053, Taiwan; 2) Center for Fundamental Science, Kaohsiung Medical University, Kaohsiung 80708, Taiwan.

We proposed a retrospective mega-analysis framework for gene- or region-based multimarker rare variant association tests. Our proposed mega-analysis association tests allow investigators to combine longitudinal and cross-sectional family- and/or population-based studies. This framework can be applied to a continuous, categorical or survival trait. In addition to autosomal variants, the tests can be applied to conduct mega-analyses on X-chromosome variants. Tests were built on study-specific region- or gene-level quasi-score statistics and, therefore, do not require estimates of effects of individual rare variants. We used the generalized estimating equation approach to account for complex multiple correlation structures between family members, repeated measurements, and genetic markers. While accounting for multilevel correlations and heterogeneity across studies, the test statistics were computationally efficient and feasible for large-scale sequencing studies. The retrospective aspect of association tests helps alleviate bias due to phenotype-related sampling and type I errors due to misspecification of phenotypic distribution. We evaluated our developed mega-analysis methods through comprehensive simulations with varying sample sizes, covariates, population stratification structures, and study designs across multiple studies. To illustrate application of the proposed framework, we conducted a mega-association analysis combining a longitudinal family study and a cross-sectional case-control study from Genetic Analysis Workshop 19.

3540W

A flexible linear mixed model for estimating heritability of categorical phenotypes. G. Darnell¹, N. Mukherjee², S. Mukherjee^{2,3}, B. Engelhardt⁴. 1) Lewis-Sigler Institute, Princeton University, Princeton, NJ; 2) Department of Statistics, Duke University, Durham, NJ; 3) Department of Mathematics, Computer Science, Duke University, Durham, NJ; 4) Department of Computer Science, Center for Machine Learning and Statistics, Princeton University, Princeton, NJ.

Heritability of human complex traits such as disease risk have long been estimated using models based on principles of quantitative genetics. However, many traits like age-related macular degeneration, schizophrenia, and autism cannot be naturally represented on a continuous scale or with binary values, and have complex diagnosis patterns that are typically represented with a category-based system. Even traits that are inherently binary and have a monogenic basis can trigger an effect on related non-binary traits through variation in population gene expression levels [Barbeira, 2018]. It is crucial for understanding the genetic architecture of complex traits to estimate genetic correlations among traits that are represented on heterogeneous scales. Many sources of data that include important signal but were previously underused, including responses to survey questionnaires such as that of the UK Biobank, are categorical in nature. We developed a flexible probabilistic framework that estimates heritability of, and the degree of pleiotropy between, categorical traits that does not require any recoding of the phenotype to a continuous, ordinal, or binary scale. Our model is based upon the multivariate generalized linear mixed model, and extends this framework further to be able to partition variance in categorical traits. Simulated trait studies show that our model is highly accurate and robust in recovering heritability and pleiotropy relative to current approximations of these estimates of heritability. We apply our framework to the UK Biobank data, which contains categorical survey responses and disease stages, covariates such as age, sex, and BMI, and genetic marker information in order to estimate the underlying heritability of categorical traits. In these data, we gain insights into the degree to which complex traits have a common genetic basis, and we can begin to uncover the pleiotropic effects of regulatory mechanisms.

3541T

Sparse canonical correlation analysis (sCCA) significantly improves power of cross-tissue transcriptome-wide association studies (TWAS). H. Feng^{1,2}, B. Pasaniuc^{3,4}, M. Major^{3,4}, P. Kraft^{1,2}. 1) Program in Genetic Epidemiology and Statistical Genetics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, United States of America; 2) Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, United States of America; 3) Department of Human Genetics, University of California Los Angeles, Los Angeles, California, United States of America; 4) Department of Pathology and Laboratory Medicine, University of California Los Angeles, Los Angeles, California, United States of America.

TWAS test the association between imputed gene expression and an outcome, which relies on accurate estimation of genetic effects on gene expression and suffers from low power at small eQTL sample sizes or when data from relevant tissues is not available. Genetic regulation of expression is similar across tissues for many genes, therefore cross-tissue TWAS tackles these issues by leveraging eQTL data across multiple tissues. Here, we use sCCA to integrate data across tissues and build a new feature (i.e. a cross-tissue weighted average of expression values with high genetic correlation, which can be used to identify heritable genes and as an intermediate trait in TWAS analyses) and show that sCCA-TWAS outperforms traditional TWAS. We simulated gene expression data from 22 tissues using empirical cross-tissue correlation patterns from GTEx, assuming different genetic correlation patterns across tissues, and simulate GWAS summary statistics assuming expression levels in a single tissue are associated with outcome, and compared sCCA-TWAS to single-tissue TWAS; cross-tissue TWAS adjusting for multiple testing with Bonferroni and Generalized Berk-Jones (GBJ); and summarizing cross-tissue expression patterns using Principal Component Analysis (PCA). All methods control the Type I error both when there is no genetic effect on expression or when expression is not associated with outcome. When expression drives trait, sCCA-TWAS had the greatest power to detect a gene associated with outcome, even when the causal expression trait was not directly measured: e.g. when gene expression explains 2% of the variability in outcome and the GWAS sample size is 20,000, the average power difference between sCCA and single-tissue, multi-tissue, and PCA approaches was 0.23, 0.26, and 0.40. The large gain in power is likely due to sCCA cross-tissue features being more likely to be significantly heritable than single-tissue or PC expression features (2.78x, and 3.72x respectively). Our results suggest that aggregating eQTL data across multiple tissues using sCCA can improve the specificity and sensitivity of TWAS. We also present a comparison of these approaches applied to multiple large, publicly available GWAS of complex traits.

3542F

Improving gene expression prediction accuracy in transcriptome-wide association studies. *J. Fryett¹, A.P. Morris², H.J. Cordell¹.* 1) Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, United Kingdom; 2) Department of Biostatistics, University of Liverpool, Liverpool, United Kingdom.

Transcriptome-wide association studies (TWAS) are a popular approach for integrating genome-wide association study (GWAS) data and gene expression data to investigate the role of gene expression in complex traits. In this method, gene expression values (attributable to genetic regulation) are first predicted from SNP genotype data, and then tested for association with a trait of interest. The power of this approach to identify expression-trait associations depends partially on gene expression prediction accuracy. Here, we use genotype and blood expression data from the Geuvadis project with 10-fold nested cross-validation to examine and compare the ability of seven statistical approaches (Ridge regression, LASSO, Elastic net with alpha pre-set to 0.5, Elastic net with alpha determined by cross-validation, a Bayesian Sparse Linear Mixed Model (BSLMM), the Best Linear Unbiased Predictor (BLUP) and Random Forest) to predict gene expression from SNP data. Overall, we find that expression of most genes cannot be predicted accurately using SNP genotype data. However, approximately 2% of genes showed reasonable prediction accuracy (cross-validation $R^2 > 0.3$), representing a set of genes where the TWAS approach may have power to detect associations. We also observe that the sparse statistical models (LASSO, Elastic net, BSLMM) predict expression marginally more accurately than polygenic models. Through an application to GWAS data for type 1 diabetes from the Wellcome Trust Case Control Consortium, we find that all seven methods perform similarly at detecting associations. We also investigate the effects of sample size, ancestry and tissue specificity on prediction accuracy. We find that prediction accuracy is consistently reduced when fewer samples are used to train the gene expression prediction models, or when populations of different ancestries are used to train the prediction models and to test them. Finally, by training gene expression prediction models using data from developing brains, we find that prediction accuracy is reduced when the tissues used to train and to test the models are different. We conclude that TWAS prediction accuracy can be improved through the use of sparse statistical models, larger sample sizes, and by using data of appropriate ancestry and tissue. This may lead to improved ability to detect associations with the TWAS approach.

3543W

Mapping multiple complex and omics trait-associations using summary statistics from correlated or independent samples. *K.J. Gleason¹, F. Yang², L.S. Chen¹.* 1) Dept. of Public Health Sciences, The University of Chicago, Chicago, IL; 2) Dept. of Biostatistics and Informatics, Colorado School of Public Health, Aurora, CO.

To date, genome-wide association studies (GWAS) have identified more than 60 thousand unique SNP-trait associations, but the functional mechanisms underlying many of these associations remain unknown. The analysis of the joint associations of a SNP to complex trait(s) and omics-phenotypes (e.g., mRNA expression, DNA methylation, and protein abundance) has the potential to elucidate mechanisms underlying known associations or to reveal novel relationships between genetic variants and complex traits. In this work, we propose an integrative genomics approach for mapping multi-trait associations using summary statistics from correlated, overlapping or independent samples, and develop an R package: "primo." Different from existing literature, primo can analyze the joint associations of more than three sets of summary statistics. Primo estimates the probability that a SNP is associated with any arbitrary combination (e.g. at least 1, at least 2, some but not others, or all) of traits and omics-phenotypes of interest. Empirical false discovery rates are estimated and controlled through permutation of test statistics. We first applied primo to identify multi-omics QTLs in tumor samples from TCGA in breast tissues, and ovary tissues, separately. We joint analyzed the eQTL, meQTL, and pQTL summary statistics, and have identified multiple genes with e-me-pQTLs, several of which were known to be cancer-related. We further integrated the three breast tissue QTL sets with summary statistics from a breast cancer GWAS, and identified several breast-cancer-risk-associated SNPs that were also multi-omics QTLs. Those results illustrated that some SNPs, in particular those in non-coding regions, may affect cancer risks and other complex traits via certain omics phenotypes. In another integrative analysis by primo, we estimated broad sharing (~73%) of eQTLs between tumor breast tissues from TCGA and normal breast tissues from GTEx. Similar sharing (~78%) was estimated in the analysis of ovary tissues. Finally, we used primo to assess pleiotropy by integrating molecular QTL summary statistics with results from GWAS of multiple complex traits. SNPs identified by primo to be associated with multiple diseases and omics phenotypes highlight possibly shared mechanisms that underlie disease processes. The package is freely available for download at <https://github.com/kjgleason/primo>.

3544T

A semi-supervised approach for predicting cell type/tissue specific functional consequences of non-coding variation using massively parallel reporter assays, with applications to complex trait genetics. Z. He¹, L. Liu², K. Wang³, I. Ionita-Laza². 1) Dept of Neurology, Stanford University, Stanford, CA; 2) Dept of Biostatistics, Columbia University, New York, NY; 3) Dept of Statistics, Columbia University, New York, NY; 4) Raymond G. Perleman Center for Cellular and Molecular Therapeutics, Children's Hospital of Philadelphia, PA.

Predicting the functional consequences of genetic variants is a challenging problem, especially for variants in non-coding regions. Projects such as ENCODE and Roadmap Epigenomics make available various epigenetic features, including histone modifications and chromatin accessibility, genome-wide in over a hundred different tissues and cell types. In addition, recent developments in high-throughput assays to assess the functional impact of variants in regulatory regions (e.g. massively parallel reporter assays, CRISPR/Cas9-mediated in situ saturating mutagenesis) can lead to the generation of high quality data on the functional effects of selected variants. We propose here a semi-supervised approach, GenoNet, to jointly utilize experimentally confirmed regulatory variants (labeled variants), millions of unlabeled variants genome-wide, and more than a thousand cell type/tissue specific epigenetic annotations to predict functional consequences of non-coding genetic variants. Through the application to several experimental datasets, we demonstrate that the proposed method significantly improves prediction accuracy compared to existing functional prediction methods at the tissue/cell type level, but especially so at the organism level. Importantly, we illustrate how the GenoNet scores can help in fine-mapping at GWAS loci using as example a leading schizophrenia risk locus flanking *MIR137*, and in the discovery of disease associated genes through an integrative analysis of lipid phenotypes using a MetaboChIP dataset on 12,281 individuals. The ability to infer cell-type specific functional elements that regulate target genes will greatly enhance our understanding of what genes contribute to cellular identity and cell lineage differentiation, and better prioritize non-coding variants from GWAS or NGS in the contexts of the cell types that are relevant to the disease under study. It also allows the design of functional genomics experiments that target specific subsets of non-coding regions based on the cell type under study. As more systematic and comprehensive lists of experimentally validated variants become available across a large number of tissues and cell types over the next few years, semi-supervised methods like GenoNet can be used to provide increasingly accurate functional predictions for variants genome-wide and across a variety of tissues and cell types.

3545F

A universal and nearly optimal permutation testing approach and an application to association analysis of quantitative traits in whole-genome sequencing studies. J. Hecker¹, I. Ruczinski², M. Cho^{3,6}, H. Loehlein Fier⁴, D. Prokopenko^{5,6}, E. Silverman³, B. Coull¹, C. Lange^{1,3}. 1) Harvard T. H. Chan School of Public Health, Boston, MA; 2) Bloomberg School of Public Health, Baltimore, MD; 3) Brigham and Women's Hospital, Boston, MA; 4) University of Bonn, Bonn, Germany; 5) Massachusetts General Hospital, Boston, MA; 6) Harvard Medical School, Boston, MA.

While permutation-based approaches are ideal tools to address the statistical issues related to association testing in whole-genome sequencing (WGS) and other large-scale genomic studies (e.g. potential violations of asymptotic distribution assumptions, non-normality of phenotypes, design imbalances), the actual application of such approaches is usually prohibitive due to the computational burden or the restriction to specific scenarios and test statistics. Whereas current approaches use approximations of the permutation distribution or are based on adaptive heuristic criteria to reduce the number of permutations that are computed, we propose here a framework for permutation testing that is based on sequential testing theory and directly tests the permutation-based p-value against a pre-specified significance level. This allows for the rigorous control of both error probabilities, type 1 and 2 error, and is nearly optimal in terms of the number of expected permutations. Our framework can be applied to arbitrary association test statistics for which permutation testing can be implemented, e.g. association tests for single variants or regions in WGS studies. In an application to a WGS study for a quantitative trait of lung function, we illustrate the performance of our approach and its practical implementation for single variant analysis. Given the moderate, additional computational burden, permutation-based testing might be considered the first choice of analysis in many large-scale genetic/genomic studies where the limitations of asymptotic theory could invalidate the findings of the statistical analysis.

3546W

Comparison of Mendelian randomisation and Bayesian network approaches for causal inference. *R. Howey, H.J. Cordell.* Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, Tyne and Wear, United Kingdom.

Mendelian randomisation (MR) is a popular method for inferring the causal relationships between variables in genetic studies. Results from genome-wide association studies often generate genetic variants that are associated with an intermediate trait. These variants can then be used as instrumental variables to investigate if the intermediate trait has a causal influence on another (outcome) trait. However, a valid instrumental variable must satisfy certain conditions, in particular, that there are no direct relationships between the instrumental variable and the outcome variable or any other potential confounders. An alternative approach for inferring the causal direction between two variables is with the use of Bayesian networks (BN), which have no such explicit causal assumptions and benefit from using more than three variables in an analysis. We compare MR and BN using simulated data in a four variable model where X is (usually) causal on Y and G and Z are instrumental variables for X and Y respectively. We consider MR analyses using G or Z as an instrumental variable and BN analyses using only G or Z as well as both G and Z. We consider three different models: (i) a model with no confounding variables; (ii) a non-genetic confounding model, where a variable influences both X and Y, and (iii) a genetic confounding model, where G influences Y via an additional intermediate variable. Our results show that for models (i) and (ii) BN showed comparative power to MR and that the use of both instrumental variables with BN showed greater power than using either separately. Receiver operating characteristic (ROC) curves were considerably better for BN than MR. The non-genetic confounder model did not unduly affect either the MR or BN analyses. The genetic confounder model caused problems for both MR and BN. For MR with G as an instrumental variable there was a huge inflation in type I error due to the instrumental variable assumption not holding. BN analysis type I error was well controlled when no effect from X to Y was present but the type I error was moderately inflated when there was a causal effect from Y to X and G was the only instrumental variable. We have shown that while MR is easy to perform and widely understood, BN offers an alternative approach with some advantages over MR such as better ROC curves, ability to utilise more variables and less adverse effects when confounding is present.

3547T

Controls: How large is too large? A survey of methods for optimal subsampling in large-scale case-control studies. *Y. Jiang, C. Tian, D. Hinds, 23andMe Research Team.* 23andMe, 899 West Evelyn Ave, Mountain View, CA.

Large-scale case-control studies provide unparalleled opportunities to advance the understanding of the genetic architecture of complex diseases. One emerging challenge of analyzing such enormous datasets is that the number of controls collected is substantially larger than the number of cases for most of the phenotypes, particularly rare diseases. 23andMe's case-control cohorts are close to the general population-based cohorts, where the control-to-case ratios are over 100 for many of the collected phenotypes. Currently, existing methods include all available controls in the analyses. This approach not only demands huge computational resources, but without proper adjustment, it can also lead to significantly inflated type I error rate induced by the imbalanced case-control ratio. In this study, we examined the hypothesis that gains in power are marginal after control-to-case ratio reaches a certain threshold for fixed sample size and allele frequency, and we attempted to identify the saturate ratio. We focused on investigating strategies that deliberately sample a subset of controls to lower the computational burden and reduce the imbalance ratio, which can improve power in certain scenarios. Even though a few subsampling methods have been proposed, the best practice to select controls remains to be elucidated. Propensity score and coarsened exact matching are commonly used methods in social sciences. We adopted these strategies and selected various proportions of samples to include in a GWAS. Our results showed that for most diseases, once the control-to-case ratio reaches 10, the selected sub-sample GWAS can recover the hits of GWAS using the full database. However, some studies have argued that the estimation is inconsistent if the model is not correctly specified. "Local case-control" sampling and "A-optimal" sampling use a post-hoc analytic adjustment to improve the biased pilot estimate, which provides a consistent estimation. The pilot fit, however, introduces tremendous computational cost for large-scale GWAS study where millions of SNPs are tested. Using 23andMe's dataset of over 3 million consented research participants, we use empirical evidence to systematically evaluate and discuss the usage of multiple subsampling strategies. We provide practical guidelines for control selection across different disease prevalence, minor allele frequency, and computational speed.

3548F

An analysis of the genetic overlap of 20 complex traits under a non-infinite model. R.D. Johnson¹, H. Shi², K. Burch², B. Pasaniuc^{2,3,4}, S. Sankararaman^{1,2,3}. 1) Dept of Computer Science, University of California, Los Angeles, CA 90024; 2) Bioinformatics Interdepartmental Program, University of California, Los Angeles, 90024; 3) Dept of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA 90024; 4) Dept of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California, Los Angeles, 90024.

Quantifying the shared genetic component of two complex traits can elucidate the genetic basis underlying the comorbidity of various traits. Large-scale genome-wide association studies conducted for thousands of traits have provided a catalog of genetic variants that are implicated across multiple traits. Current methods that quantify the genetic overlap between a pair of traits through GWAS association data either assume an infinitesimal model where every SNP is causal or relax this assumption but restrict the number of causal SNPs due to computational constraints. We previously introduced UNITY (Unifying Non-Infinite Trait analysis), a flexible, unifying framework that quantifies the overlap between two traits through GWAS summary association data. UNITY calculates the proportion of shared causal SNPs between two traits, as well as the proportion of causal SNPs that are specific to each trait. UNITY makes minimal assumptions where we do not assume every SNP is causal, also called the non-infinite model. Additionally, our inference does not restrict the number of inferred causal variants a priori, thus allowing for a variety of genetic architectures. In this work, we used UNITY to systematically quantify the proportion of trait-specific and shared causal variants between 380 pairs of complex traits through publically available GWAS summary statistics. We show that traits that were believed to be extremely polygenic actually only contain a small fraction of the genome to be causal. For example, our analysis of height showed that <5% of the genome is estimated to be causal, thus motivating the use of methods that do not impose the infinitesimal assumption.

3549W

Leveraging variance QTLs in a twins cohort identifies epistatic interactions influencing gut microbiome composition. A.R. Marderstein^{1,2}, E.R. Davenport³, C.V. Van Hout⁴, A.G. Clark^{3,4}. 1) Tri-Institutional Program in Computational Biology & Medicine, Cornell University/Weill Cornell Medicine, Ithaca/New York, NY; 2) Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY; 3) Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY; 4) Regeneron Genetics Center, Tarrytown, NY.

Efficiently identifying epistatic (GxG) and genotype-by-environment (GxE) interactions in humans remains an important challenge. Since exhaustive all-pairwise interaction testing involves $n(n-1)/2$ tests (for n SNPs), it is computationally intensive and the multiple test correction severely limits statistical power. To address this gap, we screened for genetic associations with the variance of a phenotype (vQTLs), which can be caused by both gene x gene (GxG) and genotype x environment (GxE) interactions. Screening for vQTLs involves just n tests, so it provides a powerful inroad to discovering genetic interactions by nominating loci as promising candidates for an interaction. In our study, we first used simulations to demonstrate that two variants affecting the mean of a phenotype through an interaction are also likely to be associated with the single-SNP stratified phenotypic variances. We then applied our method in a sequential approach to identify epistatic interactions impacting gut microbiome composition in >2000 individuals from the TwinsUK cohort. First, we performed a *two-step squared-residual linear mixed model* approach to identify genome-wide associations (1.1 million SNPs) with the variances of 945 bacterial relative abundances. Second, to discover novel GxG interactions, we tested significant vQTL loci with all other variants using mean-based interaction models with robust Huber-White standard errors. This robust standard error approach better maintains the genomic inflation factor across phenotypes in comparison to a default standard error approach ($\lambda_{GC, Normal}: \mu=1.20$, $\square=0.14$; $\lambda_{GC, Robust}: \mu=1.02$, $\square=0.02$). We discovered 252 genome-wide significant interactions ($P < 5 \times 10^{-6}$) across the 945 microbial phenotypes, including a study-wide significant host genetic interaction ($P < 4.7 \times 10^{-11}$) impacting Clostridiales abundances. This interaction is between an intergenic eQTL for three nearby genes in the bitter taste receptor family (*TAS2R3*, *TAS2R4*, *TAS2R5*) and another intergenic SNP in LD with an active regulatory region located 75 kb upstream of the *APOB*, which has been previously linked with LDL levels ($P = 10^{-182}$). This interaction was also genome-wide significant with 29 other taxa within the Clostridiales order, including taxa in the Ruminococcaceae and Lachnospiraceae families. Our results demonstrate that vQTLs provide a practical and powerful approach to discovering GxG and GxE interactions in human genome-wide association studies.

3550T

Estimating differential networks in microbiome data: Penalized precision matrix estimation in a multinomial model. *K. McGregor^{1,3}, A. Labbe², C.M.T. Greenwood³*. 1) McGill University, Montreal, Quebec, Canada; 2) HEC Montreal, Montreal, Quebec, Canada; 3) Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, Quebec, Canada.

The microbiome is the collection of microorganisms colonizing the human body, and plays an integral part in human health. A growing trend in microbiome analysis is to construct a network to estimate the level of co-occurrence between different taxa. Current methods do not facilitate investigation into how these networks change with respect to some phenotype. We propose a model to estimate network changes with respect to a binary phenotype. The counts of individual taxa in the samples are modelled through a multinomial distribution whose probabilities depend on a Gaussian random effect term. A sparse precision matrix over all the random effect terms determines the co-occurrence network among taxa. The model fit is obtained and evaluated using Monte Carlo methods. The performance of the model has been evaluated through an extensive simulation study, and has been shown to outperform existing methods in terms of estimation of network parameters. We also demonstrate an application of the model to estimate changes in the murine intestinal microbial network topology with respect to macrolide exposure. A new method to approximate the posterior distribution will be introduced to fit the model without the computational burden of Monte Carlo sampling.

3551F

A selection operator for summary association statistics reveals allelic heterogeneity of complex traits. *Z. Ning¹, Y. Lee², P. Joshi³, J. Wilson^{3,4}, Y. Pawitan¹, X. Shen^{1,3}*. 1) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Nobels vag 12A, SE-171 77 Stockholm, Sweden; 2) Department of Statistics, Seoul National University, Seoul 151747, South Korea; 3) Center for Population Health Sciences, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Old Medical School, Teviot Place, Edinburgh, EH8 9AG, Scotland, United Kingdom; 4) MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, EH4 2XU, Scotland, United Kingdom.

In recent years, as a secondary analysis in genome-wide association studies (GWAS), conditional and joint multi-variant analysis (GCTA-COJO) has been successful in discovering additional association signals within detected loci. This suggests that many loci mapped in GWAS harbor more than a single causal variant. We develop a penalized selection operator (SOJO) within each mapped locus, based on LASSO regression derived from summary association statistics. We show that SOJO achieves better variable selection accuracy and prediction performance. Our empirical results indicate that human height is not only a highly polygenic trait, but also has high allelic heterogeneity within its established hundreds of loci. An R package named "sojo" is made to implement this method.

3552W

MRMix: A mixture model approach to robust and efficient Mendelian randomization analysis. G. Qi¹, N. Chatterjee^{1,2}. 1) Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; 2) Department of Oncology, Johns Hopkins School of Medicine, Baltimore, MD.

Background: Many commonly used methods for Mendelian Randomization (MR) analysis rely heavily on the assumption that genetic correlations across risk-factors and health outcomes mainly arise from underlying causal relationships. Some recent methods are more robust to the violation of this assumption, but they can lack efficiency and can still be susceptible to bias if a large majority of the genetic variants violate the assumption. **Methods:** We consider a “working model” to describe the bi-variate effect-size distribution of SNPs across X and Y using a normal mixture model (MRMix) that assumes the existence of only a fraction of SNPs which are valid instruments, i.e. they have only direct effect on X , while other SNPs can have both direct and indirect effects or have no effects at all. Under this model, the residual of the form $\beta_X - \theta\beta_Y$ - here β_X and β_Y denote the total effect of a SNP on X and Y , respectively, and θ is the putative causal effect - has a mixture normal distribution with an enriched point mass at 0 only at $\theta = \theta_0$, the true value of the causal effect. Based on this observation, we propose a simple method for estimating the causal effect θ_0 through a grid-search procedure to maximize the probability concentration of observed residuals $\beta_X - \theta\beta_Y$ at the “null” component of a two-component normal mixture model. **Results:** Simulation studies showed that MRMix provides nearly unbiased or/and substantially more robust estimates of causal effects compared to alternatives under various scenarios. Further, the studies showed that MRMix is sensitive to direction and can achieve much higher efficiency (up to 3–4 fold) relative to a non-parametric mode based ratio estimator that is found to be comparably robust. We applied the proposed methods for conducting MR analysis using largest publicly available datasets across a number of risk-factors and health outcomes. Notable observations include confirmation of negative relationship between genetically determined BMI and the risk of breast cancer; a strong detrimental effect of BMI on the risk of major depressive disorder; no causal effect of HDL on coronary artery disease; detrimental effect of higher level of HDL on the risk of breast cancer; and positive effect of higher level of LDL on bone mineral density. Results from more large-scale MR analyses will be presented at the meeting. **Conclusion:** MRMix provides a robust and efficient method for estimation of causal effects in two-sample MR analysis.

3553T

Genotype imputation using the pBWT. S. Rubinacci¹, J. Marchini^{1,2}. 1) Department of Statistics, University of Oxford, Oxford, Oxfordshire, United Kingdom; 2) Wellcome Center for Human Genetics, University of Oxford, Oxford, Oxfordshire, United Kingdom.

Genotype imputation is the process of predicting genotypes that have not been typed in a sample of individuals. We have developed two new methods that aim to reduce the memory requirements and speed of the IMPUTE model. Firstly, we have implemented a new version of the IMPUTE v4 method that uses a reference panel stored in a positional Burrows-Wheeler Transform (pBWT) data structure. This approach conditions on all the haplotypes in a reference panel. When using the Haplotype Reference Panel (HRC) with 64,812 haplotypes this new method reduces the memory by 75% compared to IMPUTEv4, with a corresponding modest increase in compute time of 23%. Secondly, we have developed a new version of IMPUTE that very quickly selects a small subset of haplotypes to condition on. As the reference panel grows we expect to find an ever smaller set of good conditioning haplotypes. By adopting this idea we aim to develop a method that has sub-linear scaling in reference panel size. Again we use a pBWT data structure for the reference panel and insert each sample haplotype into this structure in order to identify locally best matching haplotypes. This operation has complexity independent of the reference panel size. The method then uses the selected haplotypes as conditioning states within the IMPUTE model. We have tested our method using the HRC chromosome 20 reference panel and compared the results to IMPUTE v4. We imputed genotypes into 10 individuals of European ancestry, based on genotypes on Illumina 1M-Duo3_C genotyping array, and compared the results to genotypes derived from high-coverage sequencing. Imputation was carried out in windows of 5Mb. Our method has a parameter L indicating the maximum number of possible copying states to add at each marker, and we have used values of $L=1, 8, 12, 25$ for each window, and this results in an average of 115, 503, 777 and 1,623 copying states respectively. Our method is more than 10x faster than IMPUTE v4, with almost identical memory requirements for every value of L tested. The new copying state selection takes only 0.56% of the time of an imputation task and produces the same accuracy as IMPUTE v4. For example, for imputed variants with frequency 0.1% IMPUTE v4 and our new method with $L=12$ produce mean imputation r^2 of 0.67 and 0.68 respectively. With $L=1$ $r^2=0.63$, highlighting that closest matches in the pBWT data structure contain a lot, but not all of the relevant information. .

3554F

Accounting for cryptic relatedness across families between subjects with no genotype data. *M. Saad¹, E. Ullah¹, E.M. Wijsman^{2,3}*. 1) Qatar Computing Research Institute, Doha, Qatar; 2) Department of Biostatistics, University of Washington, Seattle, WA, USA; 3) Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA, USA.

Complex traits continue to provide challenges for identifying disease-risk genes. Whole genome sequencing has become inexpensive and fast. This provides opportunities for investigation of rare variation. In the search for disease-associated rare variation, family-based designs have again become common, because rare, highly-penetrant genotypes can segregate in pedigrees. Association testing is often used to identify variants of interest. When related subjects are included in the sample, a kinship matrix must be used to account for relatedness between subjects. Relationships may be known from the pedigree structures, or can be inferred using observed genotypes. For subjects without genotype data, however, no approach is able to infer relatedness between subjects in different pedigrees. Pedigree-based imputation increases the sample size and thus association power, and also provides genotype probabilities in subjects without genotype data. When phenotypes for such subjects exist, including them in the analysis adds information for detecting association between genotypes and phenotype. However, knowing the relationship between such subjects and the remaining subjects, especially across pedigrees, is crucial to control the type 1 error. We propose a solution for inferring cryptic relatedness between subjects with completely missing genotype data. We use GIGI to estimate the probabilities of missing genotypes, and incorporate them in an Expectation-Maximization approach to estimate the kinship coefficients. Through simulation, our approach succeeds in inferring many types of relationships with relatively high confidence. It yields average kinship estimates of 0.19, 0.1, and 0.05 for underlying kinships of 0.25, 0.125, and 0.0625, respectively, for subjects without genotype data.

3555W

Enrichment analysis in genome-wide association studies through modelling of effect-size distribution in relation to annotations and genomic propensity scores. *A. Saha¹, Y. Zhang¹, N. Chatterjee^{1,2}*. 1) Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD; 2) Department of Oncology, School of Medicine, Johns Hopkins University, Baltimore, MD.

Background Heritability enrichment analyses in genome-wide association studies (GWAS) are commonly used to gain understanding of the contributions of various population genetic and functional genomic characteristics of the genome on trait variation. Assessment of the effect-size distribution (ESD) associated with genome-wide panel of SNPs in relationship to annotation information can provide more detailed insights into the genetic architecture of complex traits. The effort to disentangle contributions of different annotations, however, can be complicated by their complex interrelationships. **Method** We assume a normal-mixture model for the ESD with varying variance component parameters to allow for clusters of SNPs with distinct effect-sizes. We allow the SNP annotations to influence the ESD through the underlying mixing probabilities in a log-linear modelling framework. Under the proposed model, we show that the distribution of summary-statistics of GWAS markers can be approximated by an alternative normal-mixture form with the mixing probabilities depending on the probability distribution of number of SNPs from different clusters the GWAS markers tag. Further, in this representation, the variance component parameters depend on weighted LD-scores with weights depending on local annotation information. We propose utilizing the model to characterize enrichment by each individual annotation, while adjusting for others through underlying "genomic propensity scores", which could be constructed based on simple regression modelling among the annotations themselves. **Results** We applied the proposed method for the analysis of publicly available GWAS summary statistics across a wide variety of traits together with the annotation information available from various databases. Preliminary analyses show that the method can identify evidence of negative selection, i.e., inverse relationship between effect-sizes and minor allele frequencies, across a wide variety of traits. Further, it identifies evidence of strong enrichment for certain annotations like transcription start site and conserved region. Results from more detailed analysis by individual traits, annotations and underlying cell-type context will follow. **Conclusion** Modelling of effect-size-distribution in relationship to genomic annotations and underlying propensity scores can provide novel insights into genetic architecture of complex traits.

3556T

High-definition likelihood inference for heritability and genetic correlation using GWAS summary statistics. X. Shen^{1,2,3}, Z. Ning¹, Y. Pawitan¹. 1) Karolinska Institutet, Stockholm, Sweden; 2) Sun Yat-sen University, Guangzhou, China; 3) University of Edinburgh, Edinburgh, United Kingdom.

Motivation: In the past three years, LD Score Regression has been an influential technique, allowing estimation of heritability and genetic correlation using only GWAS summary statistics. LD Score Regression was also reported to be able to distinguish polygenicity from population structure in a GWAS study. However, the definition of LD score only uses part of the linkage disequilibrium information, thus the current LD Score Regression method has a limited efficiency of estimation. **Results:** In order to improve LD Score Regression as a method based on a "low-definition" likelihood (LDL) function, we develop a full high-definition likelihood (HDL) model that accounts for substantially more information in the LD structure. Simulation studies show that HDL has about 7 times efficiency compared to LD Score Regression, e.g. for genetic correlation estimation. With this, HDL is able to report accurate heritability and genetic correlation estimates with much smaller standard errors, identifying more genetically correlated complex traits. With the full likelihood method, bootstrap implementation is avoided so that computational efficiency is also improved. **Conclusion:** The HDL method, as a full likelihood model of summary association statistics and LD structure, substantially improves the accuracy of heritability and genetic correlation estimation. Some other remaining issues of LD Score Regression can also be solved directly when the full likelihood is considered.

3557F

A statistical framework to quantify shared genetic variability in *de novo* mutations for complex diseases. Y. Shi¹, Q. Lu², SC. Jin³, X. Zeng³, W. Dong³, B. Li¹, J. Warren¹, R. Lifton^{3,4}, M. Brueckner³, H. Zhao^{1,5}. 1) Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA 06511; 2) Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI, USA 53792; 3) Department of Genetics, Yale University, New Haven, CT, USA 06510; 4) Laboratory of Human Genetics and Genomics, Rockefeller University, New York, USA 10065; 5) Program of Computational Biology and Bioinformatics, Yale University, New Haven, CT, USA 06510.

Damaging *de novo* mutations (DNMs) in protein-coding genes are known to be critically involved in the etiology of many developmental diseases. In particular, exome sequencing studies of thousands of proband-parent trios have identified a large number of causal genes for congenital heart disease (CHD) and autism spectrum disorder (ASD). These results suggested substantial contributions of DNMs – 10~15% of CHD and ASD diagnoses can be explained by pathogenic DNMs. Interestingly, a substantial fraction of genes were enriched for DNMs in both CHD and ASD, hinting at shared genetic pathways. However, these results were obtained in hypothesis-driven analyses using huge samples. It is therefore of practical interest to develop a powerful framework to quantify shared genetics using DNM data of two diseases and search for common pathways while adjusting for other factors contributing to mutation rate (e.g. gene length and sequence context). In this project, we introduced a hierarchical Poisson model to jointly model DNM counts in two diseases and quantify shared DNM effects via an enrichment correlation (EC) metric. We have implemented both frequentist and Bayesian methods for parameter inference and demonstrated the effectiveness of our method using extensive simulations. We then applied our method to real DNM data from 1,213 CHD trios and 2,508 ASD trios acquired from denovoDB. We identified strong significant correlation between loss-of-function (LoF) DNM counts of CHD and ASD (EC=0.38; 95% credible interval=[0.22, 0.50]). As a positive control, we randomly partitioned 1,213 CHD trios into two equal-sized subsets and our method identified a strong correlation (EC=0.61; 95% credible interval=[0.39, 0.90]) in LoF DNMs. We also repeated the analysis in a control dataset (N=1,789) and did not observe significant correlation (EC=-0.03; 95% credible interval=[-0.31, 0.21]). Additionally, we tested correlation of CHD and ASD using 546 genes in the chromatin remodeling pathway – a key pathway for both CHD and ASD. LoF EC between CHD and ASD was further amplified (EC=0.70 [0.38, 0.91]) in this pathway. In summary, our method provides a novel, robust metric to quantify correlation in DNM enrichment between diseases. This new method has the potential to be widely used in future DNM studies to systematically screen for shared genetic components.

3558W

Joint analysis of GWAS summary statistics in East Asians and Europeans provides insights into population-specific and shared causal variants of complex traits. *H. Shi¹, K. Burch¹, R. Johnson³, M. Freund⁴, G. Kichaev¹, N. Mancuso², A. Manuel⁵, N. Dong⁶, B. Pasaniuc^{1,2,4}.* 1) Bioinformatics Interdepartmental Program, University of California Los Angeles, Los Angeles, CA; 2) Dept of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA; 3) Dept of Computer Science, University of California Los Angeles, Los Angeles, CA; 4) Dept of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA; 5) Dept of Biology, Florida International University, Miami, FL; 6) Dept of Biomedical Engineering, Boston University, Boston, MA.

Trans-ethnic genetic studies have been shown to improve detection of trait-associated loci and fine mapping of causal variants (Morris, *Genet Epidemiol*, 2011; Kichaev, *Am J Hum Genet*, 2016). However, although GWAS risk regions are largely replicated across different continental populations, whether the replication is due to shared causal variants or different causal variants that impact the same genes remain unclear. Here, we introduce POSC, a Bayesian approach to both estimate genome-wide proportion of population-specific/shared causal variants and identify individual SNPs that are population-specific/shared from GWAS summary statistics, while accounting for linkage disequilibrium (LD) of the two continental populations. In detail, POSC performs inference in two stages. In the first stage, POSC estimates the genome-wide prior probability that a SNP is population-specific/shared using all SNPs regardless of their statistical significance. In the second stage, POSC quantifies posterior probability of each of these SNPs to be population-specific/shared using the prior obtained in the first stage. We show through extensive simulations that POSC yields approximately unbiased estimates of genome-wide proportions of population-specific/shared causal variants, and robustly infers individual population-specific/shared causal variants. We apply POSC on GWAS summary statistics data of 9 complex traits and diseases in East Asians (Kanai et al., *Nat Genet*, 2018) and Europeans (average $N_{EAS} = 94,621$, $N_{EUR} = 103,507$). Across all analyzed traits, on average, 84% (82%) of the causal variants in East Asians (Europeans) are also causal in the other population, and 16% (18%) of the causal variants are specific to East Asians (Europeans). The large fraction of shared causal variants is in concordance with the high trans-ethnic genetic impact correlation ($R_g = 0.85$ on average) of these traits (Brown et al., *Am J Hum Genet*, 2016). Next, we find that genomic regions having GWAS-significant associations in only one population, on average, show a 2.8X enrichment of shared causal variants than expected from the prior, providing evidence that GWAS-significant associations not present in GWAS studies in other populations are likely in part due to different LD structures. Finally, we show that regions surrounding genes specifically expressed in trait-relevant tissues (Finucane et al., *Nat Genet*, 2018) harbor a disproportionate amount of both shared and population-specific causal variants.

3559T

Enhanced case-control association analyses leveraging family history.

C.R. Solis-Lemus¹, S.T. Fischer¹, E.J. Leslie¹, D. Ghosh², D.J. Cutler¹, M.P. Epstein¹. 1) Department of Human Genetics, Emory University, Atlanta, GA; 2) Department of Biostatistics and Informatics, Colorado School of Public Health, Aurora, CO.

Standard methods for case-control association studies of rare and common variation often treat disease outcome as a dichotomous phenotype. However, both theoretical and experimental studies have demonstrated that cases with a family history of disease can be enriched for risk variation relative to sporadic cases. Likewise, controls with a family history of disease can harbor more risk variation than controls without such history. Assuming family-history information is available, this observation motivates the idea of replacing the standard dichotomous outcome variable in case-control studies with a more informative outcome variable that distinguishes cases and controls with and without family history. Potential options for enhanced outcome variables include ordinal categorical variables and conditional means from liability-threshold models. We explore both options to create modified association tests for common and rare variation that leverage family-history information. Preliminary results suggest that leveraging family-history information via ordinal categorical variables can greatly improve power in rare-variant association tests compared to standard dichotomous modeling of disease phenotypes that ignore such information. We further illustrate our methods using GWAS and NGS data from a genome-wide study of craniofacial defects.

3560F

Building multi-predictor models of quantitative phenotypes from the Trinity Student Study with tiled regression. A.J.M. Sorant¹, J.A. Sabourin¹, C.M. Justice¹, L.C. Brody², J.E. Bailey-Wilson¹, A.F. Wilson¹. 1) Computational and Statistical Genomics Branch, National Human Genome Research Institute, NIH, Baltimore, MD, USA; 2) Gene and Environment Interaction Section, National Human Genome Research Institute, NIH, Bethesda, MD, USA.

The Trinity Student Study obtained DNA and metabolite measurements from blood samples provided by a cohort of healthy Irish college students, as well as health and lifestyle information from questionnaires [Molloy et al. 2016]. After data cleaning, 757,533 genotypes were available for analysis on 2232 individuals. Imputation of the genotyped data resulted in a set of 2,350,763 imputed genotypes. Significant results of genome-wide association studies have been published for several quantitative phenotypes including methylmalonic acid concentration (MMA) [Molloy et al. 2016] and serum vitamin B12 concentration (SB12) and holo-transcobalamin (HoloTC) [Velkova et al. 2017]. Previous results were determined with simple linear regression (SLR) analyses (single-marker tests) with imputed genotypes. Because of linkage disequilibrium, single-marker studies may pick up many associations which might or might not be due to a single causal effect. In the present study, tiled regression (TR), which combines variable selection with a staged approach [Wilson et al., 2009], was used to form a multiple linear regression (MLR) model, determining combinations of predictors that jointly and independently contribute to the variation in each phenotype. Although the selected model may include a correlated predictor rather than a causal one, multiple predictors chosen from the same region can indicate multiple causal effects. Another difference is that significance, as measured by an SLR p-value, may be different from that representing its significance in an MLR model. As in the previous studies, a $\log_{10}$ -transformed version of each phenotype, adjusted for age and sex, was analyzed, initially with raw genotypes. Tiled regression analysis was performed using TRQUANT v. 0.3 [Sorant et al., 2017], with stepwise selection criteria 10^{-5} , 10^{-6} and 10^{-7} at the tile, chromosome and genome-wide stages, respectively. With one exception, each variant reported as significant in the previous results either was included in the final selected MLR model or a nearby one was. (The exception was significant only at a suggestive level in the original analysis.) In some cases, additional variants, sometimes but not always from the same regions, were also found to jointly and independently contribute significantly to the model. Interactions among variants selected at the genome-wide stage were also considered, but they did not improve on the main effects model.

3561W

Heterogenous effects linear mixed model improves whole genome association testing in multi-ethnic populations: Applications to the TOPMed and PAGE Consortia. T.A. Thornton¹, M.P. Conomos¹, S.G. Gogarten¹, A.P. Reiner². 1) Biostatistics, University of Washington, Seattle, WA., Select a Country; 2) Epidemiology, University of Washington, Seattle, WA., Select a Country.

Whereas genetic studies of complex traits have primarily examined populations of European ancestry, recent studies, such as the Trans-Omics for Precision Medicine (TOPMed) program and the Population Architecture using Genomics and Epidemiology (PAGE) study, have generated an abundance of whole-genome genotyping and sequencing data from multi-ethnic populations for complex trait mapping. These data offer an unprecedented opportunity for the identification of novel rare and common genetic variation underlying phenotypic diversity among populations, as well as the potential to provide new insights into health disparities of minority populations for many complex diseases. However, beyond the obvious computational issues of analyzing millions of genetics variants from whole-genome data in tens of thousands of individuals, there are substantial statistical challenges for complex trait mapping in multi-ethnic samples with heterogenous genomes from a variety of study designs. Linear mixed models (LMMs) have been shown to be a powerful and effective approach for genome-wide association testing in samples with population structure, family structure, and/or cryptic relatedness. Existing LMM implementations, however, have largely been used and evaluated in relatively homogenous European populations. We demonstrate that widely used LMM implementations, such as GCTA, EMMAX and GEMMA, can fail when applied to samples from multi-ethnic populations because these methods do not appropriately model heterogenous variances of traits across multi-ethnic. To address the shortcomings of existing LMM methods, we developed HEMMAT (heterogenous effects mixed model association test) for genetic association testing of quantitative traits in multi-ethnic populations. In applications using real genotype and phenotype data from the multi-ethnic TOPMed and PAGE studies, we demonstrate that HEMMAT provides a significant improvement, in terms of both type-1 error and power, over existing LMM approaches for complex trait mapping in multi-ethnic populations.

3562T

Genome-wide priors for gene prioritization. *E. Weeks¹, J. Ulirsch^{1,2}, K. Tashman¹, J. Engreitz¹, H. Finucane¹.* 1) Broad Institute of MIT and Harvard, Cambridge, MA; 2) Harvard T.H. Chan School of Public Health, Boston, MA.

Identifying causal genes for complex traits remains one of the biggest challenges in the post-GWAS era. For the vast majority of GWAS associated loci, the identity of the causal gene(s) driving the association is unknown. One approach to begin to understand the underlying causal mechanisms behind these loci is via gene prioritization methods such as colocalization methods or methods that use enhancer-promoter interaction data. Here, we propose a gene prioritization method that leverages genome-wide enrichments of gene features such as cell-type specific gene expression, pathways, and protein-protein interaction networks. When computing the prioritization score for a gene, we leave out the GWAS summary statistics for the chromosome containing the gene, which makes the results of our method ideal for combining with other methods that use locus-specific data. Specifically, we envision that our method can be integrated into other gene prioritization methods as a way of assigning priors to genes for being causal. In this way, we expect to be able to combine our broader genome-wide approach with locus-specific methods to more accurately predict the causal genes for complex traits. Our method consists of two steps: first, we assign a prior probability of being causal to each gene using the GWAS summary statistics for the trait of interest. Second, we update our prior using the gene features in a regularized weighted logistic regression framework. The output of our model is, thus, biologically informed probabilities for every gene that it is causal for a given phenotype computed without using the GWAS data at the locus containing the gene. We apply our method to summary statistics from 73 complex traits and 593 gene features derived from several publicly available sources of gene expression data. We evaluate our results by correlating the output probabilities with our prior probabilities and show that our method has a significant correlation for 67 out of 73 phenotypes ($P < 0.05/73$). For two traits, inflammatory bowel disease and white blood cell count, we more rigorously evaluate our results using a set of gold standard genes with fine-mapped coding variants. For inflammatory bowel disease, the mean score was 0.0603 for our gold standard genes and 0.0518 for all other genes ($P = 2.5 \times 10^{-6}$). For white blood cell count, we saw a more modest difference with a mean score of 0.0585 for our gold standard genes and 0.0518 for all other genes ($P = 1.8 \times 10^{-6}$).

3563F

Simulation study of double FDR method on LD block structure in GWAS. *J. Xu, D. Mehrotra.* BARDS, Merck & Co., North Wales, PA 19454, PA.

P-value threshold after multiplicity correction is critical in genome wide association study (GWAS) to call out true signals while controlling false positives. The current well accepted threshold is 5×10^{-8} , which aims to control 5% family wise error rate (FWER) using Bonferroni method assuming one million independent SNPs. This threshold has been broadly applied in GWAS field given its simplicity. However, the between SNP correlation structure has not been considered and FWER is much more restricted than false discovery rate (FDR). In this simulation study, we propose applying double FDR method on linkage disequilibrium (LD) block structure. This method yields much higher power than $p < 5 \times 10^{-8}$ while controlling FDR at 5%.

3564W

Dimension reduction and remedy for multicollinearity of DNA sequencing data using the lasso based clustering. Y.J. Yoo. Dept Mathematics Education, Seoul National University, Seoul, South Korea.

For genetic association analysis based on multiple SNP regression using high density DNA sequencing data, multicollinearity can be a serious issue. We propose an algorithm to reduce the dimension of the sequencing data by selecting a clusters in multicollinearity and replacing the cluster with by principal components. In this algorithm, we first detect SNPs causing multicollinearity based on variance inflation factor (VIF). SNP variable with highest VIF over a threshold value is clustered with certain SNPs selected by lasso regression of the highest VIF SNP on the rest of SNP variables. The SNPs forming a cluster which are in multicollinearity are replaced by principal components explaining most of the variability caused by these SNPs. This procedure continues in greedy manner until the highest VIF of update variables are below the threshold. We applied the method to 1000 Genomes Project data and found that the degrees of freedom are reduced to 7~80% based on the threshold options. Through simulation study, we investigated how the power of multi-marker tests constructed from multiple SNP regression is affected by this dimension reduction procedure. We found that Wald test can gain power by this procedure, and linear combination (LC) test and multiple linear combination (MLC) test which is a hybrid test between Wald and LC test yielded similar power with the results using entire SNP variables.

3565T

Thresholding Fisher's p-value combination method (TFisher) for set-based genetic association studies. H. Zhang¹, Z. Wu². 1) Merck & Co., Inc., Kenilworth, NJ; 2) Worcester Polytechnic Institute, Worcester, MA.

Genetic association studies have successfully identified many genetic variants that associate with complex traits and diseases. It is believed that grouping the individual effects of multiple variants may yield extra insight into the complex diseases. Many p-value combination methods have been developed to detect the overall effect of a set of genetic variants, e.g. Fisher's p-value combination method and its derivatives like truncated product method. We propose a unified family of statistics, called TFisher, with general p-value thresholding and weighting schemes in which a soft-thresholding scheme is shown to have optimal statistical power for signal detection in a large variety of signal patterns. Analytical calculation of the p-value for arbitrary correlation structure is given. When prior information of signal pattern is unavailable, an omnibus test, oTFisher, is proposed to provide robust results. Simulations evidence the accuracy of calculation and validate the theoretical advantages of the soft-thresholding method. The TFisher tests are applied to analyzing a whole exome sequencing data of amyotrophic lateral sclerosis, as well as Genetic Factors for Osteoporosis Consortium (GEFOS) using summary statistics. The proposed methods have been implemented into an R package TFisher and published on the CRAN <https://cran.r-project.org/package=TFisher>.

3566F

Genetic mapping of human diseases using gene damage scores. X. Zhu¹, S. Gustafsson², W. Wong^{1,3}, E. Ingelsson^{4,5}, J. Zou^{2,6,7}. 1) Department of Statistics, Stanford University, Stanford, CA; 2) Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Sweden; 3) Department of Biomedical Data Science, Stanford University, Stanford, CA; 4) Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA; 5) Stanford Cardiovascular Institute, Stanford University, Stanford, CA; 6) Department of Computer Science, Stanford University, Stanford, CA; 7) Department of Electrical Engineering, Stanford University, Stanford, CA.

Genetic scores that aggregate information from multiple variants have proven successful for understanding the role of genetic variation in complex traits (e.g. Purcell et al, 2009; Gamazon et al, 2015). Here we considered a "Gene Damage Score" (GeneDS) that quantifies total mutational damage of a gene in an individual. Specifically, we defined GeneDS of an individual for a target gene as the minor allele sum of all functional variants in this gene, weighted by variant-level deleterious metrics. In this work, we selected functional variants as SNPs predicted to cause moderate-to-severe damage (including various mutations from missense to stop-gain), and used CADD scores (Kircher et al, 2014) as variant weights. To test the utility of GeneDS for mapping diseases, we conducted a whole genome-phenome association scan of 16,080 genes and 745 diseases on 337,536 unrelated individuals of European ancestry within UK Biobank. We correlated GeneDS of each gene with diagnosis of each disease via logistic regression, controlling for age, gender and genotype-based principal components. Our large-scale analysis illustrates multiple advantages of GeneDS in association studies when compared to conventional methods. First, GeneDS association signals are almost independent, since correlations between GeneDS of nearby genes are much smaller than linkage disequilibrium between nearby SNPs. Second, unlike variant-based methods, GeneDS output can be directly used in gene set enrichment analyses, which may yield additional biological insights. Third and perhaps most importantly, in contrast to most gene-based tests, GeneDS provides an interpretable direction of effect for the observed associations, which may help elucidate the mode of action for a therapeutic target gene. As a proof of concept, increased damage in *SCN1A* was associated with reduced risk in epilepsy ($z=-3.92$, $FDR<0.1$); this negative correlation is consistent with the fact that *SCN1A* is the target of four different approved blockers for epilepsy. Finally, although defining GeneDS requires individual-level genotypes, GeneDS-based association analyses can be performed on GWAS summary statistics. This feature greatly facilitates replication studies. For example, we identified a positive association between GeneDS of *HGFAC* and coronary artery disease risk ($z=6.98$, $FDR<0.1$) in UK Biobank, and then replicated this finding ($z=2.67$) using external summary data from the CARDIoGRAMplusC4D consortium (Nikpay et al, 2015).

3567W

A Bayesian mixture model to accommodate multiple sources of uncertainty into transcriptome-wide association studies. J.D. Rosen¹, Y. Li^{2,3}. 1) Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599; 2) Department of Genetics, University of North Carolina, Chapel Hill, NC 27599; 3) Department of Computer Science, University of North Carolina, Chapel Hill, NC 27599.

In the past several years various methods have been proposed for exploring associations between genes and complex traits. Sometimes referred to as transcriptome-wide association studies (TWAS), these methods share a common framework: training of a predictive model of gene expression based on eQTL data, prediction of gene expression using GWAS data for which gene expression values are unavailable, and finally using these predicted values to perform an association analysis with desired phenotypes. These methods largely differ in the predictive model employed. However most still use point estimates from their respective predictive models without considering various sources of uncertainty. These include uncertainty in SNP effect estimates in the predictive model, uncertainty in gene expression, and uncertainty in the genotype measurement (typically from imputation). We propose a predictive Bayesian mixture model to address some of these issues. Similar to the previously reported BayesR methodology, the use of a mixture model addresses the issue of multiple markers that contribute small effects to genetic variation. Early simulation studies have demonstrated predictive performance gains, measured by squared correlation, over other methodologies. For example, we observe a 15.3% relative increase in predictive performance over elastic net as implemented in PrediXcan. Uncertainty due to imputation, quantified by estimated R^2 , is incorporated into the distribution for each specific SNP, which has the effect of demanding increasing evidence of having a non-zero coefficient with decreasing imputation quality. Furthermore, the use of Gibbs sampling allows for straightforward calculation of both (unbiased) mean and variance of SNP effects, facilitating the use of a multiple imputation framework for association analysis to improve power and validity. Preliminary analyses of actual data reveal multiple genes associated with blood cell related traits.

2665T

Natural selection at the Adenylate cyclase 3 (ADCY3) gene. I. Yoshiuchi^{1,2}.

1) Diabetes, Medicine, and Genetics, Yoshiuchi Medical Diabetes Institute, Ikoma, Nara, Japan; 2) Diabetes and Medicine, Yoshiuchi Medical Diabetes Family Clinic, Ikoma, Nara, Japan.

Diabetes and obesity cause serious complications including cardiovascular disease, and become the global health burden in the world. Diabetes is strongly associated with obesity. There are highly heritabilities of diabetes and obesity. According to the thrifty gene hypothesis, natural selection for diabetes- and obesity-related genes is important during feast and famine because such genes control glucose and fat levels. Loss of function variations in the Adenylate cyclase 3 (ADCY3) gene are associated with obesity and type 2 diabetes. Mutations in ADCY3 cause monogenic severe obesity. ADCY3-deficient mice showed severe obesity, impaired insulin sensitivity, reduced physical activity and hyperphagia. Here, we aimed to uncover evidence of selection at ADCY3. We performed a three-step method to detect selection at ADCY3 using the 1000 genome population and HapMap population data. We used Wright's F-statistics as a measure of population differentiation, the long-range haplotype (LRH) to test extended haplotypes, and the integrated haplotype score (iHS) to detect selection at ADCY3. We observed high population differentiation at one ADCY3 height-adjusted BMI-associated SNP using Wright's F-statistics. The LRH and the iHS also showed evidence for selection at ADCY3. These findings provide evidence for natural selection at ADCY3. Further discoveries are warranted on the evolution of obesity- and type 2 diabetes - associated genes.

2666F

ChocoGen: Genetic ancestry and health in the Colombian Pacific. I.K.

Jordan^{1,2,3}, A.C. Conley^{2,3}, A.T. Chande^{1,2,3}, L. Rishishwar^{2,3}, E.T. Norris^{1,2,3}, J. Rowell¹, L. Mariño-Ramírez^{2,4}, A. Valderrama-Aguirre^{3,5}, M.A. Medina-Rivas^{3,6}.

1) School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA; 2) IHRC-Georgia Tech Applied Bioinformatics Laboratory, Atlanta, Georgia, USA; 3) PanAmerican Bioinformatics Institute, Cali, Valle del Cauca, Colombia; 4) National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland USA; 5) Biomedical Research Institute, Faculty of Health, Universidad Libre-Seccional Cali, Cali, Valle del Cauca, Colombia; 6) Centro de Investigación en Biodiversidad y Hábitat, Universidad Tecnológica del Chocó, Quibdó, Chocó, Colombia.

The ChocoGen research project was initiated to facilitate genomic studies of the predominantly Afro-Colombian population of Chocó, a state located on Colombia's Pacific coast. The project aims to (1) characterize the genetic ancestry of the population of Chocó, and (2) explore the relationship between ancestry and genetic determinants of health in the region. We compared genome-wide patterns of genetic ancestry for Chocó to six other admixed American populations, with an emphasis on a Mestizo population from the neighboring state of Antioquia. Chocó shows mostly African genetic ancestry (76%) with a nearly even split between European (13%) and Native American (11%) fractions, whereas Antioquia has primarily European ancestry (75%), followed by Native American (18%) and African (7%). Sample donors from Chocó self-identify as having more African ancestry than can be genetically inferred, as opposed to what we previously found for Antioquia, where individuals tend to overestimate levels of European ancestry. Despite the clear differences between Chocó and Antioquia at the level of continental ancestry, the two populations show overall patterns of subcontinental ancestry that are highly similar. When the subcontinental ancestral origins within Africa, Europe and the Americas are considered jointly, the populations of Chocó and Antioquia were found to be most closely related, to the exclusion of the other admixed American populations that we analyzed, pointing to a shared genetic legacy that underlies a common, unifying identity that binds the country. We investigated the role that genetic ancestry plays in the epidemiology of type 2 diabetes (T2D) for these two Colombian populations. Chocó has significantly higher predicted genetic risk for T2D compared to Antioquia, and the elevated predicted risk for T2D in Chocó is correlated with higher African ancestry. Despite its elevated predicted genetic risk, the population of Chocó has a three-times lower observed T2D prevalence than Antioquia, indicating that environmental factors better explain differences in T2D outcomes for Colombia. However, Chocó has substantially lower socioeconomic status (SES) than Antioquia, which is unexpected given that low SES is widely considered to be an environmental risk factor for T2D. Our exploration of the dietary and lifestyle factors that characterize low SES in Chocó illustrates how poverty can divergently manifest as a T2D protective factor in Colombia.

2667W

Cross-altitude analysis suggests a turning point at the elevation of 4,500m for polycythemia prevalence in Tibetans. *H. Zhang*^{1,2}, *Y. He*^{2,4}, *C. Cui*¹, *NA. Ouzhuluobu*¹, *NA. Baimakangzhuo*¹, *NA. Duoizhuoma*¹, *NA. Dejiqizong*¹, *NA. Bianba*¹, *NA. Gonggalanzi*¹, *Y. Pan*¹, *NA. Qula*¹, *M. Kang*¹, *NA. Cirenyangji*¹, *NA. Baimayangji*¹, *C. Bai*¹, *W. Guo*¹, *NA. Yangla*¹, *Y. Peng*², *X. Zhang*², *K. Xiang*², *Z. Yang*², *S. Liu*³, *X. Tao*³, *NA. Gengdeng*³, *W. Zheng*^{2,5}, *Y. Guo*^{2,5}, *T. Wu*², *X. Qi*², *B. Su*². 1) High Altitude Medical Research Center, School of Medicine, Tibetan University, Lhasa, China; 2) State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China; 3) National Key Laboratory of High Altitude Medicine, High Altitude Medical Research Institute, Xining, China; 4) Kunming College of Life Science, University of Chinese Academy of Sciences, Beijing, China; 5) College of Animal Science and Technology, Gansu Agricultural University, Lanzhou, China.

Although Tibetans are considered well adapted to high altitude, whether there is a limit of elevation they can live has not been tested. Tibetans have lower hemoglobin (Hb) levels compared with lowlanders moving to high altitude, which is considered a protection from overproduction of red cells (polycythemia), a condition that increases the risk of heart attack, stroke, and fetal loss during pregnancy. To detect if there is a turning point of changes of physiological traits with altitude, we collected Hb data from nearly 9,000 Tibetans, representing 20 geographic populations living in different elevations from 2,227m to 5,018m. The results indicated a clear turning point at the elevation of 4,500m where the Hb level and polycythemia incidence went up dramatically. We proposed that the elevation of 4,500m marks the altitude limit for Tibetans where the protective blunting effect on hemoglobin disappears.

2668T

High throughput pipeline for HLA typing identifies >800 novel alleles in populations from Africa. *T. Porter*^{1,2}, *M.O. Pollard*^{1,2}, *A.J. Mentzer*³, *A. Dilthey*⁴, *C. Pomilla*^{1,2}. 1) Global Health and Populations, Human Genetics, The Wellcome Sanger Institute, Cambridge, United Kingdom; 2) Department of Medicine, University of Cambridge, Cambridge, United Kingdom; 3) Wellcome Trust Centre for Human Genetics, University of Oxford, United Kingdom; 4) Institut für Medizinische Mikrobiologie und Krankenhaushygiene, Universitätsklinikum Düsseldorf.

The HLA Diversity in Africa Project aims to characterise the diversity of the HLA region amongst a wide variety of ethnic groups across Africa. Current HLA panels and reference databases are limited in data from diverse global populations, and are predominately European centric, limiting their utility in providing a global resource for characterising genetic diversity in this region. Our aim is to overcome this by providing a freely available resource (in the form of an imputation panel) to facilitate disambiguation of clinical types and help understand the role of the HLA region in response to infection, auto-immune disease and vaccine response in African and non-African populations. We present the largest African high resolution HLA panel to date. To this end, through extensive research and development, we have developed a robust and efficient high throughput pipeline using commercially available PCR primers to target HLA genes for PacBio sequencing for typing the Class I and Class II loci of 431 samples (4741 amplicons) from 19 ethno-linguistic groups across Africa, highlighting the utility of the pipeline for clinical applications. We show that our pipeline exceeds conventional sanger based typing approaches in accuracy and resolution of typing, at comparable cost. We also report more than 800 novel alleles discovered and validated using our pipeline, across the allele frequency spectrum. By sequencing the full length of these alleles we are able to distinguish them from highly homologous European alleles. These accurate high resolution types provide a rich resource for imputation, potentially improving power to discover associations within the MHC to facilitate HLA linked disease research. This also provides an important resource for disambiguation of clinical types obtained using sanger based technology (SBT). In summary, here, we present an optimised workflow for high resolution HLA typing among African populations using long-read sequencing approaches. Our pipeline allows inference of high levels of HLA diversity among populations across Africa, providing a resource for medical genetics and for better clinical HLA typing in global populations.

2669F

Different selected mechanisms attenuated the inhibitory interaction of KIR2DL1 with C2^{*} HLA-C in two indigenous human populations in Southern Africa. N. Nemat-Gorgani¹, H.G. Hilton¹, B.M. Henn², M. Lin², C.R. Gignoux³, J.W. Myrick³, C.J. Werely⁴, J.W. Granka⁵, M. Möller⁶, E.G. Hoal⁶, M. Yawata⁶, N. Yawata⁷, L. Boelen⁸, B. Asquith⁸, P. Parham¹, P.J. Norman¹. 1) Department of Structural Biology, Stanford University, Stanford, CA, CA., USA; 2) Department of Anthropology, UC Davis, CA, USA; 3) Colorado Center for Personalized Medicine and Department of Biostatistics, University of Colorado, Denver, USA; 4) SAMRC Centre for Tuberculosis Research, DST/NRF Centre of Excellence for Biomedical Tuberculosis Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Tygerberg, South Africa; 5) Department of Biology, Stanford University, Stanford, USA; 6) Department of Pediatrics, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Life Sciences Institute, National University of Singapore; and Singapore Institute for Clinical Sciences, Agency for Science, Technology and Research, Singapore; 7) Ophthalmology, Department of Medicine, Fukuoka Dental College, Japan; 8) Section of Immunology, Imperial College London, London, United Kingdom.

The functions of human natural killer (NK) cells in defense against pathogens and placental development during reproduction are modulated by interactions of killer cell immunoglobulin-like receptors (KIR) with HLA-A, -B and -C class I ligands. Both receptors and ligands are highly polymorphic and exhibit extensive differences between human populations. Indigenous to Southern Africa are the KhoeSan, the most ancient group of modern human populations, who have highest genomic diversity worldwide. We studied two KhoeSan populations, the Nama pastoralists and the #Khomani San hunter-gatherers. Comprehensive next-generation sequence analysis of *HLA-A*, *-B* and *-C* and all *KIR* genes identified 248 different *KIR* and 137 *HLA class I*, which assort into ~200 haplotypes for each gene family. All 74 Nama and 78 #Khomani San studied have different genotypes. Numerous novel *KIR* alleles were identified, including three arising by intergenic recombination. On average, KhoeSan individuals have 7-8 pairs of interacting KIR and HLA class I ligands, the highest diversity and divergence of polymorphic NK cell receptors and ligands observed to date. In this context of high genetic diversity, both the Nama and the Khomani San have an unusually conserved, centromeric *KIR* haplotype that has arisen to high frequency and is different in the two KhoeSan populations. Distinguishing these haplotypes are independent mutations in *KIR2DL1*, that both prevent KIR2DL1 from functioning as an inhibitory receptor for C2^{*} HLA-C. The relatively high frequency of C2^{*} HLA-C in the Nama and the #Khomani San appears to have led to natural selection against strong inhibitory C2-specific KIR.

2670W

Ancient and modern admixture enhanced the interactions of HLA class I with killer cell immunoglobulin-like receptors in East Asian populations. P.J. Norman^{1,5}, J. Zhen^{2,3}, G. Zhang², R. Chen², G. Sun², W. Hong², B. Zhu⁴, Q. Yu², N. Nemat-Gorgani⁶, L.A. Guethlein⁵, L. He², M. Tang⁶, X. Gao⁷, S. Cai⁶, M. Carrington⁷, H. Zou², P. Parham⁵, Z. Deng². 1) Division of Personalized Medicine, University of Colorado, Aurora, CO; 2) Immunogenetics Laboratory, Shenzhen Blood Center, Shenzhen, Guangdong, P. R. of China; 3) Central Laboratory, Shenzhen Baoan Maternal and Child Health Hospital, Shenzhen, Guangdong, P. R. of China; 4) Department of Forensic Genetics, School of Forensic Medicine, Southern Medical University, Guangzhou, Guangdong, P. R. of China; 5) Department of Structural Biology, Stanford University School of Medicine, Stanford, CA, USA; 6) Clinical Laboratory, Wuzhou Red Cross Hospital, Wuzhou, Guangxi P. R. of China; 7) Frederick National Laboratory for Cancer Research (FNLRC), Frederick, MD.

Natural killer (NK) cells are white blood cells essential for controlling human infection, cancer and fetal development. NK cell functions are modulated by interactions between inhibitory killer cell immunoglobulin-like receptors (KIR) and HLA-A, -B and -C molecules expressed on tissue cells. Both gene families are highly polymorphic; the *KIR* locus (chromosome 19q13.4) varies in gene number and sequence, and *HLA* (chromosome 6p21) are the most polymorphic human genes. All *HLA-C* alleles encode a KIR ligand and contribute to reproduction and immunity. In contrast, only some *HLA-A* and *-B* alleles encode KIR ligands and they are focused on immunity. By targeted high-resolution analysis of *KIR* and *HLA-A*, *-B* and *-C* genes, we show the Chinese Southern Han are significantly enriched for interactions between inhibitory KIR and HLA-A and -B. This enrichment has substantial input from population admixture, from both ancient and modern humans, who contributed multiple *HLA class I* and *KIR* haplotypes. Consequently, over 80% of Southern Han *HLA* haplotypes encode more than one KIR ligand. Complementing the high number of KIR ligands, the *KIR* locus combines a high frequency of genes expressing potent inhibitory KIR, with low frequency of those expressing activating KIR. The centromeric region of the *KIR* locus encodes strong and conserved inhibitory HLA-C specific receptors, and the telomeric region provides a high number and diversity of inhibitory HLA-A and -B specific receptors. We identified two *KIR* telomeric region haplotypes that have duplicated *KIR3DL1/S1* as present both in archaic humans and Southern Han. In all these characteristics, the Southern Han represent other East Asians, whose NK cell repertoires are thus enhanced in quantity, diversity and effector strength, likely through natural selection for resistance to endemic viral infections.

2671T

VarCount: Automated *in trans* variant calling reveals novel epilepsy genes and selection patterns between populations. A. Cox^{1,2}, F. Grady², G. Velez², V. Mahajan³, P. Ferguson², A. Kitchen², B. Darbro², A. Bassuk². 1) Yale School of Medicine, New Haven, CT; 2) University of Iowa, Iowa City, IA; 3) Stanford University, Palo Alto, CA.

Compound heterozygotes (CH) occur when different mutations at the same locus on both maternal and paternal chromosomes produce a recessive trait. It is suspected that undetected compound heterozygosity explains a significant portion of missing heritability for diseases and other phenotypes. Here we present the tool VarCount for the quantification of mutations at the individual level. We used VarCount to characterize homozygous and CH variants in the Epi4k epilepsy cohort and in the 1000 genomes participants. In Epi4k, we found enrichment for rare, CH mutations in six genes, including three involved in neuronal growth and development. PRTG, TNC and MACF1 are potential novel recessive epilepsy genes. Comparing the 1000 genomes superpopulations, we found significant differences in several genes, including immune response genes. NWD1 modulates androgen receptor signaling and is up-regulated in prostate cancer so recessive mutations may contribute to cancer development in African men. Mutations in the virus inhibitor APOBEC3B are enriched in East Asians, reflective of host-virus coevolution. SSC5D mutations are enriched in Africans, indicating that recessive mutations were selected against during global migrations and possibly as a factor of diet. Our results highlight that compound heterozygous mutations should be considered in sporadic epilepsy. Our findings also suggest that selection on recessive mutations occurred during the establishment of global populations, and the resulting genetic differences reflect host-virus co-evolution and underlie susceptibility to diseases like prostate and cervical cancer. .

2672F

Improved pathogenic variant localization using a hierarchical model of sub-regional intolerance. T.J. Hayeck^{1,2}, N. Stong¹, C.J. Wolock¹, B. Copeland¹, S. Kamalakara¹, D. Goldstein¹, A. Allen^{1,3}. 1) Institute of Genomic Medicine - Columbia University, New York, NY; 2) Department of Biostatistics - Columbia University, New York, NY; 3) Department of Biostatistics and Bioinformatics, Duke University, Durham, NC, USA.

Different parts of a gene can be of differential importance to development and health. This regional heterogeneity is also apparent in the distribution of disease mutations which often cluster in particular regions of disease genes. The ability to precisely estimate functionally important sub-regions of genes will be key in correctly deciphering relationships between genetic variation and disease. Previous methods have had some success using standing human variation to characterize this variability in importance by measuring sub-regional intolerance, i.e., the depletion in functional variation from expectation within a given region of a gene. However, the ability to precisely estimate local intolerance was restricted by the fact that only information within a given sub-region is used, leading to instability in local estimates, especially for small regions. Here, we overcome that limitation by borrowing information across genes using a Bayesian hierarchical model. We fit the model using standing human variation comprised of 123,136 individuals found in the genome Aggregation Database (gnomAD). We show that our approach stabilizes estimates, leading to lower variability and improved predictive utility. Specifically, using pathogenic mutations from ClinVar we show that our approach more effectively identifies regions enriched for pathogenic variants and are able to show significant correlations between sub-region intolerance and the distribution of pathogenic variation in disease genes, with AUC up to 0.86 for de novo missense variants in Online Mendelian Inheritance in Man (OMIM) genes. This later result immediately suggests that considering the intolerance of regions in which variants are found may improve diagnostic interpretation within exons. We also illustrate the utility of integrating regional intolerance into gene-level disease association tests with a study of known disease genes for epileptic encephalopathy.

2673W

Functional validation of missense variants in *VMAT1* reveals evolutionary trajectory of human psychology. D. Sato¹, Y. Ishii¹, T. Naga², M. Kawata¹.

1) Department of Ecology and Evolutionary Biology, Tohoku University, Sendai, Miyagi, Japan; 2) Department of Biomolecular Sciences, Tohoku University, Sendai, Miyagi, Japan.

VMAT1, which encodes vesicular monoamine transporter 1, regulates the transport of monoamine into neuronal synaptic vesicles, and several functional mutations in *VMAT1* have been reported to date. Among them, Thr136Ile belongs to a luminal loop domain, which interacts with G protein, and has been known to upregulate the monoamine transport and to affect our psychological conditions and/or psychiatric disorders, such as anxiety and bipolar disorders. In our previous study from evolutionary perspective (Sato & Kawata, in revision), it was estimated that the 136Ile had emerged later in modern humans and has been subject to strong selective pressure recently. Moreover, we found that this polymorphism does not exist in other mammals including the other great apes (they have monomorphic 136Asn instead), and that the substitution from the 136Asn to the 136Thr had been under positive selection in the archaic human lineage. Thus, it is quite interesting to investigate how the functional mutations have affected our psychology and personality traits during the human evolution. In the present study, we compared the functional differences among the 136Asn/Thr/Ile using fluorescent false neurotransmitters (FFNs), which act as fluorescent substrates of VMAT, to visualize the uptake of neurotransmitters. Plasmids encoding YFP-tagged VMAT1 were transiently transfected into HEK293T cells, and cells were then incubated with FFN206 for an hour. After washing, the fluorescence intensity of FFNs was quantified by a fluorescent microscope (with the ImageJ software) and a microplate reader. We also quantified the fluorescence intensity of YFP to estimate the expression level of YFP-VMAT1. The result showed that the uptake of FFNs in cells expressing the 136Asn or the 136Ile of VMAT1 was greater than that of the 136Thr. This suggests that the efficiency in monoamine transport had once decreased in the archaic human lineage but has increased again in recent ages. In summary, our study presents an unusual pattern of evolution of monoaminergic pathways in humans, suggesting the unique evolution of human psychology.

2674T

Understanding the hidden complexity of Latin American population isolates. J. Mooney¹, C. Huber², S. Service³, J.H. Sul⁴, C. Marsden⁵, Z. Zhang^{5,6}, C. Sabatti^{7,8}, A. Ruiz-Linares^{9,10}, G. Bedoya¹¹, N. Freimer³, K. Lohmueller^{1,2}, *Costa Rica/Colombia Consortium for Genetic Investigation of Bipolar Endophenotypes*.

1) Department of Human Genetics, University of California Los Angeles, Los Angeles, CA; 2) Department of Ecology & Evolutionary Biology, University of California Los Angeles, Los Angeles, CA; 3) Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California Los Angeles, Los Angeles, CA; 4) Department of Psychiatry and Biobehavioral Sciences, Semel Center for Informatics and Personalized Genomics, University of California Los Angeles, Los Angeles, CA; 5) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 6) Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY; 7) Department of Biomedical Data Science, Stanford University, Stanford, CA; 8) Department of Statistics, Stanford University, Stanford, CA; 9) Ministry of Education Key Laboratory of Contemporary Anthropology and Collaborative Innovation Center of Genetics and Development, Fudan University, Shanghai; 10) Aix-Marseille Univ, CNRS, EFS, ADES, Marseille, France; 11) Genética Molecular (GEN-MOL), Universidad de Antioquia, Medellín, Colombia.

Most population isolates examined to date were founded from a single ancestral population. Consequently, there is limited knowledge about the demographic history of admixed population isolates. Here we investigate genomic diversity of recently admixed population isolates from Costa Rica and Colombia and compare their diversity to a benchmark population isolate, the Finnish. These Latin American isolates originated during the 16th century from admixture between a few hundred European males and Amerindian females, with a limited contribution from African founders. We examine whole genome sequence data from 449 individuals, ascertained as families to build multigenerational pedigrees, with a mean sequencing depth of coverage of approximately 24X. We find that Latin American isolates have increased genetic diversity relative to the Finnish. However, there is an increase in the amount of identity by descent (IBD) segments in the Latin American isolates relative to the Finnish. The increase in IBD segments is likely a consequence of a very recent and severe population bottleneck during the founding of the admixed population isolates. Furthermore, the proportion of the genome that falls within a long run of homozygosity (ROH) in Costa Rican and Colombian individuals was significantly greater than that in the Finnish, suggesting more recent consanguinity in the Latin American isolates relative to that seen in the Finnish. Lastly, we found that recent consanguinity increased the number of deleterious variants found in the homozygous state, which is relevant if deleterious variants are recessive. Our study suggests there is no single genetic signature of a population isolate.

2675F

Ancient DNA sequencing shows population heterogeneity of Neolithic Central China. C. Li¹, J. He², X. Zhang³, F. Wu², X. Zhang¹, B. Feng², P. Underhill⁴, C. Zhang⁵, C. Wu⁶, F. Tang², B. Su¹, H. Zhao³, X. Wu², Y. Pang², Y. Huang^{2,6}. 1) Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, China; 2) Biodynamic Optical Imaging Center (BIOPIIC), School of Life Sciences, Beijing Advanced Innovation Center for Genomics, Peking University, Beijing 100871, China; 3) School of Archaeology and Museology, Peking University, Beijing 100871, China; 4) Department of Genetics, Stanford University, Stanford, CA94305, USA; 5) Institute of Oil and Gas, Peking University, Beijing 100871, China; 6) Peking-Tsinghua Center for Life Sciences, College of Engineering, Peking University, Beijing 100871, China.

Ancient human DNA sequences are highly desired for providing concrete evidence for a better reconstruction of the past life, especially the genetic relationship and interaction among prehistoric populations. Most recently, genome-wide sequencing data of an Upper Paleolithic human of East Asia has been successfully analyzed, whereas only one mitochondrial genome sequence of Neolithic human is reported to provide a genetic link to southern China as well as insights concerning early Austronesian expansion. Nuclear genome information of Neolithic human in East Asia, especially the central China which is one of the main hubs of the development of civilization in human history, has been still deficient for a long time due to the scarcity of samples and their unfavorable preservation environments, limiting our understanding of human genetic relationships and interactions in that period. We used high-throughput sequencing to study ancient DNA samples from the remains of 41 Neolithic human individuals, excavated from a 5,000-y-old secondary burial site in Baligang M13, China. The majority of these samples contained no more than 0.1% of endogenous human DNA, thus we explored various protocols to construct sequencing libraries from these poorly preserved aDNA samples from such a mass burial. We developed various nucleic acid probe-mediated approaches to enrich the mtDNA and obtained 13 complete MT genome sequences which clearly ruled out a direct maternal relationship between any two individuals, predicting the post-matriarchal structure of the late Neolithic society in China. The relatively diversified haplogroup types imply these Neolithic humans living in the Central Plain of China already a mixture of distantly related ethnic and genetic groups rather than a small population derived from a single founder. Furthermore, the quality of DNA from a petrous bone is feasible to detect a few single nucleotide variations associated with haplogroup sites in chromosome Y, helping us to unravel a close genetic relationship of the Neolithic Baligang M13 individuals to present people in China and East/Southeast Asia, and suggesting the consistence between paternal and maternal phylogeny as well. The genomic interrogation of this Neolithic burial facilitates our understanding of 5,000-y-old prehistoric culture in central China. .

2676W

Unique and conserved gene pool provides higher risk of rare genetic disorders in populations of Jammu and Kashmir-India. S. Sharma¹, E. Rai¹, A. Angural¹, A. Sapolia¹, S. Razdan², K.K. Pandita³. 1) Human Genetic Research Group, School of Biotechnology, Shri Mata Vaishno Devi University, Katra, Jammu, India; 2) Neurology Clinic, Bhagwati Nagar, JAMMU, India; 3) Health Clinic, H. No. 62, Lane 11, Swam Vihar, Muthi, JAMMU, India.

Jammu and Kashmir (J&K) is the North most State of India, located in Himalayan region and geographically isolated. It is anticipated that this region has acted as corridor to various invasions and migrations, to and from mainland India, Eurasia or northeast Asia but underrepresented in any of the study so far in literature. The isolated gene pool of the region on one had provides Uniqueness to the population groups of the region, at the same time various endogamous and consanguineous practices provide higher risk of genetic diseases in the populations. These disorders are predominantly rare, poorly understood and most often remain uncharacterized or patients are misdiagnosed due to lack of specific clinical resources. Understanding the basics of inheritance is essential in such cases as it helps to figure out the plausibility of a disorder as an inherited or genetic disease. Though identification and characterization of such disorders is complicated, Next generation Sequencing has come up as a tool in recent times and is of great help. We have devised a "Bottom UP" approach and clubbed it with NGS/WES, as an affective method in understanding uncharacterized Mendelian Disorders in the region. It may be of great help especially in developing countries and regions like J&K where such disorders remain undiagnosed/misdiagnosed due to lack of specialized testing. We have collected huge number of highly extended families representing various rare genetic disorders and trying to elucidate the genetic cause and biology of the diseases in these families with this approach.

2677T

Reconstructing the peopling of old world south Asia: From modern to ancient genomes. N. Rai¹, K. Thangaraj². 1) Ancient DNA Laboratory, Birbal Sahni Institute of Palaeosciences, Lucknow, Uttar Pradesh, India; 2) Centre for Cellular and Molecular Biology, Hyderabad, Telangana, India.

South Asia was one of the first geographic regions to be peopled by modern humans after their African exodus. Today, the diverse ethnic groups of South Asia comprise an array of tribes, castes, and religious groups, who are largely endogamous and have hence developed complex, multi-layered genetic differentiation. From such a complex structure, several questions have stood out from the research of our group and others that are only beginning to be resolved using modern sequencing techniques and targeted sampling of populations and archaeological specimens. Here, for the first time we have used ancient genomics approach to understand the deep population ancestry of Indian Sub- continent. Despite the rich sources available of modern Indian populations, success from ancient DNA specimens in the subcontinent have been limited. We have successfully analysed several museum samples and fresh excavation from the different part of India which provides us a wonderful opportunity to be able to relate these modern populations genetically with those in the past and build complex models of population mixture and migration in India. Using ancient genomics data from the human remains who have lived about 4-5 thousand years before present in North West and South of India, we are trying to understand the population history of Iron age people and their genetic relation with the North West of Indians and Iranian Farmers. Furthermore, we are providing a solid Genetic evidence that substantiates archaeological and linguistic evidence for the origins of Dravidian languages and the language of the Indus valley people.

2678F

Reconstructing the health landscape of a medieval hospital cemetery: A holistic interdisciplinary approach. C.L. Scheib^{1,2}, A.W. Wohns³, X. Ge⁴, J.S. Bates², P. Maheshwari-Aplin², B. Haines², S.A. Inskip², J. Dittmar², C. Cessford², B. Mulder², A.A. Rose², M. Metspalu¹, T.C. O'Connell², P. Mitchell², J.T. Stock⁵, J.E. Robb², T. Kivisild^{1,2}. 1) University of Tartu, Tartu, Estonia; 2) University of Cambridge, Cambridge, United Kingdom; 3) University of Oxford, Oxford, United Kingdom; 4) University of Manchester, Manchester, United Kingdom; 5) University of Western Ontario, London, Ontario, Canada.

Established for the care of the poor and infirm, with the exception of pregnant women, lepers, the wounded, crippled and insane, the assemblage of 400 internments (c. 1204-1511) from the Hospital of St. John, Cambridge, provides a laboratory for testing osteological/aDNA methods and for understanding the health and diseases of a broadly representative British medieval population as it experienced the spread of Black Death from continental Europe during 1347 – 1349. Four hundred skeletons from the cemetery site were assessed for osteological traits and paleopathology. A subset was sampled for isotopic and ancient DNA (aDNA) analysis and compared to a time-transect of sites in the local area. DNA was extracted from teeth, built into double-stranded Illumina-compatible libraries and sequenced on the NextSeq500 75-cycle single-end platform. The raw libraries were screened in collaboration with the Jena Max Planck Institute for the Science of Human History using MALT. Surprisingly, though the cemetery's prime usage overlaps the main plague period, no skeletons (n = 86, 0%) tested positive for the presence of *Y. pestis* in the teeth, while a contemporaneous site less than 500 meters away contained at least 5 individuals (n = 26, 19%) whose teeth tested positive for *Y. pestis* aDNA. In line with the charter, within the St. John Hospital cemetery, no signs of active *M. leprae* infection (osteologically or aDNA) were found; however, osteological evidence of possible *M. tuberculosis* infection is abundant (post-cranial remains are currently being extracted and screened for pathogen aDNA) as well as a number of trauma. Ancient DNA scans did, however, find evidence of opportunistic infections, pneumonia and septicemia, more indicative of 'diseases of the elderly' that effect immunocompromised individuals in today's hospital settings and analyses of the human genome provide insight into heritable disease risk present in the population. By combining the aDNA with dietary isotopic information and skeletal analysis, this collection provides a valuable experiment in holistic Bioarchaeology and an opportunity to explore the relationships between diet, culture, and heritable disease risk on the discovery and expression of heritable and communicable disease in the archaeological record.

2679W

Multi-omics approach to elucidating the C57BL mouse substrains. Y. Ren, A. Palmer. University of California, San Diego, San Diego, CA.

Since C.C. Little initially generated the C57BL inbred strain in the 1920's, the inbred substrain C57BL/6 (B6) became the most frequently used mouse strain in biomedical research. B6 mice are used in a wide variety of research areas including cardiovascular biology, developmental biology, diabetes and obesity, genetics, immunology, neurobiology, and behavioral research. The popularity of B6 inbred mice led to the establishment of many colonies at different vendors and academic institutions around the world. Once a new B6 substrain is established, spontaneous mutations will arise in both the original colony and the new colony, and a subset of those mutations will spread through the colony by genetic drift and become fixed. The longer individual substrains are separated from each other, the greater the number of genetic differences between them. These genetic differences can lead to significant phenotypic and gene expression differences. A number of studies have reported these phenotypic differences among the C57BL/6 substrains, especially related to several behaviors and tolerance for glucose, alcohol, and drugs. In order to characterize the substrain differences, we have performed whole-genome sequencing and RNA-sequencing on 9 commonly used B6 substrains and 5 commonly used B10 substrains. This has led to the discovery of many substrain specific accumulated mutations (SSAM), substrain specific accumulated gene expression (SSAGE), and cis-regulatory elements (CRE) that are important for any research using B6 substrains.

2680T

Visualization of population structure history with Scale-Invariant Geometric Data Analysis (SIGDA). M. Robinson, A. Zimmer, T. Farrah, D.E. Mauldin, N.D. Price, L. Hood, G. Glusman. Institute for Systems Biology, Seattle, WA.

We present exploratory data analyses and visualizations of single nucleotide variants (SNVs) from the 1000 Genomes cohort on the most relevant dimensions of population history: time, population structure, and chromosomal position. We demonstrate that our algebraic method, Scale-Invariant Geometric Data Analysis (SIGDA), provides the following improvements over a principal components analysis (PCA, often applied to population structure through EIGENSTRAT) of the same SNV data: 1. Like canonical correlation analysis (CCA), a leading method for contingency table analysis, SIGDA maps individuals (rows in the data matrix) and SNVs (columns) against population structures on the same scale. In contrast, PCA is only designed to map individuals on the coordinate axes; SNVs are analyzed in very different, incompatible coordinates, and are rarely visualized in a population structure analysis. 2. Due to the common coordinate system, SIGDA provides intuitive visualizations of SNVs as markers of population structure by allele frequency, a proxy for allele age. Visualization of an individual's alleles illustrates their individual population history against the background of genotypes in the cohort. SNV coordinates from SIGDA can also be visualized against their chromosomal positions, providing a visualization of an individual's population history along each chromosome in a manner similar to chromoPainter. 3. Using a resampling strategy, we show that SIGDA identifies more robust elements of population structure than PCA does on the same data. 4. In an analysis of the entire 1000 Genomes cohort, SIGDA robustly identifies population structures among the South Asian (SAS) populations which are only observed by PCA when the cohort is restricted to the SAS sub-cohort. This reflects the fact that PCA is more affected than SIGDA by within-cohort differences in effective population size. Importantly, these analyses reflect the algebraic structure of the data, independent of any population genetic model, and we anticipate that future work will allow hypothesis testing of population size, mutation rates, migration rates, and selection coefficients in an algebraic statistics context. Finally, SIGDA is a generic exploratory data analysis method; it provides unique advantages for analysis gene expression data, contingency tables, and has applications across all of data science. This work is supported by NIH 1U54EB020406 and the MJ Murdock Charitable Trust.

2681F

Detecting polygenic signals and molecular mechanisms of adaptation to local environments. *F. Luca¹, A. Findley¹, D. Witonsky², R. Pique-Regi¹.* 1) Wayne State University, Detroit, MI; 2) University of Chicago, Chicago, IL.

Human migration led to the settlement of population groups in varied environmental contexts worldwide. The extent to which adaptation to local environments has shaped human genetic diversity is a longstanding question in human evolution. Recent studies have suggested a role for archaic human introgression in modern human adaptation to environmental pressures such as pathogen load. With a few notable exceptions, most tests for polygenic adaptation in humans have relied on SNP associations with organismal-level phenotypes, commonly measured in GWAS. However, functional genomic studies have demonstrated that variation in gene expression across individuals and in response to environmental perturbations is a main mechanism underlying complex trait variation. Here we consider gene expression response to in vitro treatments as a molecular phenotype to identify genes and regulatory variants that played an important role in adaptations to local environments. We apply Berg and Coop's test for polygenic adaptation to 52 populations in the Human Genome Diversity Panel and we test for selection on eQTLs for genes differentially expressed. We first validate our approach using Crohn's disease as an example: we consider differentially expressed genes between healthy and diseased ileal tissue samples and show that selection on eQTLs from GTEx for these genes mimics the signal observed when considering SNPs associated with Crohn's disease in GWAS. We then expand to eQTLs for genes differentially expressed in a panel of 52 cellular environments, resulting from 5 cell types and 35 treatments, including hormones, vitamins, drugs, and environmental contaminants. We find evidence of polygenic adaptation for response to the antibacterial triclosan in vascular endothelial cells and iron in PBMCs (FDR = 5%). We then investigated if Neanderthal introgression in the human genome may have contributed to adaptation to local environments through gene regulatory changes. Considering the same set of eQTLs and gene expression responses as described above, we found an enrichment for genes regulated by Neanderthal SNPs in genes differentially expressed in response to 16 treatments, including dexamethasone, caffeine, and vitamin D. These findings expand the set of environments where archaic introgression may have contributed to modern human adaptations to local environments.

2682W

Self-reported race/ethnicity as a function of genetic ancestry in the Kaiser Permanente (KP) Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort. *Y. Banda¹, D. Ranatunga², C. Schaefer³, N. Risch^{1,2,3}.* 1) Institute for Human Genetics, University of California, San Francisco, San Francisco, CA; 2) Kaiser Permanente Northern California Division of Research, Oakland, CA; 3) Department of Epidemiology and Biostatistics, University of California, San Francisco, CA.

In a vast number of epidemiological and genetic association studies of disease phenotypes, ethnic origin has been shown to be associated with the phenotypes. In most biomedical research studies, either self-reported race/ethnicity or genetic ancestry data, or both, is used for ethnic origin assessment. Preference of proxy depends on the type of analysis being conducted and the nature of the question being addressed. Of particular importance is the self-identification of individuals with mixed genetic ancestry, in terms of how many and which race/ethnicity categories are endorsed. We analyzed the correspondence between self-reported race/ethnicity and genetic ancestry in a sample of 97,671 individuals who are participants in the KP GERA cohort. Genetic ancestry was determined from a set of 43,988 SNPs from GWAS arrays. We focused on individuals who had at least 5% genetic ancestry from more than one continent, and calculated the rate of self-identification with a particular race/ethnicity group as a function of genetic ancestry. We observed that rates of self-identification as African American, East Asian or Latino rise dramatically with a modest amount of African, East Asian and Native American genetic ancestry, respectively. By contrast, the rate of self-identification as White rises only when the European genetic ancestry is very high. This indicates that the majority of people who are genetically mixed, even those with primarily European genetic ancestry, self-identify with the minority ancestral group. We also examined the rate of endorsing more than one race/ethnicity category. Here we observed a difference between those with more recent admixture (e.g. in the parental or grandparental generation) versus older admixture, with the more recently admixed individuals being more symmetric in their self-reported race/ethnicity and endorsing more than one category. These results reflect multiple social and political factors, including social homogeneity, migration, mating patterns, and historical considerations of genetically admixed individuals in a majority white population. These results also have implications for the interpretation of self-reported race/ethnicity covariates in epidemiologic studies.

2683T

Study of Touch DNA in simulated situations with computer keyboards for forensic purposes. *F.T. Goncalves, D.O. Francisco, C. Fridman.* Universidade de Sao Paulo Faculdade de Medicina, Sao Paulo, Sao Paulo, Brazil.

In a crime scene, partial fingerprints can be used to obtaining genetic profiles based on genomic polymorphisms, assisting the elucidation of criminal investigations. Variables may interfere with the obtaining of profiles, as the type of the surface where the biological material was deposited. This study aimed to analyze the feasibility of obtaining STR profiles from fingerprints left on computers keyboards. Eight simulated experiments were conducted in pairs with four women and one man in two different keyboards. Each keyboard were used daily by a woman (primary donor) and then were manipulated during five minutes by the secondary donor (a woman or a man). The genetic material provided from the keyboards were collected with the double swab method and the DNA was extracted by two different methodologies: Swab Solution (Promega Corporation) and DNA IQ (Promega Corporation). The samples were quantified and then amplified by multiplex PCR using the kit for Human Identification *PowerPlex® Fusion* System (Promega Corporation) that contain 24 STR markers. Results showed that the recovery of the genetic material deposited on the substract was higher when a man was present in the mixture, independent of the extraction method. Besides, in all experiments the profile predominant in the mixture was provided by the primary donor. Stochastics effects were observed in all profiles, impairing the determination of the real number of donors present in the samples. Therefore, our group has been working to improve these results and and assist the elucidation of cybercrimes.

2684F

Systematic detection of conserved and divergent brain proteins during human evolution. *G. Dumas^{1,2,3,4}, S. Malesys^{1,2,3,4}, T. Bourgeron^{1,2,3,4}.* 1) Institut Pasteur, Human Genetics and Cognitive Functions Unit, Paris, France; 2) CNRS UMR3571 Genes, Synapses and Cognition, Institut Pasteur, Paris, France; 3) University Paris Diderot, Sorbonne Paris Cité, Human Genetics and Cognitive Functions, Paris, France; 4) Centre de Bioinformatique, Biostatistique et Biologie Intégrative (C3BI, USR 3756 Institut Pasteur et CNRS), Paris, France.

The human brain organization differs from that of other primates, but the mechanisms accounting for its evolution remain unclear. Using a systematic analysis of all protein-coding genes in non-human primates and in hominins such as Neanderthal, we confirm that brain-related genes are among the most conserved genes of the human genome. Although brain diseases are expectedly enriched in highly conserved genes, we also highlight several genes associated with neurodevelopmental conditions such as micro/macrocephaly, autism or dyslexia that display a high degree of coding-sequence divergence during primate evolution. These genes involved in key processes of cortical and cerebellar development might have played roles in the specificity of the brain of modern humans. This resource is available online and can be used to estimate the degree of evolution of a single gene or a set of genes.

2685W

Feeding behaviors determine acidic chitinase mRNA levels in mammalian and poultry stomachs. F. Oyama¹, A. Kashimura¹, A. Kikuchi¹, H. Masuda¹, R. Miyahara¹, Y. Hiruma¹, S. Wakita¹, M. Ohno¹, M. Sakaguchi¹, Y. Sugahara¹, E. Tabata^{1,2}. 1) Department of Chemistry and Life Science, Kogakuin University, Hachioji, Tokyo 192-0015, Japan; 2) Research Fellow of Japan Society for the Promotion of Science (DC1)Koujimachi, Chiyoda-ku, Tokyo 102-0083, Japan.

Chitin, a linear β -1, 4-linked polymer of *N*-acetyl-D-glucosamine (GlcNAc), is the second most abundant polysaccharide in nature and functions as a major structural polymer in many organisms. Recently, we reported that acidic chitinase (Chia) is highly expressed in mouse, chicken and pig stomach tissues and that it can digest chitin in the respective gastrointestinal tracts (GIT). However, the gene expression and enzymatic activity level in other mammals are still unknown. Chia mRNA expression levels were much lower in bovine (herbivores) and dog (carnivores) than those in mouse, pig and chicken (omnivores) stomachs. Moreover, the chitinolytic activities of recombinant bovine and dog Chia enzymes were slightly but significantly lower when compared with those of mouse, pig and chicken Chia. Thus, feeding behavior seems to be directly linked to Chia expression levels as well as chitinolytic activity of the enzyme in the particular animals. .

2686T

New insights into the genetic basis and evolutionary history of lactase persistence in Africa. A. Ranciaro¹, M.C. Campbell², E. Eyerman¹, S. Thompson¹, W. Beggs¹, S.W. Mpoloka³, G.G. Mokone⁴, T. Nyambo⁵, D.W. Meskel⁶, G. Belay⁶, M. Meredith Yeager⁷, S. Channock⁸, S.A. Tishkoff^{1,9}. 1) Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA; 2) Department of Biology, Howard University, Washington, DC 20059, USA; 3) Department of Biological Sciences, University of Botswana, Gaborone, Botswana; 4) Department of Biomedical Sciences, University of Botswana School of Medicine, Gaborone, Botswana; 5) Department of Biochemistry, Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania; 6) Department of Biology, Addis Ababa University, Addis Ababa, Ethiopia; 7) Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, MD 20892, USA; 8) Frederick National Laboratory for Cancer Research, Leidos Biomedical Research Inc., Frederick, MD 21701, USA; 9) Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA Department of Biology, School of Arts and Sciences, University of Pennsylvania, Philadelphia, PA 19104, USA.

Lactase persistence (LP) is a recent adaptive trait in humans and has evolved mainly in populations that practice milk production. Prior studies of African populations have identified novel common variants in intron 13 of the *MCM6* gene (a known candidate enhancer region for *LCT*) associated with LP primarily in pastoralists. While these variants account for a large proportion of the phenotypic variance, there may be additional polymorphisms that contribute to the LP trait. To test this hypothesis, we resequenced 555bp of intron 13 of *MCM6*, in ~1,500 African individuals from diverse populations in Ethiopia, Tanzania and Botswana. We also genotyped these same individuals using the Illumina Omni 5M SNP array, and integrated our sequencing data with the 5M genotype data to detect signatures of selection. In addition, we performed the lactose tolerance test for a subset of ~450 individuals. Based on our analyses, we identified for the first time the presence of the C-14010 and G-13915 variants in the Hadza hunter-gatherers from Tanzania. We also observed the G-13915 variant in the Burunge and Maasai populations from Tanzania, which has never previously been reported. Furthermore, we confirmed the presence of the G-13907 and the G-13915 alleles in Ethiopia. In addition, we detected a strong association between LP and genomic regions on several chromosomes in the Maasai pastoralists based on Fisher's Exact Test, suggesting candidate regions that might contribute to LP in this population. Overall, this large-scale study offers new insights into the genetic basis and evolutionary history of LP in Africa.

2687F

Evolutionarily conserved genes regulating hypoxia tolerance. D. Zhou¹, J. Xue¹, Y.H. Hsiao¹, T. Stobdan¹, G.G. Haddad^{1,2,3}. 1) Dept. of Pediatrics, University of California San Diego, La Jolla, CA; 2) Dept. of Neuroscience, University of California San Diego, La Jolla, CA; 3) The Rady Children's Hospital, San Diego, CA.

Hypoxia represents a critical element in the pathogenesis of many human diseases, such as ischemic stroke, myocardial infarction and solid tumors. Understanding the mechanisms regulating hypoxia tolerance or susceptibility is essential in developing effective strategies for medical interventions. In this regard, we have generated a hypoxia-tolerant *Drosophila melanogaster* strain through laboratory-directed evolution over >300 generations and analyzed the whole genomes of these flies. Due to the fact that the environmental conditions can be well controlled in the laboratory evolution experiments, the main variable in our experiments is the level of oxygen, which constitutes the only selection pressure applied. Therefore, these hypoxia-tolerant flies are a powerful model to delineate the mechanisms regulating hypoxia responses, and not a mixture of stressors. In parallel, we took advantage of the natural evolution of humans at different high altitude regions of the world and analyzed the whole genomes of Ethiopian and Andean highlanders. Through a comparative genomic approach, we obtained a group of evolutionarily conserved genes (28 human genes/23 *Drosophila* homologs) that are potentially involved in regulating hypoxia tolerance in human highlanders and the hypoxia-tolerant flies. Indeed, we confirmed the role of a subset of these candidate genes as regulating hypoxia tolerance experimentally. We discovered that ubiquitous knock-down of the expression of genes (Dm/Human), i.e., *gm/GATA3*, *Mkk4/MAP2K4*, *pyd/TJP1*, *RapGAP1/RAP1GAP2* and *shep/RBMS3*, dramatically enhanced hypoxia tolerance *in vivo* in *Drosophila*. These mechanisms have a strong potential to be translated into novel targets for developing therapeutic strategies to treat hypoxia-related diseases.

2688W

Detecting adaptive introgression in human populations without knowledge of the archaic samples. A. Akbari, V. Bafna. UC San Diego, La Jolla, CA.

Motivation. Most approaches that capture signatures of selective sweeps in population genomics data do not identify the specific mutation favored by selection. We recently proposed a method in Nature methods (Akbari et. al. 2018), iSAFE (for “integrated selection of allele favored by evolution”), that enables researchers to accurately pinpoint the favored mutation in a large region. iSAFE revealed a specific genomic pattern in some putative targets of the selective sweep, including EDAR1 and LCT loci, that cannot be explained by hard/soft Selective Sweep (SS), or even multiple selective sweeps. Here, we show that Adaptive Introgression (AI) or Adaptation on a Recombination Suppressor (ARS) mechanism (like chromosomal inversion) can create such a pattern in the population. We present a novel method, based on properties of iSAFE, to accurately infer the favored haplotype and label the target region whether it is neutral, SS, AI, or ARS. This method does not require, archaic samples, knowledge of demography, the phenotype under selection, or functional annotations of mutations. **Method.** The iSAFE method can accurately pinpoint the favored mutation in a large region (~5 Mbp) by using a statistic derived solely from population genetics signals. iSAFE treats each mutation as a binary classifier, i.e. carriers and non-carriers. Therefore, if two mutations are fully linked they get exactly the same iSAFE score regardless of their chromosomal position in the target window. This enables us to easily infer the favored haplotype. We use length and number of mutations on the favored haplotype to estimate the branch length of the favored haplotype and separate SS, AI, and ARS. Finally, we use an extension of the coalescent theory to infer the divergence time of the human population and, possibly unknown, archaic hominin in the AI cases. **Result.** We validated our method with extensive simulations and also on known examples of AI in the human population, like EPAS1 in Tibetans highlanders. The results, on both simulations and real data, are very accurate and promising. We also found some novel loci in all populations of 1000 genome projects that are predicted, with a high confidence, to be Adaptive Introgression from a possibly unknown archaic hominin.

2689T

De novo mutation rates from 1465 ancestrally diverse trios highlight rapid founder population changes and inter-ancestral conservation in mutational processes. M.D. Kessler¹, J. Perry¹, M. Daya², N.L. Heard-Costa³, B. Cade³, H. Wang³, J. Ziniti³, S. Datta⁴, S.T. Weiss⁵, K. Barnes⁶, S. Redline³, V. Ramachandran⁵, A. Johnson⁴, R. Mathias⁴, J. O'Connell¹, X. Tian², B. Mitchell¹, J.G. Wilson⁶, D.A. Nickerson², S. Zoellner⁶, S. Browning⁷, G. Abecasis⁸, R. Hernandez¹⁰, T.D. O'Connor⁹, TOPMed Consortium. 1) University of Maryland Baltimore, Baltimore, MD; 2) University of Washington, Seattle, WA; 3) Harvard Medical School, Boston, MA; 4) National Heart, Lung and Blood Institute's The Framingham Heart Study, Framingham, MA; 5) Boston University School of Medicine, Boston, MA; 6) University of Mississippi Medical Center, Jackson, MS; 7) University of Houston Clear Lake, Houston, TX; 8) University of Colorado Denver, Denver, CO; 9) University of Michigan, Ann Arbor, MI; 10) University of California San Francisco, San Francisco, CA.

De novo Mutations (DNMs), or mutations that appear in an individual despite not being seen in their parents, are an important source of genetic variation whose impact is relevant to studies of human evolution, demography, and disease. Utilizing the largest high-coverage human whole genome sequencing dataset to date as part of the Trans-Omics for Precision Medicine (TOPMed) program, we directly estimate and analyze DNM counts, rates, and spectra from 1465 trios across an array of diverse human populations. We find a weak but significant positive correlation between genome-wide heterozygosity and DNM count ($R^2=0.005$, $P<0.02$), though we do not find significant differences in DNM rate between individuals of European, African, and Latino ancestry. When we explore the intersect of DNMs and ancestral segments within admixed populations, we do not find significant differences in DNM accumulation rates between ancestral origins. However, interestingly, we do find significantly fewer DNMs in Amish individuals compared with other Europeans, even after accounting for parental age and sequencing center (Amish mean = 54.78, European mean = 59.15, p -value < 0.005). Specifically, we find significant reductions in the number of A→G/ T→C mutations in the Amish that seem to underpin their overall reduction in DNMs (p -value < 0.015). Given the lack of local differences in DNM rate seen across ancestrally distinct genomic segments, and that the correlation between heterozygosity and DNMs is too weak to explain the degree of difference in DNMs seen between the Amish and other Europeans, these data suggest that either environmental variables or uncommon genetic elements have led to a reduced accumulation rate of DNMs in the Amish.

2690F

Human prehistoric demography revealed by polymorphic pattern of CpG transitions. X. Liu. Human Genetics Center, University of Texas School of Public Health, Houston, TX.

Background: Thanks to the advance of archaeological and genetic studies of human populations, we have learned a lot about our own prehistoric demographic events, including the out-of-Africa migration and the eventual distribution throughout the continents. However, many details are still missing and under hot debate. One missing yet important piece of the puzzle is when and how human populations expand during the Holocene, especially around the time of Neolithic Revolution. **Materials and Methods:** We inferred demographic histories from 1 to 40 thousand years ago (kya) for 20 population samples from the 1000 Genomes Project phase 3 dataset, using a newly developed model-flexible method called Stairway plot with 36 million non-coding CpG sites genome-wide. Simulation studies were conducted to test the robustness of the inference, and 4 independent population samples were used to verify some of our findings. **Results:** Our results show many population growth events likely due to the Neolithic Revolution (i.e. shifting from hunting and gathering to agriculture and settlement): Han Chinese experienced a dramatic ~10 fold population growth around 8 kya to 12 kya; some South European and South Asian populations also began their long-term population growth around 10 kya but in a more gradual fashion; British and Western European populations began their takeoff around 6-7 kya. The potential agriculture-associated population growth of Luhya in Webuye, Kenya (LWK) came relatively late (no earlier than 3 kya) compared to other African populations, in contrast to the assumption that they are the direct descendants of Bantu-speaking immigrants from West Africa. We also observed several population growth events dated before the introduction of agriculture. **Conclusion:** In this study, we estimated detailed recent human demographic prehistories using a model-flexible approach with the observed polymorphic pattern of CpG transitions. Those prehistories help to shed light on the origins of different populations and the impact of agriculture on human demography. As an exploratory study, our results can be used not only for testing existing hypotheses, as we demonstrated, more importantly they provide new findings and hypotheses for future studies.

2691W

Full genome sequencing of Native American ancestry individuals reveals adaptation to high altitude in the Andes. M. Muzzio¹, A. Dobbyn², J.M.B. Motti³, M.C. Yee⁴, S. Kohli², K. Slivinski², M.R. Santos¹, V. Ramallo¹, E.L. Alfaro⁵, J.E. Dipierri⁵, G. Bailliet¹, C.M. Bravi¹, E.E. Kenny². 1) IMBICE CONICET UNLP, La Plata, Buenos Aires, Argentina; 2) Mount Sinai School of Medicine, New York, EEUU; 3) CONICET - Laboratorio de Ecología Evolutiva Humana, NEIPHA-FACSO-UNCPBA; 4) Stanford University, California, EEUU; 5) INECHO (Instituto de Ecorregiones Andinas) UNJu-CONICET. Instituto de Biología de la Altura; Universidad Nacional de Jujuy, San Salvador de Jujuy, Jujuy, Argentina.

We carried out Whole Genome Sequencing to 4X coverage on blood samples from 14 high-altitude Andean individuals and 11 low-altitude Chaqueño individuals with over 90% Native American ancestry on the Illumina HiSeq2500 platform. We called variants using a pooled analysis in GATK, implementing Unified Genotyper (UG) as the variant calling method. We identified samples with high missingness rates and re-ran the GATK pipeline excluding these individuals, leaving us with 13 highlanders and 9 lowlanders for all subsequent analyses. After filtering, we had a total of 5,232,922 variants remaining. We annotated our variants for genic and functional status with ANNOVAR. FST statistics were calculated using vcftools. The resulting FST statistics were winsorized, and p-values were generated for the resultant exponentially distributed statistics in R. We used DEPICT (Data-driven Expression-Prioritized Integration for Complex Traits) to conduct gene set prioritization and tissue enrichment analyses. We used input variants at multiple FST FDR thresholds (0.01, 0.1, 1 and 5%) and ANNOVAR annotations. At an FDR threshold of 0.01 percent, we identified four genomic regions that are highly differentiated between the low and high altitude populations and a number of genes in these regions could potentially be relevant in the response to high altitude hypoxia. Expanding the analysis to include windows at an FDR threshold of 0.02 percent, we identified an additional nine genomic regions, one of which has previously been implicated in a study of Chronic Mountain Sickness in an Andean population. Finally, in order to gain further biological insights from our high FST variants, we conducted tissue and gene set enrichment analyses using DEPICT. Taking variants at an FST FDR threshold of 1 percent, using all nonsynonymous/stoploss/stopgain variants as input, the tissue enrichment analysis shows the most significant (nominal P-value FDR<5%) enriched tissue to be fetal blood. Taking variants at an FST FDR threshold of 5 percent, again using all nonsynonymous/stoploss/stopgain variants as input, the tissue enrichment analysis shows the three most significantly enriched tissues to be fetal blood, esophagus, and blood, and the most significant enriched gene sets to include multiple lung and oxygen-related processes, such as reactive oxygen species metabolic process (GO:0072593), abnormal vascular endothelial cell physiology (MP:0004003), and superoxide metabolic process (GO:0006801)..

2692T

Distribution of local ancestry and evidence of positive selection in Brazilian individuals. R. Secolin^{1,5}, A. Mas², L.R. Araújo², F.R. Torres¹, T.K. Araujo¹, M.L. Santos¹, C.S. Rocha^{1,5}, B.S. Carvalho^{3,5}, F. Cendes^{4,5}, I. Lopes-Cendes^{1,5}, D. Comas². 1) Department of Medical Genetics, School of Medical Sciences, University of Campinas, Campinas, São Paulo, Brazil; 2) Departament de Ciències Experimentals i de la Salut, Institute of Evolutionary Biology (CSIC-UPF), Universitat Pompeu Fabra, Barcelona, Spain; 3) Department of Statistics, Institute of Mathematics, Statistics and Scientific Computing, University of Campinas, Campinas, São Paulo, Brazil; 4) Department of Neurology, School of Medical Sciences, University of Campinas, Campinas, São Paulo, Brazil; 5) Brazilian Institute of Neuroscience and Neurotechnology (BRAINN), Campinas, São Paulo, Brazil.

Background: Admixed populations can be used to identify causal variants related to disease due to the creation of local ancestry tracts across the genome. Although local ancestry information is available for admixed American populations, it is still poorly explored in Brazilian individuals. **Objective:** To describe the distribution and to identify deviations of local ancestry across the genome in a Brazilian sample. Furthermore, if significant deviations in local ancestry are found we aim to further explore the issue by searching for signals of recent natural selection and the possibility of association with disease. **Methods:** We evaluated 264 individuals obtained within the scope of the Brazilian Initiative on Precision Medicine (<http://biomed.org/>) and compared with the 1000 Genome Project database. Genotyping calling was performed by Genome-Wide Human SNP Array 6.0 platform. We processed the data and evaluated population structure by principal components using PLINK 1.9 software. Global and local ancestry was estimated by ADMIXTURE and RFMIX algorithm, respectively. We identified haplotype inversion events by the *invClust* package in R. We tested ancestry tracts for positive selection using five neutrality statistics combined into a Fisher combined score. **Results:** As expected, Brazilian sample presented a high proportion of South European ancestry, followed by West sub-Saharan African and Native American ancestries. However, local ancestry revealed a decreased European ancestry proportion followed by an excess of Native American ancestry on chromosome 8p23.1. We demonstrated that this deviation is due to haplotypes created by inversion events. Interestingly, Brazilian non-inverted haplotypes were found to be more similar to Native Americans than European haplotypes, different from what was observed for other admixed American populations. We also found signals of recent positive selection on chromosome 8p23.1. **Discussion:** We observed that one gene (*PPP1R3B*) located within the locus at 8p23.1 is related to diet and have already been associated with type 2 diabetes and obesity, which strengthens the positive selection hypothesis. Our results reinforce the need to study additional American admixed populations, avoiding extrapolated conclusions from one to another. Moreover, we show that individual local ancestry can be used to obtain relevant information about the risk of human diseases. Supported by: CEPID-BRAINN FAPESP, and BEPE-PD-FAPESP, São Paulo, Brazil.

2693F

Genetic differentiation and population structure of five endogamous groups of Punjab (North-West India): An analysis of *Alu* insertion and restriction site polymorphisms. G. Singh^{1,2,3}, H. Singh², AJS. Bhanwer². 1) Panjab University, Chandigarh, India; 2) Guru Nanak Dev University, Amritsar; 3) Wildlife Institute of India, Dehradun.

The state of Punjab on the North-West Frontier of India has served as a main corridor for all the major invasions into India, particularly of Indo-Europeans. It resulted in the mixing of foreign gene pool into the local population. The present study was designed to explore the genetic diversities and affinities among the five major endogamous groups using autosomal markers. A total of 1021 unrelated samples belonging to Baniyas, Brahmins, Jat Sikhs, Khatri and Scheduled Castes were genotyped for four Ins/Del (*ACE*, *APO*, *PLAT*, *D1*) and six restriction fragment length polymorphisms [*ESR* (*PvuII*), *LPL* (*PvuII*) *T2* (*MspI*) and three *DRD2* (*Taq1A*, *Taq1B*, *Taq1D*)]. All the loci were found to be polymorphic among the studied populations. The average heterozygosity among the studied populations is fairly substantial ranging from 0.3816 in Baniyas to 0.4311 in Scheduled Castes. The genetic differentiation among studied populations ranged from 0.0418-0.0022 for the *PLAT* and *Taq1A* loci, respectively, with an average value being 0.0136. Phylogenetic analysis revealed that Baniyas and Khatri are genetically closest to each other. The Jat Sikhs are genetically close to Brahmins and are distant from the Baniyas. The Jat Sikhs, Baniyas, Brahmins and Khatri are genetically very distant from the Scheduled Castes. Overall, a uniform allele frequency distribution pattern and low level of genetic differentiation was observed in the studied population groups indicating that genetic drift might have been small or negligible in shaping the genetic structure of North-West Indian Populations. .

2694W

Mutation rate estimation with three-way identity by descent. X. Tian, B.L. Browning, S.R. Browning. University of Washington, Seattle, WA.

We present a new method for mutation rate estimation based on segments of identity by descent (IBD) shared among multiple individuals. The IBD segments arise from common ancestry within the past few thousand years. Therefore, compared to inter-species comparison, our method avoids the potential confounding of mutation rates in humans with those in other closely related species. Our method uses the sharing of IBD segments among three individuals to estimate the mutation rate while controlling for genotype error, hence our method is more robust than parent-offspring trio analysis to genotype errors. Our method is applicable to sequence data with or without close relatives. We demonstrate the accuracy of our method in simulated data under various demographic models and observe that our approach is robust to substantial levels of genotype error. We apply the method to estimate the mutation rate in [how many] individuals from the Framingham Heart Study who were sequenced by the NHLBI TOPMed project (data obtained through dbGaP accession phs000974.v2.p2). We obtain an estimate of 1.18×10^{-8} mutations per basepair per generation with a 95% confidence interval of $[1.12 \times 10^{-8}, 1.25 \times 10^{-8}]$. Our estimate is consistent with previous estimates based on parent-offspring trio analysis, but has a tighter confidence interval because it is based on many more meioses.

2695T

Identifying the genetic and environmental determinants of gene expression variation in Africans. *D.E. Kelly¹, R. Ma¹, N.G. Crawford¹, Y. Ren¹, R.A. Rawlings-Goss¹, G.R. Grant¹, M. Yeager², S. Chanock³, A. Ranciaro¹, S. Thompson¹, J.B. Hirbo⁴, W. Beggs¹, T.B. Nyambo⁵, S.A. Omar⁶, D.O. Meskel⁷, G. Belay⁷, C.D. Brown¹, H. Li¹, S.A. Tishkoff¹.* 1) University of Pennsylvania, Philadelphia, PA; 2) National Institutes of Health, Division of Cancer, Bethesda, MD; 3) Frederick National Laboratory for Cancer Research, Frederick, MD; 4) Vanderbilt University, Nashville, TN; 5) Muhimbili University of Health and Allied Sciences, Salaam, Tanzania; 6) Kenya Medical Research Institute, Nairobi, Kenya; 7) Addis Ababa University, Addis Ababa, Ethiopia.

Gene regulation plays a predominant role in human evolution and complex traits, and high throughput methods are making the measurement of expression variation routine. However, studies to date have failed to capture the breadth of global genetic and environmental diversity by focusing primarily on Western individuals of European descent. Studies of Africans, who harbor the most genetic variation in the world and are exposed to a diversity of environmental variables and diets, are necessary to complete our understanding of how evolution has shaped human genetic and phenotypic diversity. To identify genetic and environmental contributors to gene expression variation in whole blood we have collected RNA sequencing data from 171 individuals representing 9 diverse East African populations. Differential expression analyses uncover genes correlated with ancestry and environmental variables, while cis-eQTL mapping identifies genetic variants contributing to gene expression variation in these samples. The majority of identified eQTLs replicate in Europeans, and those that fail to replicate are enriched for variants that are at moderate frequency in Africa and are low frequency or monomorphic in Europe. Integrating allele frequency differentiation and scans of natural selection, we identify candidate genes and pathways that may have undergone positive selection in these populations, including genes related to liver function in pastoralist populations. This study represents one of the most diverse study of gene expression in Africans to date, providing insights into the impact of adaptation on human gene expression and highlighting the need to extend genomics studies to non-European populations.

2696F

Genomes of 175 Mongolian individuals reveal population specific genetic architecture and genetic relationship with other human populations.

X. Guo^{1,6}, H. Bai^{2,4}, N. Narisu³, T. Lan^{1,6}, Q. Wu⁴, B. Burenbatur⁴, H. Zhou⁵, Y. Yin^{1,6,7}. 1) BGI, Shenzhen, Guangdong, China; 2) 1. School of Life Science, Inner Mongolia University for the Nationalities, Tongliao 028000, China; 3) 4. Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA; 4) Affiliated Hospital of Inner Mongolia University for the Nationalities, Tongliao 028000, China; 5) College of Life Science, Inner Mongolia Agricultural University, Hohhot 010018, China; 6) 3. Laboratory of Genomics and Molecular Biomedicine, Department of Biology, University of Copenhagen, DK-2100 Copenhagen, Denmark; 7) 17. School of Life Science and Biotechnology, Dalian University of Technology, Dalian 116024, China.

Approximately 10 million Mongolians currently reside in Mongolia, Northern China, Russia and other neighboring countries. Mongolians played important roles in early human dispersal, like the peopling of the New World, and in the shaping of recent Eurasians, such as the expansion of Mongol Empire. Mongolians are the typical nomadic people and have population specific traditional dietary, which leads to prevalence of related traits/diseases, such as BMI, hypertension, Type 2 Diabetes and so on. However, this interesting ethnic group is still under-studied. In this study, we aim to explore the genetic architecture of the population and their relationship to other ethnic groups in neighboring regions. To address this, blood samples from 175 Mongolians representing 6 ethnic tribes across northern China and Mongolia were sampled. Genomic DNA was extracted and the whole genome was sequenced to an average depth of 22x on Illumina HiSeq2000 platform. We characterized 15.2 million single nucleotide polymorphisms (SNPs), of which 3.9 million (25.8%) are novel when compared against the 1000 Genomes Project (1000G). Using the SNP data, we constructed a high-quality genetic variation reference panel for Mongolians. Our new reference panel fills an important gap between the East Asians and Admixed Americans provided by the 1000G and significantly improves our ability to impute missing genotypes. Although Mongolians most resemble East Asians genetically, however, from the clustering analysis, we observed the evidence of strong population stratification among tribes, and variance among Mongolians exceeds that of other East Asian populations. The structure analysis and patrilineal/matrilineal analysis also showed higher population diversity than other East Asian groups. Demography inference indicated diverse demographic histories among the Mongolian tribes, which approximately correlates with the current population structure. Pairwise estimation of separation time indicates Mongolians and other East Asians separated 7-13 kya but Horchin, whereas the tribes of Mongolians started to differentiate much later. Incorporating our results with the 1000G reveals shared alleles between Finns and Mongolians/Siberians, suggesting the extensive gene flow throughout northern Eurasian populations. Finally, through phylogenetic relationships, we highlight that North, East, and Southeast Asian populations are more aligned with each other than these groups are with South Asian and Oceanian populations. In summary, genomic data and the haplotype panel of Mongolians presented here will be a valuable resource for identifying population specific genetic causes of diseases in Central/North Asia, and exploring human evolutionary scenarios throughout Asia.

2697W

Genetic missense variants at the *EGLN1* locus associated with high-altitude adaptation in Tibetans are rare in Andeans. *E.S. Lawrence¹, E.C. Heinrich¹, L. Wu¹, F.C. Villafuerte², T.S. Simonson¹, C.A. Ramirez², A.M. Coel³, G.L. Cavalleri³.* 1) Medicine, UC San Diego, La Jolla, CA; 2) Facultad de Ciencias y Filosofía, Universidad Peruana Cayetano Heredia (UPCH), Lima, Peru; 3) Department of Molecular & Cellular Therapeutics, Royal College of Surgeons in Ireland, Dublin, Ireland.

High-altitude populations have been challenged by chronic hypoxia for many generations and demonstrate some of the most rapid evolution observed in humans. Genome-wide scans have brought to light numerous hypoxic-response candidate genes that are associated with distinct traits in these populations. One example is *EGLN1*, which shows signals of selection in Tibetan and Andean populations and has been linked to hemoglobin concentration in Tibetans living at high altitude. *EGLN1* encodes the hypoxia-inducible factor (HIF) prolyl hydroxylase 2 (PHD2) protein, which serves as an oxygen-sensitive regulator of HIF transcription factor activity. We determined the frequencies of these adaptive Tibetan alleles in Quechua Andeans residents at high altitude (4,350 m) in addition to individuals of Nepali ancestry resident at sea level. The rs186996510 minor C allele previously found at high genotypic frequency in Tibetans (N=26) (G/G: 0.15; G/C: 0.08; C/C: 0.77) (Lorenzo, et al. 2014), was absent in our Andean cohort (N=106) (G/G: 1.00) and rare among our Nepali cohort (N=17) (G/G: 0.82; G/C: 0.06; C/C: 0.12). The minor G allele of rs12097901 was also found at lower genotypic frequencies in Andeans (C/C: 0.76; C/G: 0.22; G/G: 0.02) and Nepalese (C/C: 0.65; C/G: 0.24; G/G: 0.12) compared to Tibetans (G/G: 0.81; G/C: 0.19; C/C: 0.00). These results suggest that adaptation involving *EGLN1* in Andeans involves different mechanisms than those described in Tibetans. The precise Andean adaptive variants remain to be determined.

2698T

Tankyrase: A promising target to investigate high altitude adaptation/maladaptation. *M. Miglani^{1,2}, M. Rain², M.A.Q. Pasha², A. Vibhuti¹.* 1) SRM UNIVERSITY, SONEPAT, India; 2) CSIR-IGIB, NEW DELHI.

Telomeres are prone to damage by the ROS produced due to oxidative stresses encountered at high altitude (HA). Tankyrase (TRF-1 interacting ankyrin-related ADP-ribose polymerase) acts as a positive regulator of telomere length as it performs post translational modification during DNA damage. It thus could be an important marker with respect to HA adaptation or maladaptation. The present study investigated the plausible roles of genetic polymorphism of tankyrase gene and its level as a biochemical in three well-defined groups, namely High altitude pulmonary edema-patients (HAPE-p), HAPE-free sojourners (HAPE-f) and healthy highlanders (HLs) comprising 200 samples, each. Genotyping was performed using PCR-RFLP and the protein level was estimated using ELISA and correlations between genotype/allele with protein levels were analysed. Three tankyrase SNPs, rs7015700, rs12541709& rs11991621 were genotyped. The heterozygote rs7015700GA was prevalent in HLs than HAPE-f, hence it may contribute towards adaptation. Tankyrase rs11991621GA also exhibited increased prevalence in HLs over HAPE-f. For both SNPs, heterozygotes were found to be strongly prevalent in HLs over HAPE-f & HAPE-p. HLs have the highest telomere length followed by HAPE-f and HAPE-p ($p < 0.05$, HLs v/s HAPE-r; HAPE-p v/s HAPE-r). In a correlation analysis between the heterozygotes and the telomere length, the greater length was observed in HLs than the patients. This phenomenon also suggests that the telomere effect could be multifactorial. HAPE-p had significantly decreased and HLs had marginally reduced concentration of tankyrase as compared to HAPE-f ($p \leq 0.05$). For genotypic & allelic correlations with tankyrase levels, rs7015700GA genotype contributed towards higher levels ($p < 0.05$) in HAPE-f and rs11991621G allele contributed towards higher levels ($p < 0.05$) in HAPE-p. These results suggest that perhaps telomere length is more regulated by telomerase enzyme than the tankyrase. Our findings suggest that Tankyrase levels are high in HAPE-f which is attributed to their role in maintenance of telomere length hence this gene becomes vital to study adaptation/maladaptation at HA. .

2699F

Low-frequency variant functional architectures reveal strength of negative selection across coding and non-coding annotations. S. Gazal^{1,2}, P.-R. Loh^{2,3}, H.K. Finucane², A. Ganna^{2,4,5}, A. Schoech^{1,2,6}, S. Sunyaev^{2,3,7}, A.L. Price^{1,2,6}. 1) Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA; 2) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; 3) Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA; 4) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; 5) Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA; 6) Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA; 7) Department of Biomedical Informatics, Harvard Medical School, Boston MA, USA.

Common variant heritability is known to be concentrated in variants within cell-type-specific non-coding functional annotations, with a limited role for common coding variants. However, little is known about the functional architecture of low-frequency variants, which are disproportionately impacted by negative selection. Here, we partitioned the heritability of both low-frequency ($0.5\% \leq \text{MAF} < 5\%$) and common ($\text{MAF} \geq 5\%$) variants in 40 UK Biobank traits (average $N = 363\text{K}$) across a broad set of coding and non-coding functional annotations, employing an extension of stratified LD score regression with the baseline-LD model (Finucane et al. 2015 *Nat Genet*, Gazal et al. 2017 *Nat Genet*) to low-frequency variants that produces robust results in simulations. We determined that non-synonymous coding variants explain $17 \pm 1\%$ of low-frequency variant heritability (h_{ir}^2) versus only $2.1 \pm 0.2\%$ of common variant heritability (h_{c}^2), and that regions conserved in primates explain nearly half of h_{ir}^2 ($43 \pm 2\%$). Other annotations previously linked to negative selection, including non-synonymous variants with high PolyPhen-2 scores, non-synonymous variants in genes under strong selection, and low-LD variants, were also significantly more enriched for h_{ir}^2 as compared to h_{c}^2 . Cell-type-specific non-coding annotations that were significantly enriched for h_{c}^2 of corresponding traits tended to be similarly enriched for h_{ir}^2 for most traits, but more enriched for brain-related annotations and traits. For example, H3K4me3 marks in brain dorsolateral prefrontal cortex explain $57 \pm 12\%$ of h_{ir}^2 vs. $12 \pm 2\%$ of h_{c}^2 for neuroticism, implicating the action of negative selection on low-frequency variants affecting gene regulation in the brain. By performing forward simulations at parameter settings consistent with our UK Biobank results, we confirmed that the ratio of low-frequency variant enrichment vs. common variant enrichment primarily depends on the mean selection coefficient of causal variants in the annotation, enabling us to quantify the strength of negative selection across annotations. Furthermore, we used this ratio to predict the effect size variance of causal rare variants ($\text{MAF} < 0.5\%$) in each annotation, informing their prioritization in whole-genome sequencing studies. Our results provide a deeper understanding of low-frequency variant functional architectures and guidelines for the design of association studies targeting functional classes of low-frequency and rare variants.

2700W

Heritability of language and associations with population demography. N. Telis¹, H. Arbel¹, E.P. Sorokin¹, M. Zhang¹, J.M. Granka¹, E.L. Hong¹, K.G. Chahine², C.A. Ball¹. 1) Ancestry.com DNA LLC, San Francisco, CA; 2) Ancestry.com DNA LLC, Lehi, UT.

Language and population structure are intricately connected by human culture. Past studies have found that while language can roughly mirror population genetic patterns, the two diverge, likely because of the relative fluidity and recency of linguistic change. By using a large cohort with rich pedigree information, we are able to examine the relationship between genetics and exceptionally recent linguistic patterns by comparing reported language with reported grandparental language. We characterize the heritabilities of different languages by comparing persistence of reported grandparent language with reported language in their grandchildren. To further explore the connection between recent genetic history and linguistics, we also compare the ethnicity assignments of grandchildren who did and did not maintain their grandparents' languages. We identify connections between recent linguistic retention and subtle genetic differences in subpopulations of language speakers, and use language to further understand subtle population structure arising even within populations sharing language as recently as 2 generations ago.

2701T

Demographic inference with sequential Monte Carlo methods reveals large European back-migration in the Late Pleistocene. C.B. Cole¹, D. Henderson¹, S.J. Zhu², G. Lunter¹. 1) Wellcome Centre for Human Genetics, University of Oxford, Oxford, Oxfordshire, United Kingdom; 2) Big Data Institute, University of Oxford, Oxford, Oxfordshire, United Kingdom.

Background: Mutations within human genomes can be informative about our species' demographic history, and are shaped by a complex mix of migration and admixture events as well as population size changes. Existing methods to infer demography from mutations differ in the approximations made in order to make inference tractable, resulting in a range of compromises between resolution and range the type of demographic events that can be inferred. **Methods:** We explore the use of sequential Monte Carlo techniques applied to the Sequentially Markovian Coalescent (SMC²) to infer demographics from whole genome sequencing data. The approach is asymptotically exact, and since Ancestral Recombination Graphs (ARGs) are explicitly simulated, and has the ability to infer directional migration rates between populations. Complex and increasingly realistic models are straightforward to investigate as inference is fundamentally based on simulation. **Results:** Here we analyze the genomes of ancestrally African and non-African individuals using a two-island model with migration. We detect a large migration event from ancestral European individuals back to Africa approximately 50 to 70 thousand years ago. Simulation experiments support this interpretation, and reject scenarios lacking a strong back-migration pulse. We also find that modelling European and African populations jointly with back-migration explains a previously unexplained deep ancestral high diversity signal in single-population analyses of African samples, and shows better alignment of the two population's inferred population sizes through time, with an inferred split time that is closer to the accepted dating of the out-of-Africa event. **Conclusion:** As the fossil record and genetics continue to illuminate a complex out-of-Africa event, additionally powerful and exact methods are needed to describe the demographic histories of particular groups and individuals. Here we introduce a new method for demographic inference from whole-genome data, and show that migrations back to Africa may have played a more substantial role in shaping the genetic diversity of some African populations than previously realised.

2702F

Host genetic control of the microbiome in wild primates. R. Blekhman¹, L. Grieneisen¹, M. Dasari², T. Gould¹, J.R. Bjork², J.-C. Grenier³, V. Yotova³, D. Jansen², N. Gotte⁴, J.A. Gilbert⁴, L. Barreiro³, J. Tung⁵, E. Archie². 1) University of Minnesota, Minneapolis, MN; 2) University of Notre Dame, Notre Dame, IN; 3) Université de Montréal, Montreal, Québec, Canada; 4) University of Chicago, Chicago, IL; 5) Duke University, Durham, NC.

Variation in gut microbiome composition—defined as the identity and abundance of different bacterial taxa—is linked to key health outcomes in humans and other vertebrates. Variation in the gut microbiome is affected by many environmental factors, but host genetics also has a significant impact on the microbiome. Understanding the relative roles of genetic and environmental factors is a central goal in human disease research. However, it is often difficult to disentangle genetic from environmental effects because relatives often use the same resources and have similar diets. Here, we identified components of the gut microbiome that are associated with host genetics—i.e. heritable microbes—controlling for changes in the host's physical and social environments over time. Specifically, we used an unprecedented dataset of >10,000 microbiome profiles collected over 14 years from a well-studied population of wild savannah baboons (*Papio cynocephalus*) with available host genetic relatedness data. We found that gut microbial composition changes over time and across environments. Specifically, the overall microbiome community composition was affected by baboon social group membership, collection year, and hydrologic season. Controlling for these fluctuations in the physical and social environment, we identified several heritable common bacterial taxa, such that more related individuals had more similar bacterial abundances. Heritable taxa included Christensenellaceae and Rikenellaceae, bacterial families that are also highly heritable in humans, suggesting that common host genetic mechanisms may control microbiome composition across primates. Finally, the relative abundance of heritable taxa changes over time within individuals, emphasizing the importance of longitudinal studies to account for temporal and environmental effects in shaping gut microbial composition. To summarize, we present the first analysis of host genetic control of the gut microbiome in non-human primates, indicating that microbiome heritability is affected by fluctuating environment, and suggesting that heritable microbial taxa may be shared across primate species.

2703W

Patterns of positive selection inferred across the Pacific. S. Kaewert¹, G. Wojcik², E. Sorokin², J. Homburger², C. Quinto-Cortés⁴, A. Ioannidis⁵, J. Portillo⁴, J. Rodriguez⁴, K. Auckland⁶, A. Severson³, J. Blanc⁷, G. Belbin⁸, S. Oppenheimer⁹, T. Parks⁸, M. Phipps⁸, K. Robson⁸, E. Kenny⁸, A. Hill⁸, A. Mentzer⁶, C. Bustamante², A. Moreno-Estrada⁴, C. Gignoux^{1,10}. 1) Colorado Center for Personalized Medicine, University of Colorado, Aurora, Colorado, USA; 2) Department of Biomedical Data Science, Stanford University, Stanford, CA; 3) Department of Genetics, Stanford University, Stanford, CA; 4) National Laboratory of Genomics for Biodiversity, Irapuato, Guanajuato, Mexico; 5) Computational and Mathematical Engineering, Stanford University, Palo Alto, CA; 6) University of Oxford, United Kingdom; 7) University of California, Davis, Davis, CA; 8) Icahn School of Medicine at Mount Sinai, New York, NY; 9) Monash University, Malaysia; 10) Department of Biostatistics and Informatics, University of Colorado, Aurora, Colorado, USA.

The unique prehistory of the settlement of the Pacific was a complex demographic event involving numerous migrations and exposure to a wide variety of environmental pressures on new islands. These pressures include the unique disease dynamics present in isolated, high-density populations with intermittent contact. However, the region of Oceania has been understudied in human genetics. Leveraging the data from the new Oceanian Genome Variation Project (OGVP), involving 980 individuals from 88 populations measured on the MEGA platform at over 1.7M sites along the genome, we used a combination of haplotype-based (iHS, nSI, xp-EHH) and differentiation (Fst, SPA) statistics to infer patterns of positive selection unique to regions across the Pacific. The uniquely detailed populations of OGVP allow for hierarchical evidence of selection acting on populations, specific regions, or Oceania as a whole compared to mainland Asia. Enrichment analysis of top selection hits show significant enrichment in genes associated with antigen processing and presentation (Melanesia, Micronesia, Solomon Islands, Southeast Asia), cytokine production (Melanesia and Southeast Asia), cellular defense response (Mainland East Asia, Melanesia, Solomon Islands, Micronesia, Republic of Palau and part of Papua New Guinea), viral process (Coastal Papua New Guinea) and response to interferon-gamma (French Polynesia and Tuvalu)(all FDR < 0.08). These results suggest that disease has played a critical role in the settlement of Oceania. Finally, as recently established, we extend linear mixed effects models to test for significant effects of selection associated with latitude and longitude in a GWAS framework, finding associations with immune genes IL6R (FDR=0.001), IL20RA (FDR=0.004), and C1QA, B, C (FDR=0.003) as well as a general enrichment for genes associated with cellular defense response (FDR=0.03). These results, combined with the relatively low effective population size in the region, suggest that infectious disease has provided a strong selective pressure throughout Oceania.

2704T

A pharmacogenomic variation database of Chinese population based on whole genome sequencing. J. Guo, R. Guo, N. Qu, M. Jian, X. Liu. BGI-Shenzhen, Shenzhen, Guangdong, China.

Background: Genetic variations in drug-related genes are responsible for up to 40 percent of individual drug response differences. Previous studies have underlined the importance of rare single nucleotide variants in pharmacogenes which approximately contribute 30-40 percent of functional variability. However, a comprehensive view of pharmacogenetic profiling in Chinese population is lacking. **Methods:** A cohort of 968 healthy independent Chinese individuals covering 30 administrative divisions were recruited. Whole genome sequencing was performed on BGISEQ 500 platform with 100-bp paired-end reads. **Results:** Overall, we have identified 175,283 variants in 229 pharmacogenes that encode 138 phase I and phase II drug metabolizing enzymes, 51 transporters, 40 drug targets and regulators. The majority were single nucleotide variants, and the ratio of transition (Ti, n= 104005) and transversion (Tv, n= 48344) was close to two, which was consistent with the ratio across the whole genome. The total number of variants were normalized by the length of the corresponding gene, yielding 16.7 variants per kbp on average, among them, *CYP2D6* and *SULT1A1* genes had the most variants, which were 43.6 and 43.3 variants per kbp respectively while *UGT1A7* and *UGT1A9* seemed to be most conserved with 1.8 variants per kbp only. In the current study, there were 81383 novel variants discovered compared to dbSNP Build 150 and 119349 variants were firstly reported in east Asian population in comparison with phase III 1000 genome project. Bioinformatics analysis revealed that 53.5 percent (n=93750) of variants were singletons, 10.3 percent were doubletons, and 14.6 percent were rare with minor allele frequency no more than 0.01, and only 21.6 percent were common variants with minor allele frequency more than 0.01. Most of the variants (94.4 percent) were located in the intron region as expected. The composition analysis of exonic variants showed that only 16 percent were synonymous, 30 percent were missense mutations and 50 percent were in UTR which might have functional effects. **Conclusion:** Based on whole genome sequencing technique, our study developed so far, the most comprehensive pharmacogenomic variation dataset in Chinese population. These data could be useful resource for accurate analysis of pharmacogenetic genotyping incorporating not only common and known variants but also rare functional variants to facilitate personalized medicine. .

2705F

The evolution of TLR7 and TLR8 in yellow fever virus endemic areas. *N. Torosin, L.A. Knapp.* Anthropology Dept, University of Utah, Salt Lake City, UT.

In Misiones, Argentina, howler monkey populations (*Alouatta guariba clamitans*) and (*Alouatta caraya*) were devastated first in 1965, and then in 2008-2009, by Yellow Fever Virus (YFV). Unlike humans and other South American monkeys, howler monkeys are highly susceptible to YFV and die within a week after infection. To better comprehend the relationship between immune function and genetic variation in *Alouatta*, we identified and sequenced two innate immune genes that identify YFV, known as Toll-like receptors (TLR7 and TLR8). Study subjects (N = 41) were from one of three categories: 1) alive prior to the outbreak, 2) died from the virus, or 3) survived the outbreak along with their progeny. No reference genome exists for *Alouatta*, therefore we first performed *de novo* sequencing for one individual of each species and then used Ampliseq to execute targeted sequencing of the genes of interest for the remaining samples. Using these data, along with genetic data previously collected from *Alouatta palliata mexicana* and published primate immune gene sequences, optimal maximum likelihood models of sequence evolution were identified, and phylogenetic trees were created. We compared the allele frequencies of innate immune genes TLR7 and TLR8 from pre-YFV outbreak, post-YFV outbreak, and YFV affected howler individuals using Fishers exact test. Frequencies of twenty-one variants in TLR7 and nine variants in TLR8 were statistically significant between the three groups, suggesting that some TLR variants may provide individuals with better resistance to YFV than others. Our results are valuable for identifying underlying genetic variation that affects susceptibility not only to YFV, but potentially to other single stranded RNA viruses.

2706W

The genetic structure and demographic history of Mexicans of African descent. *M.C. Ávila-Arcos¹, S. Ramachandran¹, A. Moreno-Estrada², M.W. Feldman^{3,4}, C.D. Bustamante³.* 1) International Laboratory for Human Genome Research, UNAM. Querétaro, Mexico.; 2) Laboratorio Nacional de Genómica para la Biodiversidad, CINVESTAV-IPN, Irapuato, Guanajuato, Mexico; 3) Center for Evolutionary, Computational and Human Genomics, Stanford University, Stanford, CA, USA; 4) Department of Biology, Stanford University, Stanford, CA, USA.

Despite having received 200,000 Africans during the slave trade, no study in Mexico has focused on the characterization of the African genetic ancestry of its Afro-descendant population. In this study we worked together with Afro-Mexican communities to characterize their genetic ancestry using dense genome-wide genotyping. The dataset consists of 380 self-identified Afro-descendants, indigenous and mestizo population from three Mexican states. To complement the genome-wide genotype data, we collected genealogical, self-identification and phenotypes such as skin pigmentation, height, weight, hip to waist ratio and hemoglobin. Genotype data revealed that up to 50% of the genetic ancestry of some individuals is of African origin, with the remaining being mainly of indigenous origin. By exploring local ancestry and admixture patterns in this population, as well as correlations between the genetic ancestry, self-identification and phenotypes; we have characterized aspects of their demographic history and some trends of public health relevance. Lastly, since Afro-Mexicans currently suffer from poverty, discrimination, lack of recognition as a vulnerable minority, and limited access to health services, this study contributes to their appreciation as part of Mexico's mosaic of diversity and will hopefully set the stage for health interventions.

2707T

Reconstructing the history of the Native American populations from Brazilian coast. *T. Hünemeier¹, M. Castro-Araújo¹, A. Mas^{2,4}, K. Nunes¹, M.C. Bortolini², F.M. Salzano², A. Pereira³, D. Comas⁴.* 1) Genetics and Evolutionary Biology, University of São Paulo, Brazil; 2) Department of Genetics, Federal University of Rio Grande do Sul, Porto Alegre, Brazil; 3) Laboratory of Genetics and Molecular Cardiology, Heart Institute (InCor), University of São Paulo Medical School Hospital, São Paulo, Brazil; 4) Institut de Biologia Evolutiva, Universitat Pompeu Fabra, Barcelona, Spain.

When Portuguese first arrived at Northeast Brazilian coast around 500 years ago, the first native population with whom they made contact were the Tupiniquins. After the Conquest, a reduction of 95% in the autochthonous population is estimated to have happened due to the introduction of European infectious diseases, slave trade and warfare. In 1876 the total population of Tupiniquins decayed to just 55 individuals, and they were declared extinct soon after. However, in the middle of the XX century some admixed individuals were culturally identified as belonging to the Tupiniquin group. Although these group is considered culturally related to the extinct Tupiniquin population, they do not speak any native language and there is no clear evidence whether this community is genetically related to the extinct tribe of the Conquest time or another autochthonous population resulting from the admixture of different people belonging to the Tupí and Jean groups (the two main Brazilian autochthonous linguistic stocks), such as the Guaraní people that has lived in the same region during the last century. Here we investigated the genomic origins of this self-declared Tupiniquin group by studying their extant population (n=50) as well as other native populations geographically and linguistic related to them (n=133). We also used the Tupiniquin to try elucidated the dispersion these group along Brazilian coast around 2000 years before present. Our results showed that the individuals self-declared as Tupiniquin presented a Native American ancestry not closely related to any other Brazilian Native extant population, and so could be the only living representatives of the coastal Tupí branch, representing genetically the populations that were extinct after the Conquest.

2708F

How genetic disease risks can be misestimated across global populations. *M. Kim, J. Lachance.* Georgia Institute of Technology, North Ave NW, Atlanta, GA 30332.

Accurate assessment of health disparities requires unbiased knowledge of genetic risks in different populations. Unfortunately, most genome-wide association studies use genotyping arrays and European samples. Here, we integrate whole genome sequence data, GWAS results, and computer simulations to examine how genetic disease risks can be misestimated. We find that genetic disease risks are substantially overestimated for individuals with African ancestry. These patterns hold for multiple disease classes (e.g., cancer, gastrointestinal, morphological, and neurological diseases). A contributing factor to this bias is that existing genotyping arrays are enriched for SNPs that have higher frequencies of ancestral alleles in Africa. By simulating GWAS with different study populations, we find that non-African cohorts yield disease associations that have biased allele frequencies and that African cohorts yield disease associations that are relatively free of bias. We also find empirical evidence that genotyping arrays and SNP ascertainment bias contribute to continental differences in risk allele frequencies. Because of these causes, polygenic risk scores can be grossly misestimated for individuals of African descent. Importantly, continental differences in risk allele frequencies are only moderately reduced if GWAS use whole genome sequences and thousands of cases and controls. Finally, our study demonstrates continental differences in allele frequencies also depend upon SNP age. Comparisons between uncorrected and corrected genetic risk scores reveal the benefits of taking into account whether risk alleles are ancestral or derived. This demonstrates that evolutionary history can influence predictions of complex traits, including polygenic disease risks. In sum, our results imply that caution must be taken when extrapolating GWAS results from one population to predict disease risks in another population.

2709W

The impact of genetic admixture and natural selection on driving population differences in East Asia. S. Xu. CAS-MPG Partner Institute for Computational Biology, Shanghai, Shanghai, China.

Genetic admixture in human, the result of inter-marriage among people from different well-differentiated populations, has been a common phenomenon throughout the history of modern humans. Understanding of population admixture dynamics is important not only to mapping human complex traits/diseases but also to other applications, such as elucidating population history and detecting natural selection signatures. By analyzing the genomes collected from several East Asian populations together with many present-day and ancient human genomic data, we propose that gene flow (recent population admixture and ancient genetic introgression), as an evolutionary driving force and also an under-investigated one, played an important role in shaping human genetic diversity and population adaptive divergence in East Asia.

2710T

Fine-scale population structure of European Land's end: Genetics of North Western France and the Loire River. C. Dina¹, J. Gienza¹, I. Alves¹, M. Karakachoff¹, A. Saint-Pierre², P. Lindebaum¹, E. Svensson³, E. Charpentier¹, K. Rouault², E. Baron¹, S. Bonnaud¹, T. Ludwig², S. Lecointe¹, C. Berthelot¹, E. Gallic⁴, A. Charpentier¹, C. Férec², H. Le Marec¹, R. Olaso⁵, S. Chatel¹, V. Meyer⁶, A. Boland⁶, M. Jakobsson³, P. Darlu⁵, J.-J. Schott¹, J.-F. Deleuze⁶, E. Génin², R. Redon¹. 1) l'institut du thorax, INSERM, CNRS, UNIV Nantes, CHU Nantes, Nantes, France; 2) UMR 1078 Génétique, Génomique fonctionnelle et Biotechnologies, Inserm, Université de Brest, EFS, CHU Brest, Brest, France; 3) Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden; 4) Université de Rennes; 5) CNRS/ MNHN/ Université Paris Diderot, UMR 7206 Eco-anthropologie et ethnobiologie; 6) Centre National de Recherche en Génomique Humaine - CNRGH DRF - CEA, Evry France.

Characterising the genetic structure of human populations provides insight into demographical history and informs research on disease association studies. In this study, we examine the fine-scale genetic structure of French Brittany, Anjou, Poitou and Maine. This Peninsula spans from continental area to a final edge called Finistère ("end of Lands") with Loire River as the only physical barrier, covering 60 000km² and extending 470 km from west to east. Combining array genotypes, whole genome sequence, surname and genealogy distribution data and ancient DNA we characterize the fine-scale population structure of this area and identify relation with history. **Methods Results** We genotyped 3500 individuals with grandparents born in northwestern France on Axiom™ PMRA. A subset of 400 individuals, together with 4 local medieval individuals, were sequenced whole genome using Illumina Ten X. Principal Components are highly correlated with geographical coordinates ($p < 2e-16$). Genetic distances are highly correlated with surname-based and genealogy distances (Mantel test, $r = 0.8$ and $r=0.3$). Analysis based on haplotype structure, using "chromosome painting", confirmed this fine-scale structure. The initial division occurs between North and South of Loire, the second one reflects historical border of Brittany. At the finest level, we obtained 79 clusters with more than ten individuals (154 in total) revealing more detailed structure. We observed a pattern of gradual increase of average runs of homozygosity towards the end of the Brittany, with local historically interesting hotspots. Consistently, higher levels of IBD sharing appeared in Brittany, increasing towards the cape. Using IBD sharing, we were able to identify the recent exponential growth in effective population size already seen in European populations but subgroups may have followed different trajectories in the last 25 generations. The WGS SFS confirmed the previous observations and showed a decrease in singletons in Brittany, with a westwards gradient, reflecting stratification or smaller effective population size. **Conclusion** We here report existence of a fine-scale structure across western France, with evidence of distinct demographic histories between subpopulations. To our knowledge this is the first observations of a coincidence of geographical and historical limits with genetic barriers at such a small geographical level in Europe. We are presenting results combining a large variety of information .

2711F

Ancestry specific methods for reconstructing human settlement across the Pacific. A. Ioannidis¹, J. Blanco Portillo², C. Quinto-Cortés², J. Homburger³, G. Wojcik⁴, K. Auckland⁵, K. Sandoval⁶, A. Severson³, C. Barberena Jonas⁴, J.E. Rodriguez², J. Blanc⁷, G. Belbin⁸, S. Oppenheimer⁹, T. Parks⁵, E. Hagelberg⁹, M. Moraga¹⁰, J.F. Miquel-Poblete¹¹, M. Ávila-Arcos¹², A. Sockell³, C. Eng¹³, E. Burchard¹³, R. Verdugo¹⁰, M. Phipps¹⁴, K. Robson⁵, E. Kenny⁶, A. Hill⁵, A. Mentzer⁵, C. Gignoux¹⁵, C. Bustamante⁴, A. Moreno-Estrada² *On behalf of the Oceanian Genome Variation Consortium.* 1) Computational and Mathematical Engineering, Stanford University, Stanford, CA, USA; 2) National Laboratory of Genomics for Biodiversity, Irapuato, Guanajuato, Mexico; 3) Department of Genetics, Stanford University, Stanford, CA, USA; 4) Department of Biomedical Data Science, Stanford University, Stanford, CA, USA; 5) University of Oxford, United Kingdom; 6) Undergraduate Program on Genomic Sciences, UNAM, Cuernavaca, Morelos, Mexico; 7) University of California, Davis, Davis, CA, USA; 8) Icahn School of Medicine at Mount Sinai, New York, NY, USA; 9) Department of Biology, University of Oxford, Blindern, Norway; 10) Faculty of Medicine, University of Chile, Santiago de Chile, Chile; 11) Pontificia Universidad Católica de Chile, School of Medicine, Santiago de Chile, Chile; 12) International Laboratory for Human Genome Research, Universidad Nacional Autónoma de México, Queretaro, Mexico; 13) Department of Bio-engineering, University of California San Francisco, San Francisco, USA; 14) Monash University, Malaysia; 15) Colorado Center for Personalized Medicine and Department of Biostatistics and Informatics, University of Colorado, Aurora, Colorado, USA.

The settlement of Oceania represents a uniquely structured, final chapter in the historical founding events of the human species. Resolving the detailed sequences of islands settled and dates of settlement requires careful attention to both haplotype and genetic drift based methodologies. These must be developed within a framework that accounts for the confounding effects of subsequent waves of complex admixture (Melanesian and European). We will present methodologies for analyzing these island populations that we have developed across two different large studies (the Oceanian Genome Variation Project: 980 individuals from 88 populations genotyped at 1.7M sites and a eastern-Pacific dataset comprising 500 individuals from 20 populations genotyped at 700k sites). We introduce an algorithm for graph reconstruction based on combining directionality measures with dissimilarity measures, a method for dating settlement based on haplotypes, and a PCA-like matrix completion-based method for visualizing ancestry-specific patterns of population differentiation. Based on these new methods we will present new findings about the origin and timing of specific migrations, including strong evidence for a complex pre-European admixture event between Polynesians and Native Americans.

2712W

An initiative for genetic data collection from underrepresented countries and populations. K.F. McManus, M. Moreno, J. Kim, K. Bryc, J. Mountain, . 23andMe Research Team. 23andMe, San Francisco, CA.

It is well-known that currently available human genetic data do not adequately represent the breadth and depth of the world's diversity. The lack of genetic research on people of diverse ancestries greatly limits a comprehensive understanding of genetic variation in *Homo sapiens*, as well as hinders understanding of genetic disease risk in populations underrepresented in genetic research. 23andMe's Global Genetics Project, launched in 2018, seeks to genotype individuals with four grandparents from large populations throughout the world that are underrepresented in genetic research. This initiative follows our 2016-17 African Genetics Project, which enrolled over 1,100 people with four grandparents from certain African countries, including hundreds of samples from each of Ethiopia, Somalia and Sudan. Over the next few years, we aim to enroll 6,000 participants representing 62 countries in regions including central, western and southeastern Asia; northern, western and southern Africa; Oceania; and Central and South America. Consenting participants, who are recruited from the US, are asked to provide a saliva sample and complete a survey about their family's birthplaces, cultural affiliations and languages. These data will be used for a variety of ancestry and health research projects. For ancestry research, they will be used to improve understanding of population structure, diversity and migrations throughout the world. They will be used to further health research in underrepresented populations by exploring models of disease risk in these populations.

2713T

Identification and characterization of *AluYRa1* exonization event in cynomolgus macaque genome. H. Cho^{1,2}, Y. Kim^{1,2}, S. Park¹, J. Lee¹, S. Choe^{1,2}, J. Huh^{1,2}. 1) National Primate Research Center, Korea Research Institute of Bioscience and Biotechnology(KRIBB), 30 Yeongudanji-ro, Ochang-eup, Cheongwon-gu, Cheongju-si, Chungcheongbuk-do, Korea; 2) Department of Functional Genomics, KRIBB School of Bioscience, Korea University of Science & Technology(UST), 217 Gajeong-ro, Yuseong-gu, Daejeon, Korea.

Transposable elements (TEs) moving around the genome are considered as a powerful source in gene transcript diversification by being involved in RNA splicing process. *AluYRa1*, one of TEs, has previously been identified in rhesus macaque genome. In this study, we found that some of *AluYRa1* are forward-oriented and located at the end of several genes by in-silico analysis of crab-eating monkey genome. To investigate this phenomenon, further detailed computational analysis of recently registered mRNA sequences and experimental validation were performed. Computational analysis revealed that 9 genes, *GTPBP4*, *PEX26*, *IRF9*, *BLOC1S6*, *TK2*, *CMBL*, *SLC16A14*, *PDK4* and *UBE2B* are expected to contain forward-oriented *AluYRa1* at their ends. Therefore, we carried out PCR for the target site of these genes to analyze both species specificity and evolutionary insertion time of *AluYRa1* and RT-PCR to check expression pattern of them. Our results show that *AluYRa1* is distributed in rhesus monkey and crab-eating monkey, and forward-oriented *AluYRa1* at 3'UTR tends to provide proper environment for alternative polyadenylation. Additionally, this element shows polymorphic insertion pattern in *TK2* and *PDK4*. In conclusion, *AluYRa1* is one of Old World monkey specific TEs contributing to gene transcript diversity with forward oriented insertion pattern specifically by alternative polyadenylation.

2714F

Tracing changes in gene regulation during human evolution by imputing ancient gene expression. L.L. Colbran¹, M.R. Johnson², E.R. Gamazon^{1,3}, P. Evans¹, N.J. Cox¹, J.A. Capra^{1,4}. 1) Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN; 2) Vanderbilt School for Science and Math, Vanderbilt University, Nashville, TN; 3) Clare Hall, University of Cambridge, Cambridge, United Kingdom; 4) Department of Biological Sciences, Vanderbilt University, Nashville, TN.

Despite multiple interbreeding events, the majority of the Neanderthal genome does not persist in modern human genomes, and the demographic and selective pressures that determined the landscape of remaining Neanderthal sequences are unresolved. Given the importance of changes in gene regulation to the evolution of differences between species, we hypothesized that gene regulatory divergence between modern humans and Neanderthals was a driver of loss of introgressed Neanderthal DNA. We tested this by adapting the PrediXcan approach to impute the effects of absent Neanderthal DNA on modern human gene regulation and identified 1,345 genes with significant evidence for differential regulation by non-introgressed Neanderthal DNA. Integrating data from Vanderbilt University's BioVU biobank, we demonstrated that human variation in the regulation of these genes is associated with neurological and immune phenotypes, with significant enrichment for mental disorders. This suggests that effects on regulation, especially in the brain, were likely a barrier to introgression of Neanderthal DNA. We also compared the imputed gene regulation of two Neanderthals and a Denisovan, revealing variation in immune system regulation between archaic populations that likely reflects of differing environmental pressures between them. Since diverging from their last common ancestor with Neanderthals, human gene regulation has continued to evolve in response to environmental and cultural shifts. To identify potentially adaptive gene regulation changes, we expanded our analyses to include the rapidly-increasing number of samples from ancient humans. In our pilot analysis, we imputed expression and analyzed differences in 12 ancient Hungarians spanning from 5,710 calBC to 1,180 calBC that encompass substantial shifts in diet and lifestyle. We identified 196 genes that were differentially regulated between the oldest and youngest samples, including four genes associated with gluconeogenesis and seven involved in mitochondrion organization. These differentially regulated genes suggest that genetic changes influenced metabolism across the timespan and highlight specific genes and pathways. Overall, our results establish substantial differences in gene regulatory architecture between humans and archaic hominins and demonstrate the potential of this method for studying patterns of gene regulation in ancient humans.

2715W

Genomics of ancient Brazilian populations. *D.I. Cruz Dávalos^{1,2}, M. Allentoft³, S. Reis⁴, M. Bastos⁴, J. Stenderup³, C.S. Reyna Blanco⁵, S. Neuenschwander^{1,2}, V. Moreno-Mayar⁶, C. Rodrigues-Carvalho⁴, E. Willerslev³, M.C. Ávila Arcos⁶, A.S. Malaspinas^{1,2}.* 1) Department of Computational Biology, University of Lausanne, Lausanne, Switzerland; 2) Swiss Institute of Bioinformatics, Lausanne, Switzerland; 3) Centre for GeoGenetics, University of Copenhagen, Copenhagen, Denmark; 4) Setor de Antropologia Biológica, Museu Nacional, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil; 5) Department of Biology, University of Fribourg, Fribourg, Switzerland; 6) International Laboratory for Human Genome Research, National Autonomous University of Mexico, Juriquilla, Mexico.

When the Portuguese settlers arrived in Brazil, they named "Botocudos" the Central-Eastern inhabitants who wore wooden lip and ear plates. The Botocudos actually comprised several cultural groups speaking distinct languages. Most of the Botocudo groups became extinct by the end of the 19th century, when the Portuguese Crown declared war to all indigenous tribes that refused to accept the regime. The National Museum of Rio de Janeiro owns a collection of 35 Botocudo skulls. While analyzing the mitochondrial DNA of 14 of these samples, Gonçalves et al. identified two individuals with Polynesian haplogroups. The ancestry of those individuals was further explored at the nuclear level by Malaspinas et al., who determined that both samples had a Polynesian ancestry without signs of admixture from Native American or African populations. Several scenarios were put forward to explain those results including that these individuals arrived in the Americas as part of the Polynesian expansion, or as crew or passenger members in European ships. Here, we report preliminary genomic data from 21 individuals from the Botocudo collection. We perform population genetic analyses to characterize and identify the population structure of those samples. Finally, we specifically look into whether any of our samples had a genetic affinity to non-Native American populations.

2716T

Mount Lebanon provides an opportunity to study DNA from the ancient Near East. *M. Haber¹, Y. Xue¹, C. Scheib², C. Doumet-Serhal³, T. Kivisild⁴, C. Tyler-Smith¹.* 1) Wellcome Sanger Institute, Cambridge, United Kingdom; 2) Estonian Biocentre, Institute of Genomics, University of Tartu, Estonia; 3) The Sidon excavation, Saida, Lebanon; 4) Department of Archaeology, University of Cambridge, Cambridge, United Kingdom.

The hot climate of the Near East has limited the study of ancient DNA from this region. Here, we sequenced five whole genomes from ~1,600-year-old individuals who lived during the Roman period in a village in Mount Lebanon. We consistently found surviving endogenous DNA in the petrous bones of the ancient individuals and we attribute this success to the mild climate of Mount Lebanon. The ancient individuals derived most of their ancestry from a population modelled by previously-reported Bronze Age individuals who lived on the Lebanese coast or inland in Jordan. In addition, we found steppe-like ancestry in the Roman Period individuals which we have previously detected in present-day Lebanese but not in Bronze Age individuals. This supports our previous proposition that the steppe ancestry penetrated the region more than 2,000 years ago, and genetic continuity in Lebanon is substantial. Our results show that individuals inhabiting the Mount Lebanon range likely shared similar ancestry with contemporaneous people living elsewhere in the region, but have higher chances of endogenous DNA surviving today and thus provide an opportunity to study past events more easily than in lowland Near Eastern samples.

2717F

Ancient DNAs from China and the origin of the Neolithic Chinese super-grandfather Y haplotypes. S. Huang¹, Y. Zhang¹, X. Lei¹, H. Chen¹, H. Zhou². 1) Central South University, Changsha, China; 2) Jilin University, Changchun, China.

Previous studies identified 3 Neolithic Han Chinese super-grandfather Y haplotypes, F5, F46, and F11, but their relationships with the fossil and archaeology records remain unexplored. We here report genome wide DNA sequencing data (1) for 12 ancient human samples from China ranging from 6500 to 3000 years before present. They were found to lie largely within the variation range of present day East Asians and belong to 4 different groups based on autosomal distances among them, Dashanqian (DSQ) of Xiajiadian culture in the Northeast, Banpo (BP) of early Yangshao culture in the Central West, Zhengzhou Xishan (ZX) of Miaodigou culture in the Central Plains, and Others. Present day F5 samples were closer to the ZX and DSQ groups while F11, C, O1, and O2 samples were closer to the BP group. We also sequenced the Y chromosome of one of the DSQ samples, K12, and found him to carry the haplotype O2a2b1a1a1a2a-F2137. Our analyses indicated gene flows from ZX to DSQ and to Southwest Chinese carrying the F5 subtype O2a2b1a1a6. These results linked F5 with the Miaodigou culture in Central China and F11 with the early Yangshao culture, which is remarkably consistent with archaeological and historical records.

2718W

Tracing the evolution of pigmentation-associated variants in Europe.

D. Ju, I. Mathieson. Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

Human skin pigmentation is highly differentiated among populations. This variation has been shaped by both demography and selection driven by adaptation to local levels of UV radiation. Ancient DNA studies have provided direct evidence that alleles associated with light pigmentation in present-day Europe were at much lower frequencies in the past 10,000 years. However, to date, these studies have focused on a small number of large-effect loci and have not considered the wide range of variants involved in this polygenic trait. To study the evolutionary history of pigmentation-associated variants in Europe, we first compiled a list of 43 variants associated with skin pigmentation in present-day populations. From this list, we computed pigmentation polygenic risk scores (PRS) for 1,030 ancient humans dating from 45,000-100 years BP as well as 2,814 present-day individuals. The oldest ancient Europeans (45-20,000 years BP) have pigmentation-associated variation within the range of present-day African populations, as shown both by the PRS and PCA of pigmentation-associated variants. Light pigmentation-associated variation increased in frequency over time ($p=5.36E-33$), a trend which persists when controlling for geography and ancestry ($p=0.00114$). All three major populations present in Holocene Europe (Early Farmers, Hunter-Gatherers and Steppe) show similar increases in light pigmentation-associated variation. While the majority of light pigmentation associated alleles rose in frequency over time, others decreased due to drift and gene flow. To better understand the contribution of individual variants, we examined the behavior of each SNP in our PRS. We determined the factors driving each SNP with regression models and found that 17% of SNPs showed strong evidence for selection while 44% and 22% were driven by ancestry and geography respectively, demonstrating that only a subset of pigmentation-associated loci were strongly selected. Furthermore, we found that a majority of SNPs under selection were polymorphic in the earliest Europeans. We therefore demonstrate a robust trend of depigmentation in Europe over the past 45,000 years driven by both selection and demographic transitions. Finally, we investigate similar trends in the evolution of variants associated with hair and eye pigmentation.

2719T

Investigating Sardinian population history with ancient DNA. *J.H. Marcus¹, C. Posth^{2,8}, H. Ringbauer¹, L. Lai¹¹, C. Sidore^{4,5,6}, J. Beckett¹³, H. Al-asa-dī^{2,3}, A. Furtwängler⁹, M. Zoledziewska⁴, C. Chiang¹², K. Dey³, M. Gradoli¹⁴, G. Abecasis⁵, D. Schlessinger⁹, R. Skeates¹⁰, J. Krause^{7,8}, J. Novembre^{1,2}, F. Cucca^{4,6}.* 1) Department of Human Genetics, University of Chicago, Chicago, IL; 2) Department of Ecology and Evolution, University of Chicago, IL; 3) Department of Statistics, University of Chicago, Chicago, IL; 4) Istituto di Ricerca Genetica e Biomedica, CNR, Cagliari, Italy; 5) Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, MI; 6) Dipartimento di Scienze Biomediche, Università di Sassari, Sassari, Italy; 7) Max Planck Institute for the Science of Human History, Jena, Germany; 8) Institute for Archaeological Sciences, Archaeo- and Palaeogenetics, University of Tübingen, Tübingen, Germany; 9) Laboratory of Genetics, NIA, NIH, Baltimore, MD; 10) Department of Archaeology, Durham University, Durham, England; 11) Department of Anthropology, University of South Florida, Tampa, FL; 12) Center for Genetic Epidemiology, Keck School of Medicine, University of Southern California, Los Angeles, CA; 13) Independent Contractor, Cagliari, Italy; 14) School of Archaeology and Ancient History, University of Leicester, Leicester, United Kingdom.

The sequencing of ancient DNA (aDNA) has provided new understanding into human movement and demography for many regions around the globe. For mainland Europe, ancient DNA studies have revealed a dynamic history, with major inferred population influxes due to Neolithic and Bronze Age expansions. The population of the Mediterranean island of Sardinia has been notable in these studies—typically aDNA samples of the early Neolithic on mainland Europe cluster with modern Sardinian samples. The standing model is that Sardinia had a high influx of Neolithic ancestry followed by relative isolation from the mainland and subsequent Bronze Age expansions. To gain further insight, we analyze genome-wide capture data (~1.2 millions SNPs) of 26 ancient Sardinians spanning the Neolithic, Copper Age, and Bronze Age, including individuals from Sardinia's Nuragic culture. Merging this novel data with 998 previously studied aDNA samples from across Europe and throughout the last ten millennia, we are able to place the ancient Sardinian samples into the broader context of the peopling of Europe. We confirm that ancient Sardinian samples show a strong affinity to early Neolithic samples and a near complete absence of the “Steppe” ancestry associated with Bronze Age expansions on the mainland. Interestingly, we also detect elevated affinities with pre-Neolithic peoples of Europe. Moreover, we studied genetic change through time within Sardinia. To this end, we analyzed whole-genome sequence data from approximately 1,500 modern Sardinian individuals, densely sampled across much of the island. Using our ancient samples enables us to detect significant signs of recent admixture, in particular with a strong influence from the Mediterranean region. We also find that populations from the more isolated mountainous provinces of Sardinia are less admixed and have experienced high levels of genetic drift. Overall, our analysis allows us to shed new light on the intriguing history of the peopling of Sardinia.

2720F

Reversing the bottleneck: Neanderthals as a genetic reservoir for functional alleles lost in the Eurasian migration out of Africa. *D.C. Rinker¹, C.N. Simonti², J.A. Capra^{1,3}, D. Shaw², E. Hodges³.* 1) Dept Biological Sciences, Vanderbilt University, Nashville, TN; 2) Dept Biological Sciences, Georgia Institute of Technology, Atlanta, GA; 3) Vanderbilt Genetics Institute, Vanderbilt School of Medicine, Nashville, TN.

The ancestral populations of Neanderthals and modern Eurasians each experienced genetic bottlenecks during their respective migrations out of Africa (OOA), with each population likely retaining ancestral variants lost in the other. Thus, introgression of Neanderthal DNA into the ancestors of Eurasians had the potential to reintroduce variants lost in the later OOA bottleneck. To evaluate the likelihood of this scenario, we simulated the two OOA bottlenecks followed by introgression under a range of evolutionary parameters. The simulations demonstrate that reintroduction of lost alleles by introgression was likely very common, predicting one reintroduced lost allele (RA) for every two Neanderthal-derived alleles in modern Eurasians. Furthermore, our models show a disparity in the fitness consequences of these alleles depending upon their evolutionary trajectory into modern Eurasians. Ancient variants introgressed into Eurasians exclusively through Neanderthals are predicted to be more deleterious than variants carried OOA with Eurasians, yet less deleterious than alleles of Neanderthal origin. Analyses of introgressed sequences in modern Eurasians identify tens of thousands of candidate RAs and support the results of our simulations. The observed ratio of RAs to Neanderthal-derived alleles (1:1.9) is very close to the ratio observed in simulations. Moreover, variant effect predictions using both functional (RegulomeDB) and machine learning (CADD) approaches confirm that RAs are less deleterious than Neanderthal-derived variants. We also identified 34 RAs that are eQTLs in lymphoblastoid cells from both Europeans and non-introgressed Africans. This indicates that RAs are often functional. These results suggest that introgressed haplotypes harbor variants with mixtures of fitness effects based on their origins. This raises the provocative possibility that the persistence of some Neanderthal variants may be due to genetic hitchhiking with functional RAs. Research into the consequences of Neanderthal introgression in modern Eurasians has focused almost exclusively on alleles of Neanderthal origin. Here we highlight the presence of ancient alleles on introgressed haplotypes. These RAs were African in origin but lost in the Eurasian OOA migration. Neanderthal introgression reintroduced these variants back into Eurasian populations, and their less deleterious fitness effects may have facilitated the persistence of some Neanderthal variants.

2721W

Ancient DNA reveals signatures of selection on disease-associated variants from GWAS. C.N. Simonti, J.L. Lachance. Department of Biological Sciences, Georgia Institute of Technology, Atlanta, GA.

Selection during recent human evolution has altered allele frequencies of both adaptive and disease-causing variants in modern populations. Genetic data from ancient humans has provided new evidence in the study of adaptive loci thought to be under historic selection, and thus is a powerful tool for identifying instances of selection that might be missed by methods that use present-day samples alone. Using a curated set of ~2,700 variants from the NHGRI-EBI GWAS Catalog and samples from 20 ancient DNA studies, we provide an analysis to identify disease-associated variants that bear signatures of selection over time. We accounted for the fact that not every ancient individual contributed equally to modern genomes by determining the maximum potential contribution of 145 ancient samples to modern individuals of European descent. We then used a Bayesian inference method to infer allele frequency trajectories over time, and determine which disease-associated loci exhibit signatures of natural selection. Using a demographic model of constant population size over time, we find that 557 of 2,716 variants (20.5%) show signatures of modest selection ($s > 0.001$). Unsurprisingly, variants found in the MHC locus show increased signatures of selection (58 of 148; 39.2%), but do not drive the overall pattern. Additionally, we find that protective alleles are more likely to increase in frequency over time (i.e. they appear to be positively selected) than alleles that increase the risk of complex diseases; however, allele frequencies of many risk alleles have increased over time. While these patterns could arise due to phenomena such as pleiotropy or genetic hitchhiking rather than selection, they are still informative about changes in associated traits over time. Overall, these results demonstrate the potential of ancient DNA to improve our understanding of how variants associated with disease or benign phenotypes have changed in frequency over time and the likelihood that selection has acted on them. Ancient samples allow us an unprecedented look at how our species has changed over time, which can only improve with additional samples and improvements in methods used to infer selection from time series data. Our results represent an important early step in using this new source of data to better understand how disease risk has changed over time.

2722T

Detecting tracts of introgressed Neanderthal ancestry in genomic sequence data of modern humans. M. Steinruecken^{1,2,3}, J.P. Spence⁴, J.A. Kamm⁵, E. Wiecek⁶, Y.S. Song^{3,5,7}. 1) Department of Ecology and Evolution, University of Chicago, Chicago, IL, USA; 2) Department of Biostatistics and Epidemiology, University of Massachusetts, Amherst, MA, USA; 3) Department of EECS, University of California, Berkeley, CA, USA; 4) Computational Biology Graduate Group, University of California, Berkeley, CA, USA; 5) Department of Statistics, University of California, Berkeley, CA, USA; 6) Department of Mathematics, University of California, Berkeley, CA, USA; 7) Chan Zuckerberg Biohub, San Francisco, CA, USA.

Genetic evidence revealed that the ancestors of non-African humans and their hominin sister groups, notably the Neanderthals, exchanged genetic material in the past, and traces of this introgression can still be found in sequence data of modern day humans. Uncovering these tracts has received a lot of attention in recent years. Their distribution not only provides insight into the complex population structure and migration history relating these ancient populations, it also enables studying the adaptive processes that lead to the accumulation or depletion of introgressed genetic variants in certain regions of the genome. Recent efforts to detect these introgression tracts reported long regions depleted of Neanderthal ancestry that are enriched in genes, suggesting general negative selection against the Neanderthal variants. Furthermore, some allelic variants associated with genetic diseases are likely to originate from the Neanderthal population. On the other hand, enriched Neanderthal ancestry in hair and skin related genes suggests that the introgressed variants helped modern humans to adapt to non-African environments. Here we present a Hidden Markov Model framework that allows efficient detection of the introgressed tracts, using a reference sequence for the Neanderthal population. We demonstrate through extensive simulations that our method can be used to accurately and efficiently detect introgression. Furthermore, we apply our method to sequence data of modern humans from the 1000 genomes project and a high coverage Neanderthal individual from the Altai mountains. Our results are in general agreement with previously obtained results, and we will discuss similarities, differences, and functional implications.

2723F

LILRB1/ILT2 polymorphisms detected in Brazil reveal a conserved locus with polymorphism frequencies influenced by ancestry. E.C. Castelli¹, M.L.G. Oliveira², M.A. Paz¹, A.S. Souza¹, L.C. Veiga-Castelli², L. Marcorin², G. Debortoli², A.L.E. Pereira², N.C.A. Fracasso², G.V. Silva³, A.L. Simões², A. Sabbagh⁴, E.A. Donadi⁵, C.T. Mendes-Junior². 1) São Paulo State University (UNESP), Molecular Genetics and Bioinformatics Laboratory, School of Medicine, Botucatu, State of São Paulo, Brazil; 2) Departamento de Genética, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, 14049-900, Ribeirão Preto-SP, Brazil; 3) Departamento de Química, Laboratório de Pesquisas Forenses e Genômicas, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, 14040-901, Ribeirão Preto-SP, Brazil; 4) UMR 216 IRD MERIT, Faculté de Pharmacie de Paris, Université Paris Descartes, Paris, France; 5) Departamento de Clínica Médica, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, 14049-900, Ribeirão Preto-SP, Brazil.

LILRB1 (ILT2) gene is a member of the Leukocyte Immunoglobulin-like Receptor (LIR) family. *ILT2* encodes an important NK and T cell inhibitory receptor, with the highly polymorphic Human Leukocyte Antigen (HLA) class I molecules as its major ligands. Variability at both ligand and receptor might influence this interaction. *ILT2* DNA and protein variability have been poorly explored, with few studies evaluating only well-documented variants. We evaluated *ILT2* variability at exon segments, translating sequences into full-length protein molecules. Variable sites were assessed in 384 samples from the Brazilian Southeast using the Haloplex technology. Due to sequence similarity with other LIR genes, we expected genotyping errors when using second-generation sequencing. Hence, we developed an approach to score each read based on officially recognized LIR and KIR sequences, addressing them to the most likely locus. Ancestry was assessed using the SNPforID 34-plex panel. Ancestry composition of the population sample was 70.7% European, 21% African, and 8.3% Amerindian/Asian. Samples were stratified into three subgroups: EUR90, including 126 samples with a European ancestry higher than 90%; EUR50, including 307 samples with major European ancestry (> 50%); and AFR50, including 45 samples with major African ancestry (> 50%). Altogether, we detected 58 variable sites, arranged into 87 different coding sequences that encode 17 unique full-length proteins. Three of those proteins account for 72% of the sample, ten presented frequencies higher than 1% and a frequent protein (11%) was not predicted by the *1000Genomes* data. Two of the most frequent proteins here predicted were two times more frequent at the EUR groups than at the AFR50. Moreover, many frequent proteins were detected exclusively at either the AFR50 or EUR groups. This observation is in agreement with the *1000Genomes* data, which also indicates variation on frequencies influenced by ancestry. Although conserved at the protein level when compared with its major HLA ligands, *ILT2* protein frequencies are strongly influenced by the sample ancestry background, thus population stratification must be considered when conducting association studies regarding *ILT2*. Nonetheless, further studies are necessary to elucidate whether these polymorphic *ILT2* molecules interact differently with HLA antigens, and if the frequency shifts result from selective pressures or genetic drift.

2724W

Defining trait core genes with networks. B.E. Graham^{1,2}, K. Chesmore³, S.M. Williams^{1,2}. 1) Systems Biology and Bioinformatics, Case Western Reserve University, Cleveland, OH 44106, U.S.A.; 2) Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, Ohio, 44106, U.S.A.; 3) Department of Molecular and Systems Biology, Geisel School of Medicine, Dartmouth College, Hanover, NH 03755, U.S.A.

The concept that many genes affect a single trait, i.e., the omnigenetic model, has recently gained momentum due to the observation from GWAS that numerous phenotypes each associate with hundreds of variants. Height is one such omnigenetic, complex trait. It is highly heritable with ~700 GWAS associated loci in the GWAS catalog. We propose a network-based analysis to determine which height associated loci are most important to the presumed genetic architecture of the trait, i.e. the "Core" genes as proposed by Boyle, LI and Pritchard, 2017. Using proteomic, pathway and gene interaction-based analyses of the height associated loci, we found that the best networks used confidence of relatedness between genes, edge weight, to determine nodes (genes) most important to the network (i.e., the Core genes). We investigated the pleiotropy of height genes and the relationship between phenotypes linked to those genes using a multiple phenotype-gene mapping method based on the GWAS catalog. Our resulting network revealed the number of Core genes to be approximately one third of the total, with several well-defined gene clusters, some concentrated around several oncogenes, many of which associate with more than one cancer. Phenotype-gene clusters include both height and BMI related traits as well as fetal size and preterm birth related traits. This last observation is supported by these Core genes being expressed with high confidence in placental and fetal tissues. This study shows that networks are a powerful tool for the mathematical identification of Core genes for complex disease.

2725T

The identification of new protein methylation site acquisition in conserved proteins in human lineage. *D.S Kim, J.K Kim.* Department of Rare Disease Research Center, Korea Research Institute of Bioscience.

Protein methylation plays an important role in various biological processes, including chromatin structure remodeling, gene transcription, DNA repair, protein synthesis, RNA metabolism, cell cycle progression, apoptosis and signal transduction. The acquisition of a novel methylation site can have a significant effect on protein structure and function, and this can be seen as a phenotype. We analyzed the human methylation site data and multiple alignments of mammalian proteins including those from human, primate, and other placental mammals. In our analysis, we identified 21 novel protein methylation site in 20 protein that arose in the human lineage during primate evolution. Among them, Arg-320 of Proline and serine-rich protein 2 (PROSER2) and Arg-513 of Sodium channel protein type 5 subunit alpha (SCN5A), gained human specific methylation site after the human-chimpanzee divergence. Notably, the Arg-513 in Sodium channel protein type 5 subunit alpha (SCN5A) is expected to modulate the intracellular calcium levels by affecting the sodium calcium channel. The results of molecular evolutionary analysis suggested that these sites were under positive selection during human evolution. We suggest that the novel methylation sites identified in this study may be useful candidate for functional analysis to identify progressive genetic modifications to advantageous phenotypes acquired in the human lineage.

2726F

Genome-wide identification and characterization of non-coding human specific deletions. *J. Xue^{1,2}, S.K. Reilly^{1,2}, R. Tewhey³, P.C. Sabeti^{1,2}.* 1) Harvard University, Cambridge, MA 02138; 2) Broad Institute, Cambridge, MA 02142; 3) The Jackson Laboratory, Bar Harbor, ME 04679.

The identification and characterization of human-specific regulatory elements is crucial towards understanding our unique evolutionary history. Particularly compelling are highly conserved regions of the primate genome that are deleted specifically in humans. Prior work has focused on identifying large and intermediate sized (>20bp) human Conserved DEletions (hCONDELs) which have been shown to reside almost entirely in non-coding regions and are enriched near genes involved in steroid hormone receptor signaling and neural functions. However, smaller deletions, which are in much greater abundance, have not been rigorously identified or characterized for their function due to various technical challenges. Using multiple sequence alignments generated by our group from the latest available primate reference genomes, we find that the vast majority of putative hCONDELs (>95%) are very small, suggesting that most hCONDELs are yet to be explicitly described. We computationally validated our hCONDELs by ensuring that these putative deletion sites are completely removed from a diverse set of human populations from the Simons Genome Diversity Project. To characterize the functional importance of the deletions, we employed the Massively Parallel Reporter Assay (MPRA) as a non-coding screen to assess the cis-regulatory impact of 17,197 deletions. We utilized MPRA across a diverse range of cell types to test for function across different tissue backgrounds. hCONDELs which were found to be active were enriched to lie in active regulatory regions of specific tissues. Furthermore, several hCONDELs show significant differences in activity via reporter assay between their native human and chimp contexts, and are prime candidates for contributing to possible human-specific phenotypes. We further characterize these hCONDELs to show changes in epigenetic activity at these loci and also transcriptional changes in nearby genes between human and chimpanzee. On a broader scale, we explored genome-wide trends in transcription factor binding, motif usage, and chromatin activity associated with those sites displaying species-specific activity. The characterization of these sites allows us to construct a clearer picture of the functional significance for sequences lost along the human lineage.

2727W

The orangutan has all the variants you need: Monitoring specificity and sensitivity of diagnostic next-generation-sequencing by establishing orangutan DNA as reference standard. A. Rump, F. Kuhlee, L. Gieldon, K. Hackmann, E. Schrock. Institute for Clinical Genetics, Faculty of Medicine Carl Gustav Carus, TU Dresden, Germany.

Evaluating the performance of diagnostic sequencing requires well-characterized reference samples (RS) with validated genetic variations. Since human RS tend to harbor relatively few variations within their coding regions, there will always be numerous exons from almost any gene in a given panel where a sequence variation will not be available for quality control (QC). So far, the effort to determine an assay's performance is immense, in terms of both money and time. In recent published cases, where NGS sensitivity and specificity has been assessed by sequencing up to 150 clinical specimens in order to provide 200-300 disease-relevant sequence variants. While this effort might be justified for an initial assay set-up, it is certainly not feasible for consistent QC. In order to overcome these problems we tested an orangutan (OrU) as donor for QC reference materials. So far however, OrU NGS reads have only been obtained from whole genome sequencing (WGS). Therefore, information of how OrU/human sequence divergence affects targeted enrichment by human capture probes is limited. Here, we applied capture-based whole exome sequencing (WES) panels from two different suppliers to gDNA from the commercially available male OrU cell line "EB185-JC" (Sigma-Aldrich). As positive control, the well-defined male HapMap sample NA12877 was sequenced in parallel, along with a 1:1 gDNA mixture of EB185-JC and NA12877. The reads from each of these experiments were mapped to the human reference, the OrU reference and to a combined human/OrU "mix" reference. Using this approach, we could show that OrU exons can be enriched, sequenced and mapped with the same efficiency as the competing human DNA. When mapped to the human reference, the OrU exome provides 903.763 "on target" variants (average 4.4 variants/WES target) and many of these variants would be disease causing if present in a human being. Therefore, OrU DNA can be used to measure both analytical and diagnostic sensitivity for almost every human exon and exceeds the variant variety of NA12877 DNA (average 0.3 variants/WES target) by a factor of 13.5 fold. Therefore, OrU DNA is especially valuable for small, disease-focused NGS panels where human reference DNA fails to provide enough variants for a reliable QC. Furthermore, specific OrU variants, which are very unlikely to occur in humans (neither naturally nor as sequencing artefacts), can be used to trace cross-contamination of reads in highly multiplexed NGS experiments.

2728T

Prediction of eye and hair pigmentation phenotypes using the HlrisPlex system in a Brazilian admixed population sample. C.T. Mendes-Junior¹, T.M.T. Carratto¹, F.D.N. Souza¹, L. Marcorin², G.V. Silva¹, A.L.E. Pereira², N.C.A. Fracasso³, M.L.G. Oliveira², G. Debortoli², E.A. Donadi³, A.L. Simões², E.C. Castelli⁴. 1) Departamento de Química, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, SP, Brazil; 2) Departamento de Genética, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, SP, Brazil; 3) Departamento de Clínica Médica, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, SP, Brazil; 4) Departamento de Patologia, Faculdade de Medicina de Botucatu, Universidade Estadual Paulista, Botucatu, SP, Brazil.

Human pigmentation is a complex trait, probably involving more than 100 genes. Predicting phenotypes using SNPs present in those genes is of large forensic interest. For this, the Hlrisplex tool was developed for simultaneous eye and hair color prediction, achieving high accuracy among Europeans (76% and 82%, respectively). The application of HlrisPlex in admixed populations has not been deeply evaluated. Prediction of intermediate color phenotypes may be impaired in such populations, since completely different gene-gene interactions may give rise to phenotypes in an admixed background. This study aims to analyze this set of 24 markers in 340 individuals from Ribeirão Preto (SP) region, predicting eye and hair color and comparing these predictions with actual individual phenotypic data. DNA was extracted from blood samples through a salting-out procedure. Fragment libraries including the HlrisPlex markers were created using Haloplex (Agilent Technologies, Inc). Sequencing was held in MiSeq (Illumina) platform and reads were trimmed and aligned to the hg19 reference genome using cutadapt and BWA packages. GATK tool kit and VCFx packages were employed for SNP genotyping. Population genetics analysis was performed using Arlequin 3.5 software. The prediction of eye and hair color was accomplished by the HlrisPlex online tool. Using the default prediction settings, HlrisPlex presented an overall accuracy of 64.2% for hair color prediction, ranging from 29.9% for light-brown to 98.8% for black. Regarding eye color, an overall accuracy of 78.5% was achieved. Although higher accuracy levels were obtained for individuals that have blue (86.5%) and brown (97.5%) eyes, HlrisPlex completely failed in predicting intermediate eye color. Combining eye and hair predictions, the hit rate obtained was only 53.0%. Even when a threshold of $p > 0.8$ is established for eye color determination, although 20.3% of individuals are left aside, blue and brown eyes accuracy increases to 93.3% and 98.5%, respectively, with an eye color accuracy of 83.4% and overall hair/eyes combined accuracy of 58.2%. Although HlrisPlex proved to be an applicable tool for determination of the most extreme pigmentation phenotypes, it fails in determining intermediate phenotypes, decreasing the overall accuracy of combined predictions and its reliability when used in admixed populations. FINANCIAL SUPPORT: FAPESP (Grant 2013/154470) and CNPq/Brazil (Grant 448242/2014-1 and Fellowship 309572/2014-2).

2729F

Inference and visualization of geographic patterns of variation for variant sets. A. Biddanda¹, J. Novembre^{1,2}. 1) Department of Human Genetics, University of Chicago, Chicago, IL; 2) Department of Ecology and Evolution, University of Chicago, Chicago, IL.

Most summaries of population structure in genetic data convey the relative similarity of individuals in a sample. However, it is also useful to have a variant-centered view of population structure. For instance, one does not fully understand population structure until one knows the proportion of variants that are shared across all populations versus those that are geographically restricted. This information is relevant for understanding evolutionary history, for designing and interpreting GWAS studies, and more broadly, for building precision medicine solutions that are relevant to all populations. Currently, existing tools for summarizing population structure fail to convey such differences in the absolute frequencies of variants between different populations. To represent patterns of variant-level population structure, we have developed an approach with two major features 1) a visualization toolkit to succinctly display the geographic patterns of multiple variants and 2) a probabilistic clustering algorithm for grouping variants by their absolute frequencies across many populations. For our visualization method, we show heatmaps customized to represent variants by their frequency patterns. When clustering, our method infers a set of “modal frequency patterns” that capture the majority of the variation in the geographic distribution of variants, while also inferring the relative abundance of each modal frequency pattern. Following the inference of these consensus frequency patterns, we then assign variants to their most likely underlying frequency and subsequently calculate enrichment and depletion of particular categories within subsets of variants. Applying our method across a number of publicly available human datasets, we see features of human variation that are expected outcomes from the expansion out of Africa and recent explosive population growth. We also applied our method to different variant sets. For example, among variants in the NHGRI GWAS catalog, we observe an enrichment of European common variants, reflecting the predominantly European ancestry composition of many large-scale GWAS studies. Overall we show that our new methods provide a useful tool for the visualization of multi-dimensional population genomic data, allowing for the characterization of population structure from a variant-centric view that we expect will be of broad utility to the population genetic community.

2730W

Gene-level intolerance to copy number variation in UK Biobank reveals a cohort-level survival bias and novel disease associations. M. Aguirre^{1,2,3}, M. Rivas², J. Priest^{1,3}. 1) Pediatric Cardiology, Stanford University, Stanford, CA; 2) Biomedical Data Science, Stanford University, Stanford, CA; 3) Cardiovascular Institute, Stanford University, Stanford, CA.

Copy number variations (CNV) constitute a significant proportion of the total genetic differences between individuals, and many CNVs associate causally with a variety of health outcomes. Here, we use genetic and phenotype data from UK Biobank to measure genic tolerance to rare structural variation using previously developed methods for determining regional selective constraint. We perform separate constraint analyses for partial gene deletion, whole-gene duplication, as well as any overlapping copy number event as the basis of bayesian burden tests between rare genic CNVs and over 800 derived phenotypes. We observe the strongest intolerance to duplication in developmental (*CSH2* [MIM 118820], *GH1* [MIM 139250]), reproductive (*TUBB8* [MIM 616768]), and autoimmune-related genes (*HLA-DRB1* [MIM 142857]), as well as intolerance to any CNVs overlapping canonical cancer susceptibility genes (*BRCA2* [MIM 600185], *BRCA1* [MIM 113705], *APC* [MIM 611731], *ATM* [MIM 607585], *MSH2* [MIM 609309], and *MLH1* [MIM 120436]). For specific disease associations we find a significant CNV burden in *DLEU1* [MIM 605765] and *DLEU7* [MIM 605766] in leukemia ($p < 10^{-16}$), *PMP22* [MIM 601097] in hereditary peripheral neuropathy ($p < 10^{-16}$), and in *PNKD* [MIM 609023] in stroke ($p = 4.5 \times 10^{-6}$). These associations are consistent with known pathogenic consequences of LoF variants in these genes, while offering novel information on potential disease associations. Ranked intolerance scores suggest a cohort-level selection bias against CNVs conferring risk for lethal conditions with onset in childhood or early adulthood. Our preliminary results recapitulate the role of structural variation in the genetic architecture of known disease, offer novel disease-associations, and contextualize the absence of population-wide burden of CNVs associated with early-onset genetic disease in the UK Biobank.

2731T

Structural variation of the human population: Characterizing and sequence-resolving variation to identify common alleles, correct reference errors, and improve short-read analysis. *P. Audano¹, A. Sulovari¹, T. Graves-Lindsay², S. Cantsilieris¹, M. Sorensen¹, A. Welch¹, M. Dougherty¹, B. Nelson¹, S. Dutcher², W. Warren², R. Wilson^{3,4}, E. Eichler^{1,5}.* 1) Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA 98195, USA; 2) McDonnell Genome Institute, Department of Medicine, Department of Genetics, Washington University School of Medicine, St. Louis, MO 63108, USA; 3) Institute for Genomic Medicine, Nationwide Children's Hospital, Columbus, OH 43205; 4) The Ohio State University College of Medicine, Columbus, OH 43210; 5) Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA.

Although the accuracy of the human reference genome is critical for basic and clinical research, structural variants (SVs) have been difficult to assess because data capable of resolving them have been limited. To address potential bias, we sequenced a diversity panel of nine human genomes to high depth using long-read, single-molecule, real-time sequencing data. Systematically identifying and merging SVs ≥ 50 bp in length for these nine and one public genome yielded 83,909 sequence-resolved insertions, deletions, and inversions. Among these, 2,839 (2.0 Mbp) are shared among all discovery genomes with an additional 13,349 (6.9 Mbp) present in the majority of humans, indicating minor alleles or errors in the reference, which is partially explained by an enrichment for GC-content and repetitive DNA. Genotyping 83% of these in 290 additional genomes confirms that at least one allele of the most common SVs in unique euchromatin are now sequence-resolved. We observe a 9-fold increase within 5 Mbp of chromosome telomeric ends and correlation with *de novo* single-nucleotide variant mutations showing that such variation is nonrandomly distributed defining potential hotspots of mutation. We identify SVs affecting coding and noncoding regulatory loci improving annotation and interpretation of functional variation. To illustrate the utility of sequence-resolved SVs in resequencing experiments, we mapped 30 diverse high-coverage Illumina-sequenced samples to GRCh38 with and without contigs containing SV insertions as alternate sequences, and we found these additional sequences recover 6.4% of unmapped reads. For reads mapped within the SV insertion, 25.7% have a better mapping quality, and 18.7% improved by 1,000-fold or more. We reveal 72,964 occurrences of 15,814 unique variants that were not discoverable with the reference sequence alone, and we note that 7% of the insertions contain an SV in at least one sample indicating that there are additional alleles in the population that remain to be discovered. These data provide the framework to construct a canonical human reference and a resource for developing advanced representations capable of capturing allelic diversity. We present a summary of our findings and discuss ideas for revealing variation that was once difficult to ascertain.

2732F

A population genetics approach to discover genome-wide saturation of structural variants from 22,600 human genomes. *L.F. Paulin¹, A. von Haeseler^{1,2}, F.J. Sedlazeck³.* 1) Center for Integrative Bioinformatics Vienna, Max F. Perutz Laboratories, University of Vienna, Medical University of Vienna, Vienna, Austria; 2) Bioinformatics and Computational Biology, Faculty of Computer Science, University of Vienna, Vienna, Austria; 3) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA.

Structural variants (SV) are an important source of human variations and shown to impact evolution, genetic diseases (e.g. Mendelian diseases and cancer) and sometimes cell regulation. SVs are defined as 50bp or larger deletions, duplications, inversion, and translocations. The current state of the art infers locations of SVs by analysing short reads that are reported to lack sensitivity that can be improved with long reads that are often more expensive and thus cannot be applied when studying larger cohorts. A way out of this dilemma is the development of tools that predict if a certain genomic region is already saturated with SVs. Then one would simply not sequence it anymore and would only sequence not saturated regions with less expensive methods like capture approaches. Based on a multi ethnicity, whole genome sequencing collection (30x) of 22,600 human genomes we developed a statistical model to determine hotspots of diversity of SVs taking into account allele frequencies across the ethnicities. We applied the Ewens Sampling Formula (ESF) over 10kbp windows to predict the genomic regions that did not reach saturation of SVs although sequencing 22,600 WGS. ESF is a well known tool in population genetics reapplied to our question. The ESF describes the probability distribution of the frequency of how many distinct alleles were observed a given number of times under the infinite allele model. Once the parameter of the ESF is estimated from a genomic sample, we can predict the probability to find a new SV in any given window. This approach discovered regions such as HLA, a few regions distributed in the telomers and naturally the centromeres as regions unsaturated regions, with respect to SVs. These regions are potentially relevant for personalized medicine as they are often unique to a few samples and demand suggest a closer investigation with capture based sequencing. Our model is applied to large cohorts such as the 1,000 Human-Genome and the 1,001 Arabidopsis genome project and can be applied to other data-sets to estimate a lower and higher bound variation that can be expected with sequencing additional samples. Thus, our tool provides a more cost efficient way to analyze large population based studies. Already small samples provide enough evidence to sequence only potential hotspots in the genomes of interest rather blindly sequencing whole genomes.

2733W

Quantification of selective pressures on human pigmentation between populations and over epochs. X. Huang¹, S. Wang^{1,2}, Y. He³, L. Jin^{1,2}. 1) Chinese Academy of Sciences Key Laboratory of Computational Biology, Chinese Academy of Sciences-Max Planck Society Partner Institute for Computational Biology, Shanghai Institute of Nutrition and Health, Shanghai Institutes for Biological Sciences, University of Chinese Academy of Sciences, Chinese; 2) State Key Laboratory of Genetic Engineering and Ministry of Education Key Laboratory of Contemporary Anthropology, Collaborative Innovation Center for Genetics and Development, School of Life Sciences, Fudan University, Shanghai, 200433, China; 3) Institute of Biomedical Sciences, Shanghai Medical College, Fudan University, Shanghai, 200032, China.

The obvious diversity of human pigmentation—skin, hair, and eye colour—has attracted particular attention from both academic and non-academic investigators for thousands of years, as noted by Charles Darwin one century ago and as noticed by ancient Egyptians more than 4000 years ago. Why human pigmentation diverges still remains a central puzzle in human biology. Some researchers have proposed that human pigmentation diversity is adapted for ultraviolet radiation and driven by natural selection. During the last decade, many studies have detected selection signals in individual pigmentation genes; however, few studies have examined how multiple genes contributed to the evolution of human pigmentation. Moreover, none has quantified selective pressures on human pigmentation between populations and over epochs. Here we report the first systematic quantification of the differences and changes in selective pressures between distinct populations and during different periods with our new approaches. We suggest directional selection not only affected the light pigmentation in Eurasians, but also influenced the dark pigmentation in Africans. Thus, directional selection, rather than the relaxation of selection, is the more likely explanation of human pigmentation diversity, a long-standing debate in the evolution of human pigmentation. We also reveal that the selective pressures on human pigmentation changed markedly during the Out-of-Africa event, and the magnitude of the change has diminished more recently. Our results elucidate the quantitative details of human pigmentation diversity as well as its evolution, and may help researchers to facilitate studies for the evolution of other complex traits. .

2734T

Facial masculinity does not appear to be a condition-dependent male ornament and does not reflect MHC heterozygosity in humans. J.D. White¹, A.A. Zaidi^{1,2}, B.C. Mattern^{1,3}, C.R. Liebowitz¹, D.A. Puts¹, P. Claes^{4,5}, M.D. Shriver¹. 1) Department of Anthropology, The Pennsylvania State University, University Park, PA; 2) Department of Biology, The Pennsylvania State University, University Park, PA; 3) Department of Biobehavioral Health, The Pennsylvania State University, University Park, PA; 4) Department of Electrical Engineering, ESAT/PSI, KU Leuven, Leuven, Belgium; 5) Medical Imaging Research Center, MIRC, UZ Leuven, Leuven, Belgium.

Facial masculinity is thought to be a condition-dependent male ornament that may have evolved to indicate immunocompetence and other measures of 'quality' to potential mates and rival males. In non-human animals, male ornaments, such as stag antlers and beetle horns, tend to be exaggerated structures that are hypervariable and sensitive to subtle variation in growth among males. We tested whether this applies to facial masculinity by calculating an objective, vector-based measure of facial masculinity/femininity from three-dimensional images in a large sample ($N = 1,233$) of people of European ancestry. Using single nucleotide polymorphism (SNP) data from 567,787 markers, we calculated genome-wide heterozygosity as well as genomic ancestry principal components (PCs), to correct for population substructure. An additional measure of immunocompetence was calculated via MHC heterozygosity, estimated using 114 human leukocyte antigen (HLA)-tag SNPs previously reported to capture HLA haplotype variation in Europeans. After controlling for sex, age, weight, and genomic ancestry PCs, we show that facial masculinity is positively correlated with adult height in both males and females ($T\text{-statistic} = 8.81$, $p\text{-value} = 4.17 \times 10^{-18}$). This suggests that variation in growth contributes, at least in part, to variation in facial masculinity, which is characteristic of condition-dependent traits. However, facial masculinity scales with growth similarly in males and females ($b_{\text{Male}} = 0.227$, $b_{\text{Female}} = 0.299$, $Z\text{-score}_{\text{diff}} = -1.18$, $p\text{-value} = 0.120$), suggesting that facial masculinity is not specifically a male ornament, as male ornaments tend to be more sensitive to growth variation in males. Additionally, we show that while height is positively correlated with MHC heterozygosity ($T\text{-statistic} = 3.18$, $p\text{-value} = 1.54 \times 10^{-3}$), facial masculinity is not, regardless of whether height is included in the model ($T\text{-statistic} = -0.038$, $p\text{-value} = 0.970$) or not ($T\text{-statistic} = 0.586$, $p\text{-value} = 0.558$). Thus, facial masculinity does not reflect immunocompetence in humans, as previously thought. Future studies should explore the relationship between facial masculinity and other measures of immunocompetence. Overall, in this well-powered study, we find no support for the idea that facial masculinity is a condition-dependent male ornament that has evolved to indicate immunocompetence in humans.

2735F

A novel, scalable, and powerful IBD-based method to detect archaic hominin sequence in modern humans. *L. Chen^{1,2}, W. Fu³, J.M. Akey^{1,2}*

1) Ecology and evolutionary biology, Princeton University, Princeton, NJ; 2) Lewis Sigler Institute, Princeton University, Princeton, NJ; 3) 1 Microsoft Way, Microsoft, Redmond, WA.

A number of statistical methods exist to identify sequences that were inherited from archaic hominin ancestors, such as Neandertals and Denisovans. These methods typically require haplotype resolved data, make use of a reference population that does not segregate introgressed sequence, and are computationally intensive. Here, we describe a rigorous probabilistic method based on identity by descent (IBD) to detect Neandertal and Denisovan sequences in modern humans and evaluate its operating characteristics through extensive simulations. Importantly, we show that our approach can be used without a reference population, which increases power to find introgressed sequences. We also show that our IBD based method outperforms existing methods, such as *S**, and is much more computationally tractable as it operates on genotype data directly. Finally, we apply our method to the 1000 Genomes Project and other sequencing data sets from geographically diverse populations to refine the set of Neandertal and Denisovan sequences segregating in modern humans.

2736W

Localizing and classifying adaptive targets with trend filtered regression.

M.R. Mughal¹, M. DeGiorgio^{2,3}, 1) Bioinformatics and Genomics at the Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, PA; 2) Departments of Biology and Statistics, Pennsylvania State University, University Park, PA; 3) Institute for CyberScience, Pennsylvania State University, University Park, PA.

Identifying genomic locations of natural selection from sequence data is an ongoing challenge in population genetics. Current methods utilizing information combined from several summary statistics typically assume no correlation of summary statistics regardless of the genomic location from which they are calculated. However, due to linkage disequilibrium, summary statistics calculated at nearby genomic positions are highly correlated. We introduce an approach termed *Trendsetter* that accounts for the similarity of statistics calculated from adjacent genomic regions through trend filtering, while reducing the effects of multicollinearity through regularization. Our penalized regression framework has high power to detect sweeps, is capable of classifying sweep regions as either hard or soft, and can be applied to other selection scenarios as well. We find that *Trendsetter* is more robust to missing data than similar current approaches, has comparable power, and is also robust to strong background selection. Moreover, the model learned by *Trendsetter* is highly interpretable, as it represents a set of curves modeling the spatial distribution of summary statistics in the genome. Application to human genomic data revealed positively-selected regions previously discovered such as *LCT* in Europeans and *EDAR* in East Asians. We also identified a number of novel candidates and show that populations with greater relatedness share more sweep signals.

2737T

The Oceanian Genome Variation Project. C.D. Quinto-Cortés¹, A. Ioannidis², J.R. Homburger³, G.L. Wojcik⁴, J. Blanco-Portillo⁵, K. Auckland⁶, A. Severson³, C. Barberena-Jonas⁶, J.E. Rodríguez-Rodríguez⁷, J. Blanc⁷, G.M. Belbin⁸, S. Oppenheimer⁹, T. Parks³, M. Phipps⁹, K. Robson⁹, E.E. Kenny⁹, A.V.S. Hill⁹, A.J. Mentzer⁹, C. Bustamante⁴, A. Moreno-Estrada¹, G.R. Gignoux¹⁰ *On behalf of the Oceanian Genome Variation Consortium.* 1) National Laboratory of Genomics for Biodiversity, Irapuato, Guanajuato, Mexico; 2) Computational and Mathematical Engineering, Stanford University, Palo Alto, CA; 3) Department of Genetics, Stanford University, Stanford, CA, USA; 4) Department of Biomedical Data Science, Stanford University, Stanford, CA, USA; 5) University of Oxford, United Kingdom; 6) Undergraduate Program on Genomic Sciences, UNAM, Cuernavaca, Morelos, Mexico; 7) University of California, Davis, Davis, CA, USA; 8) Icahn School of Medicine at Mount Sinai, New York, NY, USA; 9) Monash University, Malaysia; 10) Colorado Center for Personalized Medicine and Department of Biostatistics and Informatics, University of Colorado, Aurora, Colorado, USA.

Oceania comprises the region from Southeast Asia throughout the South Pacific and has a rich demographic history, with the first human settlements in Papua New Guinea dating to ~40,000 years ago. Multiple migration and admixture events have occurred throughout the region, like the Lapita expansion that reached Tonga and Samoa and the colonization of Remote Oceania. Despite the richness of different peoples and cultures, this area of the world remains almost entirely left out of the global genetic mapping efforts. Here, we present the Oceanian Genome Variation Project (OGVP), the largest collection of historic genotyped DNA samples collected by the University of Oxford. The main aims of this project are to reconstruct the population genetic history of the major human migrations into the Pacific and to increase resolution on the characterization of understudied regions. We generated genome-wide SNP data for 980 individuals using the MEGA platform and identified genetic clusters of population differentiation with ADMIXTURE. They broadly correspond to Melanesia, Micronesia and Polynesia, recapitulating cultural and linguistic differences among the islands. In particular, there is decreasing continental East Asian ancestry throughout Malaysia and Indonesia, including some of the remote islands as we move east away from the mainland. The ADMIXTURE patterns seen in the westernmost islands (Tonga, Samoa, and Tuvalu) contrast with the remainder of the Polynesian islands as they have both of the components seen in Polynesia and Melanesia. In addition, Remote Polynesia shows patterns of pervasive European admixture, which contact will be dated to reconstruct the recent demographic dynamics of the region triggered by intercontinental travel. We used BEAGLE to find identity-by-descent (IBD) segments between individuals in Oceania and found that the amount of IBD shared between different populations increases in an eastward direction and that Polynesian populations share the highest amount of IBD. Lastly, we identified maternal and paternal lineages in this dataset. The most common mitochondrial haplogroups were B (namely B4a1), Q and M. Most of the Y chromosome lineages found in Oceania were O, C, K, M and R. The OGVP initiative represents an important effort to include a historically underrepresented area of the world in the publicly available genetic databases and to facilitate the genetic research in the Pacific.

2738F

Detecting signatures of adaptive positive selection from high-density genotyping data in the Lithuanian population. A. Urnikyte¹, M. Monda², E. Bosch³, A. Molyte¹, V. Kučinskas¹. 1) Department of Human and Medical Genetics, Biomedical Science Institute, Faculty of Medicine, Vilnius University, Vilnius, Lithuania; 2) Estonian Biocentre, Tartu 51010, Estonia; 3) Institut de Biologia Evolutiva (UPF-CSIC), Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Parc de Recerca Biomèdica de Barcelona, Doctor Aiguader 88, 08003 Barcelona, Catalonia, Spain.

Introduction: A characterization of the adaptive history of human populations requires knowledge of the genes that have been affected by positive natural selection which is also important for an analysis of the genetic causes behind human disease. The present study aims to analyze signatures of adaptive positive selection from high-density genotyping data in a total of 399 individuals from the Lithuanian population. **Material and methods:** We analyzed Illumina 770 K HumanOmniExpress-12 v1.1 and Infinium OmniExpress-24 array data from Lithuania combined with CEU, FIN and YRI populations from 1000 genome data for estimating genetic differentiation between continents. Three statistics were employed for scanning high-density genotyping data to infer signs of positive natural selection: Tajima's D, F_{ST} and cross-population extended haplotype homozygosity (XP-EHH) statistical method. Selected SNPs were annotated using ANNOVAR in GRCh37 (hg19), RefSeqGene, dbSNP147 and CADD version 1.3. **Results:** Based on genetic differentiation, when Lithuanian population is compared with Yoruban, CEU and FIN populations, we identified a region with *TYRP1* gene which is implicated in the lightening of skin pigmentation. In addition *SLC25A27* gene was identified, when comparing Lithuanian and reference populations, which may play role in thermoregulatory heat production and metabolism in brain. Among other genes identified, those with the top XP-EHH values include *PLA2G7* gene, located on chromosome 6, containing nonsynonymous SNP (exon7:c.T593C), which is a candidate gene for developing Asthma disease –the most common chronic disease affecting children and young adults. Other nonsynonymous specific variants for Lithuanian population were identified. **Conclusions:** The selection signatures identified in the Lithuanian population demonstrated positive selection pressure on a set of important genes with potential functions that are involved in many biological processes.

2739W

Archaic deserts: Interpreting human genomic regions depleted of archaic hominin ancestry. A.B. Wolf^{1,2}, J.M. Akey². 1) University of Washington, Seattle, WA; 2) Princeton University, Princeton, NJ.

Recent studies have identified archaic sequence in modern human genomes that is inherited from ancient hybridization with Neanderthals and Denisovans. Strikingly, the distribution of surviving archaic sequence throughout the human genome is heterogeneous. Several large regions of the human genome are significantly depleted of archaic ancestry. In particular, four multi-megabase regions are significantly depleted of both Neanderthal and Denisovan sequence for multiple geographically diverse populations with unique population histories. Understanding the formation and characteristics of archaic deserts will help interpret how archaic admixture influenced human evolution and, possibly, what genes play a role in unique human behaviors. For example, archaic deserts may represent loci where archaic sequence was deleterious and rapidly purged from modern human populations. Alternatively, the stochastic loss of archaic sequence due to drift could contribute to these archaic deserts. To better interpret the mechanisms responsible for archaic desert formation, we have performed an exhaustive set of coalescent simulations under a wide range of neutral demographic models. As well, we have used a range of forward-in-time simulations that incorporate selection on introgressed archaic sequence. Our results show that neutral demographic processes alone are unlikely to explain archaic deserts. Models that incorporate selection acting on deleterious archaic sequence provide a significantly better fit to the empirical data.

2740T

Whole genome sequencing reveals admixture history of Singapore Peranakan Chinese. D. Wu¹, P. Li^{2,3}, B. Pan^{2,3}, J. Dou¹, C. Bellis², J. Liu², R.S.Y. Foo^{2,3}, C. Wang^{1,4}. 1) Computational and Systems Biology, Genome Institute of Singapore, Agency for Science, Technology and Research (A*STAR), Singapore; 2) Human Genetics, Genome Institute of Singapore, Agency for Science, Technology and Research (A*STAR), Singapore; 3) Cardiovascular Research Institute, National University Health System, National University of Singapore, Singapore; 4) Duke-NUS Medical School, National University of Singapore, Singapore.

Peranakan Chinese, also known as Straits-born Chinese, are descendants of early Chinese immigrants settling in the Malay Archipelago between the 15th and 17th centuries. Through the years, this community has developed unique Peranakan culture, which preserves most of their Chinese traditions with strong influence from local Malays. Yet, whether genetic admixture co-occurred with the cultural mixture is debatable even among the Peranakan community. To answer this question, we recruited 177 Singapore Peranakans and performed whole genome sequencing at ~15X. We have by now analyzed 79 Peranakan genomes together with additional whole genome data of Chinese, Malays, and Indians in Singapore, as well as Southern Chinese and Northern Chinese from the 1000 Genomes Project. Principal component analysis and supervised ADMIXTURE analysis suggest that Peranakan Chinese inherit a significant level of Malay ancestry. On average, the estimated Malay ancestry proportion is $\sim 0.114 \pm 0.009$ (mean $\pm$ s.e.m, standard error of the mean) in Peranakan Chinese, $\sim 0.032 \pm 0.004$ in Southern Chinese and $\sim 0.008 \pm 0.003$ in Northern Chinese. GLOBETROTTER analysis dates the admixture event between Chinese and Malays about 6-12 generations ago. In summary, our preliminary results provide strong genetic evidence to support the hypothesis that genetic admixture co-occurred with cultural mixture in the formation of the Peranakan Chinese community.

2741F

Landscape of multi-nucleotide variants across 123,136 human exomes.

Q. Wang^{1,2,3}, E. Pierce-Hoffman^{2,3}, A. Hill⁴, A. O'Donnell-Luria^{2,3}, B. Cummings^{2,3,5}, K. Karczewski^{2,3}, L. Francioli^{2,3}, L. Gauthier^{2,3}, D. MacArthur^{2,3}. 1) Program in Bioinformatics and Integrative Genomics, Harvard Medical School, Boston, MA; 2) The Broad Institute of MIT and Harvard, Cambridge, MA; 3) The Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 4) Genome Sciences, University of Washington, Seattle, WA; 5) Program in Biomedical and Biological Sciences, Harvard Medical School, Boston, MA.

Multi-nucleotide variants (MNV) are defined as clusters of two or more nearby variants existing on the same haplotype. When multiple variants in an MNV are found within the same codon, the overall impact may differ from the functional consequences of any of the individual variants (e.g. missense + missense resulting in protein truncating variant, PTV). Moreover, the mutational mechanisms underlying MNVs remain poorly characterized. To create the largest collection of MNVs to date, we scanned phased high-quality SNP pairs within 3bp across 123,126 exome-sequenced individuals aggregated as part of the Genome Aggregation Database (gnomAD) v2 callset. Three different phasing methods showed very high (>94%) concordance when tested on 13 exome trios, conforming the quality of read-based phasing. Our analysis resulted in discovery of 106,492 MNVs (72,220 MNVs of adjacent SNVs, 32,816 MNVs of 2bp apart, and 1,456 MNVs of 3 consecutive bases). Of those, 377 were gained PTV (both SNPs are not PTV, but the resulting MNV is), 1,809 were rescued PTV (at least one of individual SNP is a PTV, but the resulting MNV is not), and 15,534 additional variants altered the protein sequence of the affected gene product, emphasizing the importance of haplotype-based annotation for the complete detection of functional variants. We found a skewed distribution of MNV frequency per base pattern. There were 7,896 CA->TG mutations, but only 52 TA->GC mutations, consistent with the >100x higher mutation rate for CpG transitions compared to non-CpG transversions. Also, we found evidence of previously known two-base substitution mechanism such as polymerase slippage in repeat regions and replication error by polymerase-zeta. We show that the major mutational origin of MNVs differs substantially by sequence pattern (>92% are two SNP event in a row for CA->TG mutation we discovered, >60% polymerase slippage for AA<->TT, and >87% polymerase zeta-induced replication error for TC->AA and their reverse complements.). Finally, we present the results of ongoing analyses investigating the contribution of MNVs to rare disease by applying the methods to >3,000 sequenced individuals from rare disease families, and their distribution in non-coding sequence by analyzing >15,000 sequenced whole genomes. This project provides a validated approach for the detection and interpretation of MNVs on very large sequence data sets, and demonstrates their importance for functional and clinical variant interpretation.

2742W

Population-specificity of eQTL effect sizes across tissues. S. Kasela^{1,2},

P. Hoffman¹, S. Castel^{1,2}, T. Lappalainen^{1,2}, GTEx Consortium. 1) New York Genome Center, New York, NY; 2) Department of Systems Biology, Columbia University, New York, NY.

Genetic variation modulating gene expression is widespread in the human genome. Previous studies have shown a high level of sharing of regulatory genetic effects across populations, based on eQTL analysis in a subset of blood cell types. Here, we conducted systematic approach to characterize population differences in effect sizes of *cis*-eQTLs in 31 tissues with up to 588 individuals of European and up to 86 individuals of African ancestry from the Genotype-Tissue Expression (GTEx) Project v8 release. We measured eQTL effect size using our previously developed metric, allelic fold change (aFC), which is robust to differences in allele frequency and expression level, and allows replication of the eQTL results using allele-specific expression data. While the vast majority of the top *cis*-eQTLs discovered in GTEx showed similar effect size in both of the populations, up to 0.6% of the top *cis*-eQTLs tested had significantly divergent magnitude of aFC per tissue at FDR 5%, with no obvious tissue enrichment for these effects. We observed that eQTLs with significant differences were likely to be shared across tissues, suggesting that population differences appear mostly shared as well. Controlling for differences in LD, we defined a set of high confidence *cis*-eQTLs with population-specific effects. One such eQTL was rs1815739 where there derived T allele creates a premature stop codon in the actinin alpha 3 (*ACTN3*) gene that is expressed only in fast-twitch muscle fibers responsible for causing force at high velocity. This relatively common variant especially in European populations is well known to trigger nonsense-mediated decay and associate to reduced sprint performance. In skeletal muscle of European Americans the SNP reduced *ACTN3* dosage substantially, as expected, log₂aFC = -5.2, 95% CI -5.7..-4.7. However, in African Americans the effect was nearly absent (log₂aFC = -0.3, 95% CI -3.2..1.5) with allele-specific expression data suggesting dosage compensation. This raises interesting questions of molecular mechanisms as well as potential population-specific phenotypic effects of this variant. Taken together, our results have widened the knowledge of population differences across variety of tissues, confirming the high overall similarity of genetic regulation of gene expression between populations but also highlighting interesting exceptions that shed light to molecular drivers of human diversity.

2743T

Ethnic specific reference genome may salvage "unmapped" reads to the global reference of GRCh38.

J. Kim¹, I.C. Yang², Y.K. Bae², J. Sung³. 1) Interdisciplinary program in Bioinformatics, Seoul National University, Korea, Republic of, MPH; 2) Korea Research Institute of Standards and Science, Center for Bio-Analysis, Daejeon, Korea, PhD; 3) Complex Disease and Genome Epidemiology Branch, Department of Epidemiology, School of Public Health, Seoul National University, Seoul, Korea, PhD, MD.

"Re-sequencing" methods relying on the Reference Human Genome (GRCh38) have enabled the current deluge of cost-efficient and high-throughput NGS analyses. The use of the GRCh38, the most complete human genome assembly with multi-ethnic mosaic, was sometimes asked whether it is truly representative of all ethnic groups, but few studies empirically showed the utility and implications of having a well-established ethnic-specific reference genome. A reference genome for Asians might be a good example given the current GRCh38 still has low contributions from Asians. The AK1 is a *de novo* genome assembly based on a Korean man, which has about 45 Mb of scaffold N50. We used fourteen individuals (5 East Asian, 4 European, and 5 African ancestry) whole genome sequencing (WGS) data aligned to the GRCh38 from the 1000 genome database which are Illumina PCR-free and with high coverage. We extracted unmapped reads from the data and aligned these unmapped reads to AK1 assembly as a reference genome. In average 95.6% of the total reads (773 million out of 809 million) NGS reads were mapped to the GRCh38 with alternative sequence leaving 4.4% unmapped. Among the unmapped (by GRCh38) reads, average 0.90% for 3 ethnic group was newly aligned to AK1. Specially, alignment rate for East Asian was 0.95% which was higher than other ethnic groups (European; 0.86% and African; 0.84%). When we divided the 2832 scaffolds of AK1 into three groups: group 1) almost uniquely mapped to single chromosome of GRCh38 (>90% to one chromosome); group 2) mapped to multiple chromosome; group 3) scaffolds not recognized by the chromosomes of GRCh38, unmapped reads were aligned to the group 3 scaffolds of AK1 (8.8% of total unmapped reads), but also to group 1) and 2). When we examined some newly mapped reads with high depth by introducing AK1 as a reference genome, (only scaffolds which are linked to chromosomes of GRCh38), some new blocks were identified, mostly flanking repeated sequences such as AluSx, MIRc, or Tigger1. It is noteworthy that AK1, a high-quality genome assembly of a Korean, revealed new genomic information and sequences for not only Asians but also European and African ancestry genomes.

2744F

L1 retrotransposition machinery is maintained via the live-and-die model of evolution in primate genomes. *P. Liang, W. Tang, J. Nanayakkara.* Brock University, St. Catharines, Ontario, Canada.

Mobile elements are genetic elements capable of moving or copying/inserting into new locations in the genome, and they make up approximately half of primate genomes. Although mobile elements used to be viewed as genome parasites, they are now known to play important roles in genome evolution and gene function. LINE1 (L1) sequences are mobile elements that are capable of encoding the molecular machinery for retrotransposition, the mobilization of DNA sequences through an RNA intermediate. Most other active transposons in primate genomes rely on L1s to provide functional proteins for their retrotransposition. However, an evolutionary model of intact L1 evolution and retention in the genomes has not been proposed. Using a bioinformatics comparative genomics approach, we performed an analysis to determine 1) the biochemical requirements of intact L1s, 2) the mode of evolution of intact L1s, and 3) *cis*-preference of intact L1s on a genome-wide scale. Intact L1s were found to require a full length ORF1, ORF2, and promoter to maintain retrotransposition activity in the genome. High rates of human-specificity and polymorphism were found among intact L1s compared to non-intact L1s. Case studies of L1 clusters revealed that current intact L1 copies came from L1s which may be currently non-intact. Based on the analysis, we established that intact L1s evolve through a "Live-and-Die" model. Under this model, L1s that are capable of encoding functional proteins may produce many new copies of themselves while "alive" in a *cis*-preference fashion. Over evolutionary time, the original parental L1s acquire random, inactivating mutations to lose transcription and/or coding capacity and "die". Thus, it leaves the new generation of intact L1s to take over the responsibility to produce transposition molecular machinery for the entire genome. Interestingly, the human genome has a much higher number of intact L1s (more than 10 times) than all other primates examined, including its close relative, chimpanzee. This suggests that humans, of all primates, have the greatest potential to further evolve via retrotransposition. Moreover, most intact L1s in humans are human-specific, among which a significant portion also represent polymorphic insertions, further demonstrating the youth and plasticity of the human genome.

2745W

A complete characterization of the *HLA-A* haplotype structure in a

Brazilian population sample. T.H.A. Lima^{1,2}, I.O.P. Porto^{1,3}, M.A. Paz^{1,3}, L.C. Veiga-Castelli⁴, M.L.G. Oliveira⁴, E.A. Donadi⁵, D. Meyer⁶, A. Sabbagh⁷, C.T. Mendes-Junior⁸, E.C. Castelli^{1,2,3}. 1) São Paulo State University (UNESP), Molecular Genetics and Bioinformatics Laboratory - Experimental Research Unity, School of Medicine, Botucatu, State of São Paulo, Brazil; 2) São Paulo State University (UNESP), Genetics Program, Institute of Biosciences of Botucatu, Botucatu, State of São Paulo, Brazil; 3) São Paulo State University (UNESP), Pathology Program, School of Medicine, Botucatu, State of São Paulo, Brazil; 4) Department of Genetics, School of Medicine of Ribeirão Preto, University of São Paulo (USP), Ribeirão Preto, State of São Paulo, Brazil; 5) Department of Medicine, School of Medicine of Ribeirão Preto, University of São Paulo (USP), Ribeirão Preto, State of São Paulo, Brazil; 6) University of São Paulo, Department of Genetics and Evolutionary Biology, São Paulo, Brazil; 7) UMR 216 MERIT, IRD, Université Paris Descartes, Faculté de Pharmacie, Sorbonne Paris Cité, Paris, France; 8) Departamento de Química, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, São Paulo, Brasil.

The Human Leucocyte Antigen (HLA) complex comprises some of the most polymorphic human genes, including *HLA-A*. *HLA-A* is a classical HLA class I gene that encodes a key molecule for antigen presentation and recognition of self and non-self antigens. It is also associated with NK cell modulation through the interaction with NK activating and inhibitory receptors. *HLA-A* variability in exons is frequently explored in many worldwide populations, but minor attention has been given to regulatory segments and introns, as well as to other exons such as the ones encoding the $\alpha 3$, transmembrane and cytoplasmic domains. Variability in these neglected segments might influence the *HLA-A* expression profile, signaling, cytotoxicity, anchorage, and stability of the molecule on the plasmatic membrane. Here we present a molecular and computational strategy to evaluate the complete *HLA-A* variability including its regulatory segments (promoter and 3'UTR) by using second-generation sequencing, surveyed on 408 individuals from the Brazilian Southeast. We detected 453 variable sites considering the entire *HLA-A* segment, arranged into 94 different haplotypes. The *HLA-A* locus was highly polymorphic and presented a high Linkage Disequilibrium (LD) pattern along the entire segment, with many polymorphisms in complete LD. The promoter region presented 100 variable sites and 25 haplotypes. This low number of haplotypes could be explained as a consequence of the LD pattern observed and the fact that many of these variable sites were in complete LD. There was a direct correlation among the promoter, coding and 3'UTR haplotypes here observed, thus few variable sites are necessary to capture which promoter/coding/3'UTR haplotype groups are present in a sample. We observed a high nucleotide diversity throughout the gene, exception made for the 3'UTR region. The *HLA-A* promoter segment seems to be as polymorphic as the coding region, but with a conserved 'core promoter'. Positive Tajima's D was detected for all segments demonstrating a balancing selection profile across the entire *HLA-A* gene segment. However, exon 4 presented evidence of purifying selection, and this exon encodes a domain that interacts with the CD8 receptor. It is possible that each promoter here detected (and their associated coding alleles) is related to a distinct expression profile. Functional evaluations are still required to fully understand the biology and expression pattern of this gene.

2746T

Black South African Bantu-speakers^(*) show evidence of significant population structure with implications for biomedical research: H3Africa SNP array data. M. Ramsay^{1,2}, D. Sengupta¹, A. Choudhury¹, S. Aron¹, S. Tollman³, S. Norris⁴, M. Alberts⁵, S. Hazelhurst^{1,6}, members of AWI-Gen and the H3Africa Consortium. 1) Sydney Brenner Institute for Molecular Bioscience, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg; 2) Division of Human Genetics, National Health Laboratory Service and School of Pathology, Faculty of Health Sciences University of the Witwatersrand, Johannesburg, South Africa; 3) MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt), School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg 2193, South Africa; 4) MRC/Wits Developmental Pathways for Health Research Unit, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa; 5) Department of Pathology and Medical Science, School of Health Care Sciences, Faculty of Health Sciences, University of Limpopo, Polokwane, South Africa; 6) School of Electrical & Information Engineering, University of the Witwatersrand, South Africa.

Genetic diversity among extant black South African Bantu-speakers (BS) is the result of multiple waves of migration during the Bantu-expansion over the past ~2000 years and complex gene flow from resident hunter-gatherer and agro-pastoralist populations. To date, disease mapping studies have essentially treated BS as a monolithic group and the impact of population structure on discovering robust genetic associations among the BS has remained largely unexplored. This is the first large-scale study on the genetic diversity of black first language Bantu-speakers in South Africa, including ~4000 participants representing over 9 major ethnolinguistic sub-groups recruited in three provinces (Soweto – urban Gauteng; Dikgale – rural Limpopo; Agincourt – rural Mpumalanga). They were genotyped on the newly designed H3Africa 2.5M array. Principal component analysis (PCA) showed population structure among the BS, with the three more northern groups, Tsonga, Pedi and Tswana showing a clear east to west cline. The Tsonga, Pedi, Zulu and Xhosa also showed a north to south cline. The distribution of these groups on PCA mirrored the most recent geographic settlement defined according to ethnolinguistic clustering and was consistent with the history of Bantu-migration. Admixture analyses detected the level of Khoesan (KS) ancestry to be a major differentiating factor for these groups, varying from minimal in Tsonga (~2%), intermediate in Pedi (11%), to high among Tswana (~22%). The level of maternal KS ancestry (L0 mtDNA haplogroup) demonstrated a similar trend from lowest in Tsonga to highest in Tswana. ALDER and MALDER were used to date the timing of gene flow from the KS populations, providing an estimate of the timeline for the arrival of various BS populations in South Africa. We report a deep relatedness within the two rural South African communities (more than 30% related at a PIHAT cutoff of 0.05) likely reflecting geographic and cultural isolation that would require consideration in GWASs. This study provides new insights into the origins and population structure of black South African Bantu-speakers. ^(*) Bantu-speakers are a substantive component of the Niger-Congo ethnolinguistic group, one of the four major ethnolinguistic groupings in Africa (Nilo-Saharan; Afro-Asiatic; Niger-Congo; Khoesan).

2747F

Initiatives to increase the diversity of populations represented in genetics research. A.J. Shastri, M.E. Moreno, N. Litterman, G.D. Poznik, A. Auton, J.L. Mountain, the 23andMe Research Team. 23andMe, Inc., CA.

Studying diverse populations is critical not only for maximizing new discoveries but also for ensuring that the results of research benefit all. Throughout its history, 23andMe has worked to increase the diversity of its genetic data. In 2011 we launched the Roots into the Future initiative, which enrolled over 10,000 African Americans in genetics research through access to 23andMe's service. In 2016, we launched the African Genetics Project, a study of individuals with recent African ancestry, which aimed to improve our ability to provide more detailed ancestry results and to contribute to the global understanding of how people have migrated within Africa and from Africa to the rest of the world. Contributions of over 700 research participants enrolled in the study have helped to provide finer resolution within our Ancestry Composition report for more than half a million customers. With NIH funding, we are sequencing more than 2,000 African American customers who consented to participate in research. Through this project, we will broaden the 23andMe sequencing panel for imputation. Participants have also agreed to make their de-identified data available through an NIH database, dbGAP, to qualified researchers conducting health-related studies. In 2017, we also launched the Global Genetics Project, which provides kits at no cost to participants with all four grandparents born in a country underrepresented in our reference data panel. The project has enrolled over 3,000 participants in just 5 months. **In our newest effort, the Populations Collaborations Program, we partner with researchers working to genotype people in communities across Africa, Asia, the Americas and beyond. We provide financial and scientific support for qualified researchers and use the data to simultaneously expand our Reference Data Panel and support researchers who are working with understudied populations, thereby expanding the breadth of genomic studies. In evaluating projects for this program, we consider scientific merit as well as how the research benefits the community being studied. We give preference to studies that engage local researchers and community members, and we provide support for community investments. Through prior collaborations of this nature, we have generated data on populations from Democratic Republic of Congo, Sierra Leone, Tanzania, and Southern Africa, which have enabled us to provide more detailed African ancestry.**

2748W

HLA-C diversity and natural selection profile surveyed in a Brazilian population sample. A.S. Souza^{1,2}, M.A. Paz^{1,3}, L.C. Veiga-Castelli⁴, I.O.P. Porto^{1,3}, E.A. Donadi⁵, A. Sabbagh⁶, C.T. Mendes-Junior⁷, E.C. Castelli^{1,2,3}. 1) São Paulo State University (UNESP), Molecular Genetics and Bioinformatics Laboratory - Experimental Research Unity, School of Medicine, Botucatu, State of São Paulo, Brazil; 2) São Paulo State University (UNESP), Genetics Program, Institute of Biosciences of Botucatu, Botucatu, State of São Paulo, Brazil; 3) São Paulo State University (UNESP), Pathology Program, School of Medicine, Botucatu, State of São Paulo, Brazil; 4) Department of Genetics, School of Medicine of Ribeirão Preto, University of São Paulo (USP), Ribeirão Preto, State of São Paulo, Brazil; 5) Department of Medicine, School of Medicine of Ribeirão Preto, University of São Paulo (USP), Ribeirão Preto, State of São Paulo, Brazil; 6) UMR 216 MERIT, IRD, Université Paris Descartes, Faculté de Pharmacie, Sorbonne Paris Cité, Paris, France; 7) Departamento de Química, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, São Paulo, Brasil.

HLA-C is a highly polymorphic Human Leukocyte Antigen class I gene associated with antigen presentation and NK cell modulation. *HLA-C* is the only classical *HLA* gene expressed on the maternal-fetal interface, where it plays an important immunomodulatory role. *HLA-C* variability at exons is frequently explored in worldwide populations, but minor attention has been given to regulatory and intron segments, as well as to other exons such as the one encoding the transmembrane and cytoplasmic domains. Variability at these neglected segments might influence the *HLA-C* expression profile and stability of the molecule. Here we present the variability and haplotype structure of the entire *HLA-C* locus, including introns and regulatory segments, surveyed in 400 individuals from the Brazilian Southeast. Variability was assessed using second-generation sequencing and in-house software to map reads taking into account *HLA* genes variability and structure. *HLA-C* presented 363 variable sites with a strong linkage disequilibrium (LD) among them, configuring 92 different extended haplotypes. The promoter segment (1,525bp upstream the first translated ATG) presented 82 variable sites and 29 haplotypes, the coding region (including introns) presented 245 variants and 83 haplotypes, and the 3'UTR segment presented 36 variants and 20 haplotypes. Each coding allele group was associated with the same promoter and/or 3'UTR haplotype groups. Nucleotide diversity was high throughout the gene, but the promoter and the 3'UTR initial segment (region of conserved targets for miRNA) presented the lowest indices detected. *HLA-C* presented signatures of balancing selection throughout the entire locus, exception made for the promoter segment and exon 1. Additionally, exon 1, which encodes the leader peptide, that is not part of the mature protein but an HLA-E ligand, presented evidence of positive selection. Together, these results indicate that *HLA-C* presents evidence of balancing selection along the entire locus, but it also presents conserved segments mainly related to expression regulation. Moreover, because of the high LD detected, few variable sites are necessary to capture the promoter, coding and 3'UTR haplotypes. Functional studies are fundamental to better understand the relationship between the *HLA-C* variability pattern and its function and expression profile, and also whether different promoters are related to a differential expression profile.

2749T

Genetic variants in *ST6GAL1* gene are associated with thyroglobulin plasma level in healthy individuals. *T. Zemunik¹, A. Matana¹, M. Popović¹, T. Boutin², V. Torlak³, D. Brdar³, I. Gunjača¹, I. Kolčić⁴, V. Boraska Perica¹, A. Punda³, O. Polašek⁴, M. Barbačić¹, C. Hayward².* 1) Department of Medical Biology, University of Split, School of Medicine, Split, Croatia; 2) MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, United Kingdom; 3) Department of Nuclear Medicine, University Hospital Split, Spinčičeva 1, Split, Croatia; 4) Department of Public Health, University of Split, School of Medicine, Šoltanska 2, Split, Croatia.

Background: Thyroglobulin (Tg) is the largest and most abundant protein in the thyroid gland that serves as a precursor for thyroid hormone synthesis. A twin study showed that there is a strong genetic component to the variability of serum Tg; however, no genome-wide association study (GWAS) of serum/plasma Tg levels has been reported so far. The aim of this study was to identify genetic variants associated with plasma Tg levels among healthy individuals. **Methods:** We performed GWAS meta-analysis within two Croatian populations, CROATIA_Split and CROATIA_Korcula. The analyses included 1094 healthy individuals. The selection of healthy individuals was based on detailed thyroid function assessment (including measurements of Tg, TSH, fT3, fT4, TgAb, and TPOAb) and anamnestic data. A total of 7 597 379 variants, imputed using the 1000 Genomes reference panel, were analysed for the association. GWAS was performed under an additive model, controlling for age, gender, and relatedness within each data set. Meta-analysis was conducted using the inverse-variance fixed-effects method. **Results:** Sixteen polymorphisms located on chromosome 3 within the *ST6GAL1* gene reached genome-wide significance. The most significant SNP was rs4012172 ($P=1.29 \times 10^{-10}$), which explained 3.19% of the variance in Tg levels. *ST6GAL1* belongs to the sialyltransferase proteins family which has a fundamental role in the synthesis of specific sialylated structures on various glycoproteins, including thyroglobulin. *ST6GAL1* could have a role in the regulation of thyroglobulin level through receptor-mediated signalling transduction, helping to recycle immature Tg from thyrocytes into the follicle lumen, subsequently causing the alteration of Tg release in the circulation. **Conclusions:** In this first reported GWA meta-analysis of Tg levels, we identified a highly biologically plausible locus that could have a role in the regulation of plasma Tg levels in healthy individuals.

2750F

Systematic comparative genomic analysis reveals differential DNA transposition as a significant mechanism in primate evolution. *W. Tang, P. Liang.* Department of Biological Sciences, Brock University, St. Catharines, Ontario, Canada.

Mobile elements generated via DNA transposition represent an abundant type of DNA elements in primate genomes, ranging from ~47% in the green monkey genome to ~52% in the human genome. As a result of past and ongoing activity, DNA transposition is known to be responsible for generating inter-species and intra-species genomic variations, and it plays important roles in shaping genome evolution and in impacting gene function. While limited analysis of mobile elements has been performed in a few primate genomes, a large-scale systematic comparative genomic analysis is still unavailable. Using a bioinformatics comparative genomics approach, we performed an analysis to identify species-specific mobile elements (SS-MEs) in 8 primate genomes, which include human, chimpanzee, gorilla, orangutan, green monkey, crab-eating macaque, rhesus monkey, and baboon. These genomes have good representations for ape and monkey group of the primates for which genome sequences are available. Our analysis identified a total of 230,855 SS-MEs from the 8 primate genomes, which collectively contribute to ~82 Mbp genome sequences. Several new interesting observations were made based on these SS-MEs. First, the DNA transposition activity level was shown to be drastically different across species with the highest (baboon genome) being more than 30 times higher than the lowest (crab-eating macaque genome) by the number of SS-MEs. Second, the compositions of SS-MEs differ also significantly across genomes. By the copy numbers of SS-MEs divided into major ME classes, SINE represents the dominant class in all genomes, but more so in the monkey genomes than the ape genomes with orangutan genome being the outlier of this trend by having LINE as the dominant class. While AluY forms the major SINE classes in the ape genomes, AluYRa1 is the dominant SINE in the monkey genomes. For LINES, each genome seems to have different dominant L1 subfamilies. Third, DNA transposons, despite being considered to be dead in primate genomes, were in fact shown to have a certain level of activity in all genomes examined with a total of ~2,400 entries as SS-MEs. Representing the first systematic comparative analysis of mobile elements in primate genomes, our study has shown that mobile elements evolved quite differently among different groups and species of primates, indicating that differential DNA transposition has served as an important mechanism during primate evolution.

2751W

Whole-genome based linkage disequilibrium maps in sub-Saharan African population: Functional clustering and relationship to disease. *R. Jabal-Ameli Forooshani, W. J. Tapper, S. Ennis, A. Collins.* Genetic epidemiology & Genomic Informatics, University of Southampton, Southampton, United Kingdom.

The extent and strength of linkage disequilibrium (LD) reflect the historical impact of recombination, selection and population history. The LD landscape in human population is known to comprise highly punctate hotspots that divide extended blocks of low haplotype diversity from narrow regions of LD breakdown. LD maps based on whole-genome sequence (WGS) data enables complete resolution of LD structure across human chromosomes. Increased marker density from WGS-derived data provides superior power for delineation of regions of the genome that are differentially recombined. We posited that recombination rate and haplotype diversity can be considered as a proxy for gene essentiality and establish the interesting relationship between gene function and LD strength. The weaker LD in sub-Saharan African population achieves unprecedented resolution for functional analysis LD maps in the context of human disorders. We constructed LD maps from WGS data across six sub-Saharan African populations according to the Malécot-Morton model of allelic association. The LD maps are constructed on the scale of linkage disequilibrium units (LDU) in which the decline of LD to 'background' levels across a variable physical distance corresponds to one LDU. The LDU scale enables direct comparison of the extent of LD in different genomic regions independent of the physical scale. The gene-specific LDU size across 17,927 genes was corrected for gene length and converted to a gene-specific Z-scores. Functional gene clustering analysis across quartiles was carried out in PANTHER (v.13.1) and taxonomic comparison of gene clusters was carried out using HomoloGene database (release 68). Our results demonstrate that the magnitude of LDU size across different populations directly reflects the paleogeographic characteristic of the respective population. We identified that gene clusters with intense LD are enriched with genes involved in developmentally important pathways such as DNA damage response or regulators of cell-cycle checkpoints. We also identified that gene clusters with extensive LD are generally evolved earlier during evolution. In contrast genes with low LD appear to be enriched with gene clusters that evolved more recently. In conclusion, fully saturated LD maps generated from whole-genome data in SSA populations provides a framework for accurate delineation of signatures of selective sweeps in the human genome.

2752T

A map of constrained coding regions in the human genome. *A. Quinlan, J. Havrilla, R. Layer, B. Pedersen.* Department of Human Genetics, University of Utah, Salt Lake City, UT.

Deep catalogs of genetic variation collected from many thousands of humans enable the detection of intraspecies constraint by revealing coding regions with a scarcity of variation. While existing techniques summarize constraint for entire genes, single metrics cannot capture the fine-scale variability in constraint within each protein-coding gene. To provide greater resolution, we have created a detailed map of constrained coding regions (CCRs) in the human genome by leveraging coding variation observed among 123,136 humans from the Genome Aggregation Database (gnomAD). The most constrained coding regions in our map are enriched for both pathogenic variants in ClinVar and de novo mutations underlying developmental disorders. CCRs also reveal protein domain families under high constraint, suggest unannotated or incomplete protein domains, and facilitate the prioritization of previously unseen rare variation and de novo mutation in studies of disease. Finally, we hypothesize that CCRs with the highest constraint exist within genes that are likely to cause yet unobserved human phenotypes owing to strong purifying selection. As such, they have the promise of revealing new genic regions, that when mutated, lead to severe developmental phenotypes. .

2753F

Bayesian factor analysis for inference of population structure and polygenic selection. *A. Kousathanas, M. Robinson.* Complex Trait Genetics Group, Department of Computational Biology, University of Lausanne, Switzerland.

Human populations display substantial within and between-population phenotypic variation. It is unknown to what extent between-population phenotypic differences have been driven by adaptive evolution to each population's local environment. Understanding local adaptation is crucial as it may underlie population differences in disease risk, especially when individuals move to new environments during their life that are different from their ancestor's. The highly polygenic architecture of most quantitative traits and common diseases, suggests that adaptive evolution should be driven by the cumulative effect of selection over many loci. Existing methods to detect selection on traits that underlie local adaptation take advantage of this expectation by using the cumulative selection signal across several loci associated with a particular trait. However, these methods typically only use loci that are strongly associated with a particular trait, suffer from ascertainment biases that may be extreme, require defined populations which is particularly difficult in cases of heavy admixture, and provide limited inference as to the genomic regions involved in polygenic adaptation. Here, we address these issues, by developing an approach based on Bayesian factor analysis, that is a generalization of the widely used principal component analysis (PCA). Our method jointly infers population structure, the effect of each locus on population differentiation (z), the effect of each locus on one or multiple traits (b), and the covariance between locus effects and genetic differentiation ($\text{cov}(z,b)$). We present theory and extensive simulation work to demonstrate how the estimation of $\text{cov}(z,b)$ provides unique inference of polygenic selection. The signal detected can be decomposed to reveal the populations that experienced local adaptation, the direction of the phenotypic differentiation, and the genomic regions involved. We analyse a range of human genotype-phenotype datasets, testing for population genetic differentiation of height, body mass index, and common complex disorders of cardiovascular disease and type-2 diabetes. Overall, our method not only unifies population genetic structure and genetic association analyses into a single framework, but provides a comprehensive way to infer the signal of polygenic selection.

2754W

A standardized statistic improves power to detect long-term balancing selection. *K.M. Siewert¹, B.F. Voight^{2,3,4}.* 1) Genomics and Computational Biology Graduate Group, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 2) Department of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 3) Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 4) Institute for Translational Medicine and Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

Balancing selection is an evolutionary force which can maintain genetic variation in a population over long evolutionary time periods. Detection of genomic targets of this type of selection can reveal selective pressures and phenotypically interesting haplotypes in the genome. Until recently, methods to detect this type of selection suffered from a lack of power or were computationally intensive. We recently reported a method (Beta) to detect balanced haplotypes without the need for genomic data from an outgroup species. This method uses a difference of two estimators of the mutation rate, one which is sensitive to balancing selection and one which corrects for the background mutation rate. Here, we present new work which incorporates between-species substitution data into a novel mutation rate estimator integrated into the Beta statistic framework. We show that this new approach outperforms existing summary statistic methods in simulations under a broad array of parameters. In addition, we derive formulas for the variance of both versions of Beta -with and without substitution data- allowing for the calculation of a z-score. This facilitates appropriate interpretation of the evidence of balancing selection across samples, populations, and allele frequencies, variables that hampered previously described methods. As a result, the Beta statistic is a well-powered approach to detect balancing selection in a wide variety of species.

2755T

Population pharmacogenomics for precision public health in Colombia.

E.T. Norris^{1,2,3}, M. Robledo-Moreno⁴, S.D. Nagar^{1,2,3}, L. Rishishwar^{2,3}, K.L. O'Neal¹, S. Vélez⁴, I. Torres⁴, M.A. Medina-Rivas⁵, A. Valderrama-Aguirre⁶, I.K. Jordan^{1,2,3}, J.E. Gallo⁴. 1) School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA, USA; 2) IHRC-Georgia Tech Applied Bioinformatics Laboratory, Atlanta, GA, USA; 3) PanAmerican Bioinformatics Institute, Cali, Valle del Cauca, Colombia; 4) GenomaCES, Universidad CES, Medellín, Antioquia, Colombia; 5) Centro de Investigación en Biodiversidad y Hábitat, Universidad Tecnológica del Chocó, Quibdó, Chocó, Colombia; 6) Biomedical Research Institute, Faculty of Health, Universidad Libre-Seccional Cali, Cali, Valle del Cauca, Colombia.

While genomic approaches to precision medicine hold great promise, they remain prohibitively expensive for developing countries. The precision public health paradigm, whereby healthcare decisions are made at the level of populations as opposed to individuals, provides one way for the genomics revolution to directly impact health outcomes in the developing world. Genomic approaches to precision public health require a deep understanding of local population genomics, which is still missing for many developing countries. We are investigating the population genomics of pharmacogenetic variants in an effort to inform healthcare decisions in Colombia. Our work focuses on two neighboring populations with distinct ancestry profiles: Antioquia (European) and Chocó (African). We discovered a number of pharmacologically relevant variants that show anomalous effect allele frequencies within and between the two populations and found these differences to be associated with their distinct genetic ancestry. For example, the C allele of the SNP rs4149056, which is associated with an increased risk of toxicity to a commonly prescribed statin, is found at relatively high frequency in Antioquia and associated with European ancestry. In addition to alleles related to increased toxicity risk, we also have evidence that alleles related to dosage and metabolism have anomalous frequencies between the two populations and are associated with specific ancestries. Using these findings, we have developed and validated an inexpensive allele-specific PCR assay to test for the presence of such population-enriched pharmacogenetic SNPs in resource-limited settings like Colombia.

2756F

A new paradigm for pharmacogenomic discoveries: Capturing drug response during surgery.

S. Wenric¹, J.M. Jeff¹, T. Joseph², K. Slivinski¹, M-C. Yee³, G.M. Belbin¹, A.O. Obeng^{1,4,5}, S.B. Ellis¹, E.P. Bottinger¹, O. Gottesman¹, M.A. Levin^{2,6}, E.E. Kenny^{1,7,8,9}. 1) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai Hospital, New York, NY; 2) Department of Anesthesiology, Icahn School of Medicine at Mount Sinai Hospital, New York, NY; 3) Carnegie Institution for Science, Dept. of Plant Biology, Stanford, CA; 4) Division of General Internal Medicine, Icahn School of Medicine at Mount Sinai Hospital, New York, NY; 5) Department of Pharmacy, Icahn School of Medicine at Mount Sinai Hospital, New York, NY; 6) Division of Cardiothoracic Anesthesia, Icahn School of Medicine at Mount Sinai Hospital, New York, NY; 7) Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai Hospital, New York, NY; 8) Department of Genetics and Genomics, Icahn School of Medicine at Mount Sinai Hospital, New York, NY; 9) Center for Statistical Genetics, Icahn School of Medicine at Mount Sinai Hospital, New York, NY.

We present a novel pharmacogenomic approach in which we define **real-time drug response to intravenous medication administered during surgery**. We have built a database of surgical procedures that capture drug administration events and succeeding physiological responses, as well as type and length of surgery, anesthetic technique, quantity of fluids administered and other patient information. These data are linked to an internal Biobank of patients from New York, with genotype, sequence, and rich phenotypic data from the electronic medical record. Leveraging this resource, we investigated real-time response to phenylephrine, a selective α 1-adrenergic receptor agonist that is used to treat hypotension during surgery. We found that administration of a bolus of phenylephrine (50-200 μ g) increases MAP an average of **17.03 mmHg (SE $\pm$ 0.44), 14.8 mmHg(SE $\pm$ 0.48), and 14.89 mmHg(SE $\pm$ 0.41), in self-reported European-Americans (EA;n=1389), African-Americans (AA;n=1221) and Hispanic/Latino (HL;n=1714), respectively. Overall EAs have a greater response to phenylephrine compared to HLs and AAs (Mann-Whitney-Wilcoxon p-value < 0.0003 and 0.0004 respectively). We performed a GWAS to test whether genetic variation influences Δ MAP after phenylephrine drug bolus. We confirmed associations in biologically relevant α 1-adrenergic receptors, *ADRA1A* (p<2*10⁻²) and *ADRA1B* (p<4*10⁻³) and overall association enrichment in *ADRA1A*. Notably, **we discovered a suggestive GWAS signal (p=5.16*10⁻⁹) in ERICH1 gene**, which codes for glutamate rich protein 1. This work demonstrates our ability to define real-time drug response, detect population differences in phenylephrine response, large effect alleles and novel genes affecting pharmaceutical response, and identify a subset of phenylephrine 'non-responders' with implications for personalized treatment during surgery.**

2757W

Genetic substructure in the State of Jammu and Kashmir, India: Insights from two populations based on autosomal, Y-chromosomal and mitochondrial DNA analyses. *M. Singh¹, A. Sarkar¹, D. Kumar², M.R. Nandineni¹.*

1) Centre for DNA Fingerprinting and Diagnostics, Hyderabad, India; 2) Central Forensic Science Laboratory, Kolkata, India.

Indian sub-continent is bestowed with cultural, linguistic and genetic diversity. Jammu and Kashmir (J&K) is the northern most frontier of the country and its strategic location at the crossroads of China and Tibet in the north-east and Afghanistan and Pakistan in the north-west provides a great opportunity for population geneticists. Here we present the analysis of two populations of J&K viz., Gujjars (GJ) and Ladakh populations (LL). GJ are nomadic/semi-nomadic tribes practicing high level of endogamy, whereas LL are from high altitude region (>3,000 m above sea level (masl)), though one of the most difficult places for inhabitation but was once considered as 'Silk-Road' that played crucial roles in trade and exchange of culture. Both GJ and LL were analyzed employing markers on autosomes (single nucleotide polymorphisms (SNPs) and short tandem repeats (STRs)), Y-chromosome (Y-STRs) and mitochondrial DNA sequences (control region). Pairwise F_{ST} values and cluster analysis based on 275 SNPs shortlisted previously showed that GJ behaved differently than the rest of the populations. Principle co-ordinate and cluster analyses based on 22 autosomal STRs highlighted that GJ and LL were not only genetically distant to each other but they were also shown to be distant to other populations studied. These observations were further substantiated by 23 Y-STRs which also highlighted the unique nature of both the populations. On comparison with other populations of the world, interestingly we observed genetic proximity of GJ with Pashtuns, Pathans and Sindhis. LL populations on other hand showed genetic affiliations with populations from Nepal, Tibet and Uighur-China. GJ were represented by Y-haplogroup; R1a, H and L whereas LL comprised of R1a, L, Q, J1, R1b and I2a1 Y-haplogroups. A total of 19 and 42 unique haplogroups based on mtDNA analysis were assigned to GJ and LL respectively and revealed much lower genetic diversity for GJ than LL, which indicated a greater heterogeneity in LL as compared to GJ. Higher ratio of patrilineal to matrilineal F_{ST} was observed in both GJ and LL which is indicative of patrilocal practices in these groups. Overall this study provides a comprehensive picture of GJ and LL which would connect the dots to trace the ancient trails of human migration in this part of the globe. .

2758T

Examining highly differentiated genotypes in the Tangier Island genetic study. *T.M. Brunetti¹, M. Daya¹, A. Shetty¹, N. Rafaels¹, M. Boorgula¹, S. Chavan¹, M. Campbell¹, C.R. Gignoux¹, T.H. Beaty², K.C. Barnes¹, R.A. Mathias².* 1) Department of Medicine, University of Colorado, Aurora, CO, USA; 2) Department of Epidemiology, Bloomberg School of Public Health, JHU, Baltimore, MD, USA.

RATIONALE: Population isolate communities continue to exist in the United States. During 1993-2001, we recruited ~60% of the 650 individuals living on Tangier Island, Virginia, into a genetics of asthma and associated traits study. Tangier Islanders have retained Cornish speech attributes, reflecting their British heritage and relative isolation. The island can be reached by 45-minute boat ride or by small aircraft, but limited occupational opportunities minimized immigration to the island in the 20th century. We used genealogical data to reconstruct the extended pedigree of 3,512 individuals to the settlement of the community founded in 1722. In this study, we aimed to identify disease relevant alleles that are common in the Tangier Island population, but rare in comparable populations of European Americans. **METHODS:** Samples were genotyped on the Illumina's MEGA. After quality control, 404 samples were available, after which PRIMUS was used to identify a maximal set of 88 independent individuals for further analysis to mitigate inherent bias within a highly related population. HLA genotypes were estimated using HIBAG and its multi-ethnic reference panel, and genotypes with posterior probability > 0.5 were assigned to subjects. HLA allele frequencies of European Americans were downloaded from allelefreqencies.net and compared with the Tangier Island frequencies. Logistic regression (PLINK) was used to identify SNPs across the genome with relatively large frequencies in Tangier Island compared to 90 Thousand Genomes Project (TGP) individuals from Great Britain (GBR). ClinVar annotations were performed using ANNOVAR. **RESULTS:** We identified multiple HLA alleles that are rare in European Americans but quite common on Tangier Island: DBP1*02:02 (0.7% vs. 10.8%), DBP1*05:01 (1.8% vs. 8.5%), DRB1*01:02 (1.3% vs. 20.45%), DRB1*01:03 (1.2% vs. 10.8%), DRB1*04:02 (1.1% vs. 8.5%), DQA1*01:04 (3.2% vs. 18.2%), B*14:01 (1.0% vs. 22.2%) and A*02:05 (0.9% vs. 20.5%). Additionally, 23 SNPs were identified in ClinVar as likely pathogenic and one SNP located on chromosome 11 at position 16691634 was found to be more common in Tangier Island compared to TGP GBR with minor allele frequencies of 0.1477 and 0.05, respectively (Bonferroni corrected p-value < 0.05). **CONCLUSION:** Tangier Island is a genetically isolated founder population, and we speculate such alleles rose to higher frequency due to founder effects and subsequent isolation, which we aim to quantify in future work.

2759F

Detecting fine scale population structure, migration and recent population expansion in the Netherlands. *R.P. Byrne¹, W. van Rheenen², L.H. van den Berg², J.H. Veldink², R.L. McLaughlin¹.* 1) Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Leinster, Ireland; 2) Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, the Netherlands.

We present a detailed genetic study of the population structure and history of the Netherlands, exploiting both genome wide SNP data (N=1626) and geographic information (N=1422), which together allow us to explore the interplay between genetics and local geography. Our study exploits recently developed methods to: i.) Explore fine-scale population structure within the country; ii.) Model local patterns of migration; iii.) Estimate changes in effective population size in recent years. To explore fine scale population structure we applied the ChromoPainter/fineSTRUCTURE pipeline, which defines genetic clusters based on patterns haplotype sharing across samples. We observed that our genetic clusters correlate closely with geography, and demonstrate regional structure at a level not previously seen, revealing subtle differences both between and within the provinces of the Netherlands. These genetic clusters are robust and demonstrate high genetic differentiation (Mean pairwise F_{st} : 0.000458), with some pairs showing pairwise F_{st} as high as 0.001. This is comparable to the level of differentiation observed between some European countries. At the coarsest level of clustering we identified a clear north/south split, roughly separated by the Rhine. To investigate whether this split reflects a barrier to gene flow, we estimated regional rates of migration across the country using the recently developed "Estimating Effective Migration Surfaces" (EEMS) method. EEMS models local migration rates based on the relationship between genetic and geographic distance. Our analysis revealed a strong migrational cold spot spanning the center of the Netherlands broadly overlapping the Rhine, supporting evidence that it may act as a barrier to gene flow. Finally we explored recent changes in the effective population size (N_e) using the IBDNe method. IBDNe harnesses information from long inferred segments of Identity By Descent (IBD) in a population to estimate N_e across recent generations. We observed a super-exponential population expansion across the past 50 generations, which rapidly increases in rate from ~1650 CE onwards, roughly corresponding to the Dutch Golden age in the 17th century. This growth may have been triggered by the increased prosperity and immigration during this period. Notably our N_e estimates from the northern provinces are systematically lower than in the south, consistent with increased IBD sharing and ROH in northern provinces reported previously.

2760W

Detection of shared balancing selection in the absence of trans-species polymorphism. *X. Cheng^{1,2}, M. DeGiorgio^{1,3,4}.* 1) Department of Biology, Pennsylvania State University, University Park, PA; 2) Huck Institutes of Life Sciences, Pennsylvania State University, University Park, PA; 3) Department of Statistics, Pennsylvania State University, University Park, PA; 4) Institute for CyberScience, Pennsylvania State University, University Park, PA.

Trans-species polymorphism has been widely used as a key sign of long-term balancing selection across multiple species. However, such sites are often rare in the genome, and could result from mutational processes or technical artifacts. Few methods are yet available to specifically detect footprints of trans-species balancing selection without using trans-species polymorphic sites. In this study, we develop summary- and model-based approaches that are each specifically tailored to uncover regions of long-term balancing selection shared by a set of species by using genomic patterns of intra-specific polymorphisms and inter-specific fixed differences. We demonstrate that our trans-species statistics have substantially higher power than single-species approaches to detect footprints of trans-species balancing selection, and are robust to those that do not affect all tested species. We further apply our model-based methods to human and chimpanzee whole genome sequencing data. In addition to the previously-established MHC and malaria resistance-associated *FREM3/GYPE* regions, we also find outstanding genomic regions involved in barrier integrity and innate immunity, such as the *GRIK1/CLDN17* intergenic region, and the *ABCA13* and *THSD7B* genes. Our findings echo the significance of pathogen defense in establishing balanced polymorphisms across human and chimpanzee lineages. Finally, we show that these trans-species statistics can be applied to and work well for an arbitrary number of species, and integrate them into open-source software packages for ease of use by the scientific community.

2761T

A comprehensive map of genetic variation in the world's largest ethnic group - Han Chinese. C. Chiang¹, S. Mangul², C. Robles², S. Sankararam². 1) University of Southern California, Los Angeles, CA; 2) University of California, Los Angeles, Los Angeles, CA.

As are most non-European populations around the globe, the Han Chinese are relatively understudied in population and medical genetics studies. From low-coverage whole-genome sequencing of 11,670 Han Chinese women we present a catalog of 25,057,223 variants, including 548,401 novel variants that are seen at least 10 times in our dataset. Individuals from our study come from 24 out of 33 administrative divisions across China (including 19 provinces, 4 municipalities, and 1 autonomous region), allowing us to study population structure, genetic ancestry, and local adaptation in Han Chinese. We identify previously unrecognized population structure along the East-West axis of China, demonstrate that a general pattern of isolation-by-distance among Han Chinese, and report unique signals of admixture across geographical space, such as European influences among the Northwestern provinces of China. Finally, we identified a number of highly differentiated loci, such as MTHFR, ADH7, and FADS, among others. These putative signals of local adaptation may be driven by immune response, climate, and diet in the Han Chinese. On the other hand, we find that Neandertal ancestry does not vary significantly across the provinces, and contrary to a previous report, Neandertal ancestry does not explain a significant amount of heritability in depression. We have published allele frequency estimates stratified by administrative divisions across China in the Geography of Genetic Variant browser. Our findings provide the largest genetic data set so far made available for Han Chinese and provide insights into the history and population structure of the world's largest ethnic group.

2762F

Transcriptome and chromatin accessibility in diverse African populations enables fine-mapping GWAS. M.K. DeGorter¹, E. Karakoc^{2,3}, T. Carstensen^{2,3}, M.P.S. Moorthy^{1,4}, V.R. Dwaracherla^{1,4}, K.S. Smith¹, C. Pomilla^{2,3}, J. Fellay^{5,6}, P. McLaren^{7,8}, D. Gurdasani^{2,3}, P. Flícek^{9,10}, S.B. Montgomery¹, M. Sandhu^{2,3}. 1) Department of Pathology, Stanford University, Stanford, CA; 2) Human Genetics, Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK; 3) Department of Medicine, University of Cambridge, Cambridge, UK; 4) School of Engineering, Stanford University, Stanford, CA; 5) School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland; 6) Precision Medicine Unit, Lausanne University Hospital, Lausanne, Switzerland; 7) JC Wilt Infectious Diseases Research Centre, Public Health Agency of Canada, Winnipeg, Canada; 8) Department of Medical Microbiology and Infectious Diseases, University of Manitoba, Winnipeg, Canada; 9) European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK; 10) Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK.

The increased genetic diversity and lower linkage disequilibrium in African populations provides a powerful opportunity to improve our understanding of genetic variants influencing gene expression, gene regulation and disease. Despite this potential, functional genomics data from African populations have been mainly limited to the Yoruba population sample. To more broadly characterize transcriptome diversity in Africa, we measure gene expression using RNA-Seq in LCLs from 599 individuals from six African populations, including 166 Maasai (MKK), 97 Luhya (LWK), 42 Yoruba (YRI), 112 Gambian (GWD), 99 Esan (ESN) and 83 Mende (MSL). Using whole genomes from the 1000 Genomes Project and 166 deeply-sequenced Maasai, we identify expression and splicing quantitative trait loci (eQTLs/sQTLs) and assess patterns of sharing between populations and with European individuals sampled as part of the GEUVADIS project. We use ATAC-Seq to assess chromatin accessibility in a subset of 100 of these individuals, including from 17 GWD, 16 LWK, 17 ESN, 17 MKK, 16 MSL and 17 YRI individuals. Using these data, we characterize differential gene expression and differential chromatin accessibility across African populations and their relationship to population differentiation. To characterize the whole range of the transcript structure in African populations, we also assemble novel transcripts in these six populations. In addition to identifying 99.4% of all known annotated transcripts, we identify novel transcripts and exons, providing insights into the transcriptomic landscape among African populations. Finally, we use gene expression and chromatin accessibility in African populations to fine-map GWAS signals for a range of complex diseases. This project provides a powerful functional genomics resource for biomedical research to identify causal variants, understand population genetic history and interpret the genomic basis of disease.

2763W

A statistical model for reference-free inference of archaic local ancestry.

A. Durvasula¹, S. Sankararaman^{1,2,3}. 1) Human Genetics, UCLA, Los Angeles, CA; 2) Computer Science, UCLA, Los Angeles, CA; 3) Bioinformatics IDP, UCLA, Los Angeles, CA.

Statistical analyses of genomic data from diverse human populations have demonstrated that archaic hominins, such as Neanderthals and Denisovans, interbred with the ancestors of modern humans. Central to these analyses are methods for inferring archaic ancestry along the genomes of present-day individuals. Methods for archaic local ancestry inference rely on the availability of reference genomes from the ancestral archaic populations for accurate inference. However, several instances of archaic admixture lack archaic genomes, making it difficult to characterize these events. We present a statistical method that combines diverse population genetic summary statistics to infer archaic local ancestry inference without access to an archaic reference genome. We pose the problem as a classification problem, where we are interested in the probability of a haplotype being archaic. We simulate data using the coalescent according to a demographic history that approximates human-Neanderthal history. Then, we calculate summary statistics informative of admixture and train a logistic regression model, learning the weights of each summary statistic. We validate the accuracy and robustness of our method in simulations and find that our method is robust to structure as well as misspecification in the demographic model. When applied to genomes of Europeans, our method recovers segments that are substantially enriched for Neanderthal ancestry without having access to any Neanderthal reference genomes. Overall, we find ~1.99% of European genomes have Neanderthal ancestry at a false discovery rate of 20% consistent with previous estimates. We also validate previous signals of high frequency archaic introgression at the *BNC2* gene and *OAS* gene cluster. Finally, we recapitulate the reduction in the frequency of archaic ancestry in selectively constrained regions of the genome. We then applied our method to west African Yoruban genomes and find that on average, ~7.97% of these genomes trace their ancestry to a deeply-diverged ancestral population. We find that these segments of ancestry are not enriched for Neanderthal ancestry suggesting that this ancestry is not a result of Neanderthal introgression. We also find no enrichment for ancestry related to Biaka and Mbuti populations suggesting that this ancestry does not derive from these hunter-gatherer populations. Our results indicate that our method can be useful in characterizing ancient admixture in modern human populations.

2764T

Population genetics and historical demography of North American immigrant communities.

D. Garrigan¹, S. Song¹, A.J. Stern^{1,3}, J.K. Byrnes¹, E.L. Hong¹, C.A. Ball¹, K.G. Chahine². 1) Ancestry.com, San Francisco, CA; 2) Ancestry.com, Lehi, UT; 3) University of California Berkeley, Berkeley, CA.

Contemporary North American populations are a genetic mosaic resulting from admixture between numerous founding groups, each of whom immigrated at different times and in different numbers. A large pairwise identity-by-descent (IBD) network among millions of immigrant North American individuals enables us to identify recent fine-scale population structure. We refer to clusters in these IBD networks as "genetic communities" and the distributions of shared pairwise IBD segment lengths within these communities can be leveraged to estimate the recent trajectory of the community effective population size. Interestingly, many of the genetic communities show a pronounced enrichment for IBD segments with lengths corresponding to coalescence events occurring between 8-13 generations before the present. Our model of historical demography attributes this observed enrichment to transient reductions in the historical effective size of a genetic community. To explore possible alternative demographic models, we examine the distribution of IBD segment lengths in simulated ancestral recombination graphs generated under models that include gene flow and admixture. The simulations suggest that the most likely explanation for this transient reduction in historical effective size is a bottleneck/founder event. Furthermore, the putative bottleneck/founder event pattern is preferentially inferred for older, or more geographically isolated, immigrant communities. Conversely, the founder event pattern is absent from genetic communities who likely immigrated to North America more recently. To complete this characterization of the demographic histories of these genetic communities, we use historical genealogical records to estimate both dispersion rate and generation time from each of the communities. We found that father-child distance is significantly higher than mother-child distance among all genetic communities and generation time differs between sex and genetic communities significantly. Finally, we consider the implications of these recent demographic events by examining how the distributions of allele frequencies of clinically relevant variants and allele ages vary across these genetic communities.

2765F

Fast and accurate identification of diploid IBD segments in large biobank cohorts. X. Geng¹, A. Naseri², X. Liu³, S. Zhang², D. Zhi¹. 1) School of Bio-medical Informatics, University of Texas Health Science Center at Houston, Houston, TX; 2) Department of Computer Science, University of Central Florida, Orlando, Florida 32816, USA; 3) Department of Epidemiology, Human Genetics & Environmental Sciences, The University of Texas Health Science Center at Houston, Houston, Texas 77030, USA.

Diploid IBD segments, or IBD2 segments, are IBDs on both copies of chromosomes between two individuals. Obviously, diploid IBDs are enriched among of full-siblings and double-cousins and are useful for tracing relatives. IBD2 is relatively less well-studied compared to haploid IBD (IBD1) but over 30 unexpected IBD2 sharing were found in 1,397 individuals from HapMap III, suggesting the presence of unnoted inbreeding. Effectively, diploid IBDs create "local identical twins", i.e., complete identical pair of chromosomal segments only up to a few mutations. Therefore, studies of diploid IBDs is potentially connected to the twin-study literature. Also, IBD2 is informative for mapping phenotypes with strong local recessive effects. IBD2 matrix can be used as an additional matrix for estimating heritability beyond additive effects. Although diploid IBDs are much-less prevalent than haploid IBDs, they are enriched in populations experienced a recent bottleneck or populations with endogamy. Therefore, they are also informative for inferring recent demographic history. However, diploid IBDs are not well-studied in population settings due to lack of sensitive and efficient detection methods. We developed an efficient method for the detection of diploid IBDs in population setting, RaPID-IBD2. Our methods is an extension of our previous methods for IBD1 segment detection, RaPID. In simulated data sets, we showed that our method scales up linearly with the sample size, and can detect IBD2 segments with high power and accuracy. We apply RaPID-IBD2 in UK Biobank data and reveal interesting patterns.

2766W

Genetic ancestry and population structure of Scotland and its surrounding isles. E.H. Gilbert¹, S. O'Reilly², M. Merrigan², D. McGettigan², V. Vitart³, P.K. Joshi⁴, D.W. Clark⁴, H. Campbell⁴, C. Hayward⁵, S. Ring⁵, J. Golding⁶, N. Timpson⁷, P. Navarro⁸, S.M. Kerr⁸, C. Amador⁸, A. Campbell⁸, C.S. Haley^{8,9}, D.J. Porteous⁸, G.L. Cavalleri¹⁰, J.F. Wilson^{3,4}. 1) Molecular and Cellular Therapeutics, Royal College of Surgeons in Ireland, Dublin, Leinster, Ireland; 2) Genealogical Society of Ireland, Dún Laoghaire, Ireland; 3) MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh, Scotland; 4) Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, Scotland; 5) Bristol Bioresource Laboratories, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK; 6) Centre for Child and Adolescent Health, Bristol Medical School, University of Bristol, Bristol, UK; 7) Avon Longitudinal Study of Parents and Children, MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK; 8) Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh, Scotland; 9) The Roslin Institute and Royal (Dick) School of Veterinary Sciences, University of Edinburgh, Edinburgh, Scotland; 10) FutureNeuro Research Centre, Royal College of Surgeons in Ireland, Dublin, Ireland.

Scotland and Ireland are separated in places by less than 20 kilometres of sea. They share related Gaelic languages and similar frequencies of particular alleles and phenotypes, hinting at shared ancestry. The population structure within England and Ireland have recently been described. However, the extent of structure within the majority of Scotland, its surrounding islands, and their links to Ireland are less understood and have been described with less resolution. We present an analysis of the British Isles and Ireland using a combined and comprehensive sample (n=2,556) of all major regions – expanding coverage in mainland Scotland (n=567), the Hebrides (n=57), the Isle of Man (n=40), Orkney (n=111) and Shetland (n=172). By analysing individuals with extended ancestry from specific regions with haplotype-based methods, we demonstrate extensive structure in all regions of the British Isles and Ireland, as well as some of the finest scale structure observed worldwide within Orkney. We resolve the shared genetic history between Ireland and Mainland Scotland, confirm the strongest differentiation of Orkney and Shetland from other populations, show the major differentiation in Mainland Scotland is between the south-west and the north-east, and reveal the distinctiveness of the Hebrides and the Isle of Man. Finally we also show a decreasing cline of Norwegian ancestry in the north of Britain and Ireland, across northern Scottish islands to Ireland and the Isle of Man, following the spread of the Norse Vikings. Our work represents a comprehensive description of genetic structure in the British Isles and Ireland and greatly expands the knowledge of genetic stratification within the north of the British Isles. This informs on the study of rare genetic variants and genetic trait associations in these populations, as well as supplementing our knowledge of Scottish history.

2767T

Population structure and demographic history of Ethiopian hunter-gatherers. S. Gopalan¹, R.E.W. Berf², G.M. Belbin³, C.R. Gignoux⁴, M.W. Feldman⁵, B.S. Hewlett⁶, B.M. Henn⁷. 1) Stony Brook University, Stony Brook, NY; 2) Colorado State University, Fort Collins, CO; 3) Icahn School of Medicine at Mount Sinai, New York, NY; 4) University of Colorado, Anschutz Medical Campus, Aurora, CO; 5) Stanford University, Stanford, CA; 6) Washington State University, Vancouver, WA; 7) University of California, Davis, Davis, CA.

The process by which populations transitioned to agropastoralism remains a major research question in both archaeology and genetics. In particular, the fate of hunting and gathering populations during the agricultural transition, by cultural adoption, extinction, or incorporation into expanding populations, may be heterogeneous within Africa. Here, we investigate the genetic signatures of a transition to agriculture by studying three populations in southwest Ethiopia: the Chabu, who were primarily hunter-gatherers until the 1990s, and their close neighbours, the Majang and the Shekkacho. The genomes of the Chabu indicate that they are a genetically distinct group from all other Ethiopians characterized so far, with strong genetic affinities to Mota, an ancient Ethiopian man who lived 4,500 years ago. They are also more related to the Majang than the Shekkacho, with Chabu individuals sharing a mean of 49,706 kb of their genomes identically by descent with Majang individuals, compared to only 1,780 kb with Shekkacho individuals. We also show that the Chabu have been experiencing a steady decline in population that has accelerated in the past 30 generations, and are currently significantly more endogamous than their neighbours. We compare and contrast the demographic histories of the Chabu with additional populations of the region, including other former hunter-gatherers. These results serve to elucidate the various consequences the spread of agriculture had on hunter-gatherer populations in Ethiopia.

2768F

Identifying and classifying shared selective sweeps from multilocus data. A.M. Harris^{1,2}, M. DeGiorgio^{2,3,4}. 1) Molecular, Cellular, and Integrative Biosciences, Pennsylvania State University, University Park, PA; 2) Department of Biology, Pennsylvania State University, University Park, PA; 3) Department of Statistics, Pennsylvania State University, University Park, PA; 4) Institute for CyberScience, Pennsylvania State University, University Park, PA.

Positive selection causes beneficial alleles to rise to high frequency, resulting in a selective sweep of the diversity surrounding the selected sites. Accordingly, the signature of a selective sweep in an ancestral population may still remain in its descendants. Identifying genomic regions under selection in the ancestor is important to contextualize the timing of a sweep, but few methods exist for this purpose. To uncover genomic regions under shared positive selection across populations, we apply the theory of the expected haplotype homozygosity statistic H_{12} , which detects recent hard and soft sweeps from the presence of high-frequency haplotypes. Our statistic, $SS-H_{12}$, is distinct from other statistics that detect shared sweeps because it requires a minimum of only two populations, and properly identifies independent convergent sweeps and true ancestral sweeps, with high power. Furthermore, we can apply $SS-H_{12}$ in conjunction with the ratio of a different set of expected haplotype homozygosity statistics to further classify identified shared sweeps as hard or soft. Finally, we identified both previously-reported and novel shared sweep candidates from whole-genome sequences of global human populations. Previously-reported candidates include the well-characterized ancestral sweeps at *LCT* and *SLC24A5* in Indo-European populations, as well as *GPHN* worldwide. Novel candidates include an ancestral sweep at *RGS18* in sub-Saharan African populations involved in regulating the platelet response and implicated in sudden cardiac death, and a convergent sweep at *C2CD5* between European and East Asian populations that may explain their different insulin responses.

2769W

Mapping African ancestry among African Americans. C.C. Jones^{1,2}, P.S. Ramos^{3,4}, S.J. Chanock⁵, S.A. Tishkoff^{6,7}, S.M. Williams⁸, M.C. Aldrich^{1,2,9}, *African American Lung Cancer Consortium*. 1) Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN; 2) Department of Thoracic Surgery, Vanderbilt University Medical Center, Nashville, TN; 3) Department of Medicine, Medical University of South Carolina, Charleston, SC; 4) Department of Public Health Sciences, Medical University of South Carolina, Charleston, SC; 5) Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD; 6) Department of Genetics, University of Pennsylvania, Philadelphia, PA; 7) Department of Biology, University of Pennsylvania, Philadelphia, PA; 8) Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH; 9) Division of Epidemiology, Vanderbilt University Medical Center, Nashville, TN.

The transatlantic slave trade led to the forced movement of Africans to the US, with distinct ports of embarkation in Africa and debarkation along the coastal Southeast. Subsequent mating between African- and European-descent populations resulted in mixed ancestry in African Americans. We present fine-scale resolution and a geographic assessment of African genetic ancestry among African Americans from across the US, with a focus on the Southeast, where most slave trade ports of entry occurred. Rural communities within the Southeastern US were a focus of our study due to potentially less admixture. We used genome-wide data from 6,998 African Americans from eight populations with distribution across the US, including the Gullah Sea Islands (N=239). About one-third of our population included dense sampling (N=2,133) across 12 Southeastern states. Global ancestry was estimated using ADMIXTURE with European (CEU, N=99) and representative African individuals from 13 countries (N=1,262) as putative ancestral groups. The genomic data were consistent with the presence of six (K=6) ancestral sub-populations among reference individuals: (1) European, (2) Pygmy, (3) San, (4) Fulani, (5) Senegambian/Sierra Leonean, and (6) Bantu ancestry. Nigerian ancestry split between Senegambian/Sierra Leonean and Bantu ancestries. In African Americans, Senegambian/Sierra Leonean/Nigerian ancestry contributed the highest median percentage of African ancestry (36%), followed by Bantu/Nigerian ancestry (30%). Compared to study populations outside the South, samples in the Southeast and the Gullah had substantially more Senegambian/Sierra Leonean/Nigerian ancestry and less European ancestry. Within the Southeast, European ancestry was significantly higher among individuals living in urban areas ($p=0.002$), while Senegambian/Sierra Leonean/Nigerian ancestry was significantly higher in rural areas ($p=0.01$). Ancestry was significantly correlated with geolocation, with Senegambian/Sierra Leonean/Nigerian ancestry increasing from West to East ($p=1.6 \times 10^{-10}$) and from North to South ($p=2.1 \times 10^{-4}$) within the Southeastern US. Our study established multiple sources of West African ancestry among African Americans and provided the first fine-scale characterization of geographic variation in African ancestry across the Southeastern US. These findings shed light on the African ancestries and population structure of African Americans.

2770T

From genome-wide to world-wide: Shared genetic etiology and ancestry among major complex diseases. T.B. Mersha, Y. Gautam, S. Ghandikosa. Cincinnati Children's Hospital Medical Center, Cincinnati, OH.

While genomic studies have been successful in identifying genetic variants that influence specific disease risk primary in European ancestry, it is important to understand the impact of genetic variants across multiple ancestry. An improved understanding of shared genetic etiologies could provide valuable insights into shared pathophysiological mechanisms and novel therapeutic targets. Our aim is to determine unique and shared genetic risk factors among complex diseases in the context of big data and ancestry variations. Cross-phenotype genetic studies can also provide a better understanding of phenotypic characteristics associated with particular genetic variants. Using genomic resources from the database of Genotypes and Phenotypes (dbGaP) and NHGRI GWAS catalog and our own datasets, we will survey and present geographically relevant risk factors and pathways in human populations around the world. We will present novel research strategies such as cross-database gene set analysis and ancestry variations among genome-wide studies. In cross/database gene set analysis, we will present how we implement over-representative analysis methods to evaluate the relative contributions of the shared gene clusters to individual disease entity by using hypergeometric distributions. Additionally, we will compare and present the SNP variation datasets and large-scale consortium gene expression patterns across multiple disease using GTEx, ENCODE, and NIH Roadmap Epigenomic for functional annotations of non-coding genome with genomic, transcriptomic, and epigenomic data resources. In this presentation, we hope to deliver research strategies to tackle the complex genetic relationships across multiple diseases and ancestry, and outline the importance of expanding current genome wide research in European ancestry to worldwide.

2771F

Building human reference genomes for Africa. *M.O. Pollard*^{1,2}, *E. Karakoc*^{1,2}, *T. Porter*^{1,2}, *J. Prado Martinez*¹, *C. Pomilla*^{1,2} on behalf of GDAP investigators. 1) Human Genetics, Wellcome Sanger Institute, Cambridge, United Kingdom; 2) Department of Medicine, University of Cambridge, United Kingdom.

The current version of the human reference GRCh38 is largely a mosaic of European haplotypes with limited contribution from genome sequences from Africa. As such, using the current human reference genome for global population studies may bias inferences, including for population genetics and genomic discovery. Recent attempts to broaden the repertoire of reference haplotypes, and develop population specific references have shown that these enable discovery of novel regions within the genome, and provide a resource for improved mapping and subsequent variant calling. The high levels of genetic diversity in Africa, highlight the need for a high resolution resource of genome assemblies representing genetic diversity across Africa. Here, we present the African reference resource, comprised of 43 individuals with highly contiguous assemblies from 23 populations across Africa. These data have been assembled using a combination of approaches, including 42 genomes on the 10x Chromium platform, and one Zulu genome sequenced at 55x coverage and 30x coverage on the PacBio and 10x Chromium platforms, respectively. We also created BioNano optical maps for this sample to perform assembly scaffolding, quality control and the discovery of structural variation. Our assemblies using different platforms show a high degree of concordance with regard to structural variants. We use the Africa reference genome resource to identify novel population specific sequences and structural variation. Development of a resource of reference genomes across Africa will enable population genetics research, and genomic discovery using sequencing studies from this region and globally.

2772W

High-throughput local ancestry inference reveals fine-scale population history. *A. Sedghifar*¹, *S. Song*¹, *Y. Wang*¹, *K. Noto*¹, *J. Byrnes*¹, *E.L. Hong*¹, *K.G. Chahine*¹, *C.A. Ball*². 1) AncestryDNA, San Francisco, CA; 2) AncestryDNA, Lehi, UT.

An individual's genome can be viewed as a mosaic of haplotype blocks from different ancestral origins, the sizes of which depend on the timing of admixture events. Recovering the length of these local ancestry blocks, together with their ethnic origin, provides information on the admixture and recombination events that shape current day genomes, thus shedding light on personal history as well as population history. As genomic databases rapidly approach sizes on the order of millions of genomes, there is an increased demand for super efficient approaches to identifying local ancestry blocks. Our team has developed Polly, an ultra fast algorithm for estimating genome-wide ancestry proportions in admixed individuals. Here, we present a modification of the Polly algorithm for accurately inferring local ancestry blocks. We evaluated the performance of our algorithm on simulated admixed individuals, and also assessed accuracy of estimated tract length distributions in admixed populations. Finally, we applied our method to estimate tract length distributions in historically admixed African American and Latin American populations.

2773T

The effect of consanguinity on between-individual identity-by-descent sharing. A. Severson¹, N. Rosenberg², S. Carmi³. 1) Genetics, Stanford University, Stanford, CA; 2) Biology, Stanford University, Stanford, CA; 3) Braun School of Public Health, the Hebrew University of Jerusalem, Jerusalem.

Consanguineous unions, in which mating pairs share a recent common ancestor, produce offspring whose two genomic copies possess increased sharing of long segments inherited identical-by-descent (IBD). In a population, consanguinity increases the rate at which IBD segments pair within individuals to produce runs of homozygosity (ROH). The extent to which such unions affect IBD sharing *between* rather than within individuals, however, is not immediately evident from within-individual levels. Using the fact that the time to the most recent common ancestor (TMRCA) for a pair of genomes at a specific locus is inversely related to IBD sharing between the genomes in the neighborhood of the locus, we study IBD sharing for a pair of genomes sampled either within an individual or in different individuals. We develop a coalescent model for a set of mating pairs in a diploid population, treating the fraction of consanguineous unions as a parameter. Considering a variety of types of consanguinity, we determine the effect of the consanguinity rate on TMRCA for lineage pairs sampled either within an individual or in different individuals. We find that consanguinity not only increases within-individual IBD sharing, it also increases between-individual IBD sharing, with the magnitude of the effect increasing with the kinship coefficient of the type of consanguineous union. Considering ROH and IBD computations in Jewish populations with consanguinity rates reported from demographic data, the model provides an explanation for an observation that increases in consanguinity and ROH levels inflate between-individual IBD sharing in a population.

2774F

Population diversity of Austro-Asiatic speakers: Complex substructure and intricate layered network of dispersals. D. Tagore¹, F. Aghakhanian², R. Naidu², P.P. Majumder¹, M.E. Phipps², A. Basu¹. 1) National Institute of Biomedical Genomics, Kalyani, West Bengal, India; 2) Jeffrey Cheah School of Medicine and Health Sciences, Monash University (Malaysia), Selangor, Malaysia.

India and Southeast Asia (SEA) are home to diverse ethno-linguistic groups; the Austroasiatic (AA) language speakers being one of them. Early genetic studies postulated that the ancestors of extant AAs arrived with the first wave of modern human migration out-of-Africa ~60,000 years before present (YBP). The AAs are sparsely distributed in North-East, East and Central India and in many countries of SEA, often as isolated small population groups. Aphorism among linguists and archeologists link the presence of the AAs with the spread of rice cultivation ~10,000 YBP, yet many AAs across the entire Indian peninsula and SEA remain predominantly hunter-gatherers. Events leading to the scattered distribution of AAs over this vast geographical space and the disparity in the times of arrival of the AA ancestors and AA language spread are both poorly understood. To address these issues we analyzed genome-wide polymorphism data on 189 individuals from 11 AA populations of India and SEA (Malaysia) along with comparable genetic data on 322 individuals neighboring the AAs but belonging to different ethno-linguistic background. We also used appropriate subsets of individuals from the 940 individuals in HGDP (Human Genome Diversity Project). AAs of Malaysia, AAs of India and the Tibeto Burman speakers (TB) share multiple distinct long Identical-by-Descendant segments suggesting evolution from a distant common ancestor. ADMIXTURE analysis inferred that the genetic ancestry attributed to the Malaysian AAs, is significantly higher amongst TB than in Indian AAs ($P < 2.117 \times 10^{-10}$) providing evidence of gene flow between geographically close but linguistically distant TB and Malaysian AAs. This is further corroborated by Treemix analysis. Indian AAs, who also are geographically close to TB form a separate cluster in Principal Components and fineSTRUCTURE analyses while TB and Malaysian AAs closely cluster indicating that neither geography nor language is the sole determinant of genetic distance and population diversity. Also, significantly higher values of D statistics ($|Z| > 3$) indicate that unlike the Indian AAs, both TB and Malaysian AAs have experienced extensive gene flow from the "East Asians of HGDP". This could increase the apparent genetic affinity between Malaysian AAs and TB of India. Apart from revealing directional migration in SEA, our study indicates that this gene flow may help to understand the linguistic history of this region better.

2775W

Differences in polygenic architecture between multiple ethnic human populations revealed by pathway and epistasis analysis. *M.C. Turchin^{1,2}, L. Crawford^{1,3,4}, S. Ramachandran^{1,2}*. 1) Center for Computational Molecular Biology, Brown University, Providence, RI; 2) Department of Ecology and Evolutionary Biology, Brown University, Providence, RI; 3) Department of Biostatistics, Brown University, Providence, RI; 4) Center for Statistical Sciences, Brown University, Providence, RI.

Genome-wide association (GWA) studies have identified thousands of significant genetic associations in humans across a number of complex traits, including height, body mass index (BMI), and type 2 diabetes. However, the vast majority of these studies have been conducted in datasets of predominantly European ancestry (Popejoy & Fullerton 2016). It has generally been thought that complex trait genetic architecture should be transferable across populations of different ancestries; but recent work has shown a number of differences between ethnic groups, including heterogeneity in both the causal variants being discovered and in the effect size estimates for many overlapping variants (Martin et al. 2017, Wojcik et al. 2017). Here we compare and reveal further evidence that the genetic architecture of complex traits is fundamentally different between human ethnic groups. To reveal these differences, we jointly leverage the benefits of pathway and epistasis analysis. Under the assumption that a given trait may have differential polygenic architectures across human ancestries, we hypothesize that differences in epistatic effects among human populations may be particularly enriched. However since polygenic traits also tend to have smaller GWA effect sizes, combining variants via pathway analysis may allow us to better reveal these signals. To accomplish this, we build on a recently published method that identifies marginal epistasis (Crawford et al. 2017), moving from testing single variants to testing groups of variants. We apply our method to multiple ancestries present in the UKBio-Bank dataset (Sudlow et al. 2015). Using the C2 gene set from the Broad Institute's MSigDB (Liberzon et al. 2011) and by analyzing height, BMI, and other morphometric traits, we find evidence for differential architecture in these phenotypes among African, Indian, Chinese, and British individuals. We also find evidence that these associations are further differentiated by heterogeneous signals for epistatic effects. This work further reveals that complex trait architecture is heterogeneous among human populations and that assumptions regarding transferability of primarily European-based results may be fundamentally flawed.

2776T

F_{ST} never satisfies the triangle inequality for biallelic markers with distinct allele frequencies. *I. Arbisser, N. Rosenberg*. Stanford University, Stanford, California, CA.

F_{ST} , the widely used population-genetic statistic proposed by Sewall Wright in 1951, describes the apportionment of genetic variation within and between populations. Pairwise F_{ST} is often treated as a measure of genetic distance between populations between two populations. However, F_{ST} is not a true distance metric. A distance metric in the mathematical sense must satisfy the triangle inequality: for any three points, the sum of the two shorter distances is always greater than the third, longest, distance. Sewall Wright knew that F_{ST} did not always satisfy the triangle inequality and provided a counterexample. Here, we prove the stronger result, that for biallelic markers whose allele frequencies differ across three populations, F_{ST} never satisfies the triangle inequality. We study the deviation from the triangle inequality as a function of the allele frequencies of three populations and show analytically that the maximal deviation occurs, in fact, at the specific counterexample suggested by Sewall Wright. We also examine the extent to which F_{ST} fails to satisfy the triangle inequality in genome-wide data from human populations. In data on CEU, CHB, and YRI from the HapMap Project, we find that some loci have frequencies that produce deviations from the triangle inequality near the maximum. We examine the consequences of the theoretical results for various types of data analysis that take distance matrices as input, including multidimensional scaling analysis and inference of neighbor-joining trees.

2777F

Cryptic Native American ancestry recapitulates population-specific migration and settlement of the continental United States. A.B. Conley^{1,2,3}, L. Rishishwar^{1,2,3}, I.K. Jordan^{1,2,3}. 1) School of Biological Sciences, Georgia Institute of Technology, Atlanta, Georgia USA; 2) IHRC-Georgia Tech Applied Bioinformatics Laboratory, Atlanta, Georgia, USA; 3) PanAmerican Bioinformatics Institute, Cali, Valle del Cauca, Colombia.

European and African descendants settled the continental US during the 17th-19th centuries, coming into contact with established Native American populations. The resulting admixture among these groups yielded a significant reservoir of cryptic Native American ancestry in the modern US population. We analyzed the patterns of Native American admixture seen for the three largest genetic ancestry groups in the US population: African American, European American, and Hispanic/Latino. The three groups show distinct Native American ancestry profiles, which are indicative of their historical patterns of migration and settlement across the country. Native American ancestry in the modern African American population does not coincide with local geography, instead forming a monophyletic group with origins in the southeastern US, consistent with the Great Migration of the early 20th century. European Americans show Native American ancestry that tracks their geographic origins across the US, indicative of ongoing contact during westward expansion, and Native American ancestry can resolve Hispanic/Latino individuals into distinct local groups formed by more recent migration from Mexico and Puerto Rico. We found an anomalous pattern of Native American ancestry from the US southwest, which most likely corresponds to the *Nuevomexicano* descendants of early Spanish settlers to the region. We addressed a number of controversies surrounding this population, including the extent of Sephardic Jewish ancestry. *Nuevomexicanos* are less admixed than nearby Mexican-American individuals, with more European and less Native American and African ancestry, and while they do show demonstrable Sephardic Jewish ancestry, the fraction is no greater than seen for other Hispanic/Latino populations.

2778W

The spatial patterns of genetic diversity in the United States. C.L. Dai¹, M.M. Vazifteh^{2,3}, C.H. Yeang⁴, R. Tachet^{2,3}, M.G. Vilar⁵, C. Ratti^{2,3}. 1) Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA; 2) Senseable City Lab, Massachusetts Institute of Technology, Cambridge, MA; 3) Department of Urban Studies and Planning, Massachusetts Institute of Technology, Cambridge, MA; 4) Institute of Statistical Science, Academia Sinica, Nankang, Taipei, Taiwan; 5) Genographic Project, National Geographic Society, Washington, DC.

Historical migration of individuals into and throughout the United States has contributed to significant mixing of distant ancestries and a heterogeneous genetic landscape. Here, we present a comprehensive view of the genetic ancestry, haplotype relatedness, and population history of 32,589 genotyped individuals from the United States participating in the National Geographic Society's Genographic Project. Varying the geographical scale of analysis, we find that diversity in ancestry origin and admixture reflects the size of present-day metropolitan areas. Haplotype sharing reveals geographically proximate decay of relatedness but also long-range relatedness, especially in populations of non-European ancestry. We further explore migration rates to understand the varying patterns across populations of different ancestries, comparing them with historical demographic events. Dating the admixture histories from ancestral haplotype tracts, we also assess the relationship between admixture patterns and geography.

2779T

Inference of population structure from ancient DNA. T.A. Joseph¹, I.

Pe'er^{1,2,3}. 1) Department of Computer Science, Columbia University, New York, NY; 2) Department of Systems Biology, Columbia University, New York, NY; 3) Data Science Institute, Columbia University, New York, NY.

One of the great successes of modern genomics is the ability to reconstruct human history from genetic data alone. Over the past decade, studies to this end have increasingly relied on ancient DNA. Yet, traditional methods to infer population history from ancient DNA either implicitly assume all samples are contemporary, or presuppose population assignment for ancient samples. To overcome this limitation, we developed Dystruct (Dynamic Structure), a time-series model of population history and stochastic variational inference algorithm to discover shared ancestry among sampled individuals (Joseph and Pe'er, In Proc. RECOMB 2018, 90-104). Individuals are modeled as mixtures of populations whose allele frequencies drift over time. Dystruct assumes multiple populations with independent histories, but correctly models later (modern) samples as mixtures of earlier (ancient) ones. We now extend Dystruct to handle more complex population histories. Using simulations, we demonstrate that the accuracy of current methods is substantially impacted by genetic drift – the random fluctuation of allele frequencies over time – when individuals are sampled over time, as is common in ancient DNA datasets. We show that Dystruct offers a correction for the effects of drift, leading to more accurate ancestry estimates. To further demonstrate the utility of our model, we applied Dystruct to a dataset of 92 ancient and 1941 modern samples genotyped at 222755 loci. Ancient samples included early hunter-gatherers groups, Neolithic farmers, and steppe herders from the Pontic-Caspian steppe. Modern samples are from global populations in the Human Origins dataset. We further compared Dystruct to ADMIXTURE, a model ubiquitous for inferring shared ancestry among modern samples, and often applied to ancient DNA datasets. We show that when later samples are mixtures of earlier ones, they appear as such in Dystruct. Conversely, ADMIXTURE tends to present such ancient samples as mixtures of their later counterparts. Most notably, Dystruct is able to capture a migration and admixture event of the steppe herders throughout Eurasia approximately 5000 years ago (Joseph and Pe'er, In Proc. RECOMB 2018, 90-104), a signal supported by the literature but missing from ADMIXTURE. We explore the extent to which modeling more complex population histories improves these results. Our model is a promising first step towards modeling the true complexity in the history of our species.

2780F

Whole-exome sequencing of 1,050 Turkish individuals reveals population structure and genetic variation. X. Song^{1,9}, Z. Coban-Akdemir^{1,9}, D. Pehlivan¹, E. Karaca¹, Y. Bayram^{1,2}, H. Du¹, T. Gambin³, S.N. Jhangiani⁴, D.M. Muzny⁴, C.A. Shaw^{1,5}, E. Boerwinkle^{4,6}, R.A. Gibbs^{1,4}, J.E. Posey¹, J.R. Lupski^{1,4,7,8}.

1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA; 2) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; 3) Institute of Computer Science, Warsaw University of Technology, Warsaw 00-665, Poland; 4) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA; 5) Baylor Genetics, Houston, TX 77021, USA; 6) Human Genetics Center, University of Texas Health Science Center at Houston, TX 77030, USA; 7) Department of Pediatrics, Baylor College of Medicine, Houston, TX 77030, USA; 8) Texas Children's Hospital, Houston, TX 77030, USA; 9) These two authors contributed equally.

Turkey, often described as a bridge between Asia and Europe, is a hub of ancient and contemporary population migration. Previous studies on healthy unrelated individuals from the Greater Middle East have demonstrated the occurrence of population diversity and admixture – but also suggested extensive consanguinity. Here, we performed whole-exome sequencing (WES) on 1,050 Turkish individuals, including 695 affected patients representing of a multitude of disease phenotypes and 355 unaffected relatives, in order to study the population structure and disease etiology of a large, but representative, Turkish cohort. We performed principal component analysis (PCA) on the Turkish samples and the African, Asian and European population samples from the 1000 Genome Project. The Turkish samples were isolated from the other three but fell closer to the Asian population. We show that 38.7% of the heterozygous and 25% of the homozygous distinct variants in the Turkish individuals were not present in ExAC, suggesting a necessity of using a population-matched control database in analyzing variant pathogenicity. We then compared the Human Phenotype Ontology (HPO) similarity for each pair of probands with their clinical features recorded in the PhenoDB database. Our analysis revealed that phenotypic heterogeneity accompanied by multi-locus variation in the Turkish population (17/30 cases) results from biallelic variation at 2 loci for an AR + AR trait combination of blended phenotypes. This increased burden of homozygous variants and a larger median span of absence of heterozygosity (AOH) of 90.4 Mb determined by BafCalculator from WES data and consistent with identity-by-descent, in the Turkish samples compared to 25.8 Mb in 5,735 non-Turkish samples locally sequenced through Baylor-Hopkins Center for Mendelian Genomics ($P = 4.98 \times 10^{-109}$). We next analyzed 167 trios in this cohort and curated a list of 606 *de novo* variants. Eleven (11/606, 1.8%) were found to be shared among 2 affected siblings in 17 families and will be investigated further for parental mosaicism. In summary, we generated a database of variants derived from 1,050 exomes for assessing rare potentially pathogenic variants of interest to disease trait biology in the Turkish population. These results show consanguinity within this cohort and emphasize the utility of AOH analysis from WES data revealing identity by descent and deleterious variant burden when evaluating the contribution of *de novo* variants to disease.

2781W

The role of recent demography in shaping patterns of genetic variation in US Hispanic/Latino populations. *M.L. Spear^{1,2}, D.G. Torgerson³, R.D. Hernandez^{2,4,5,6}*

1) Biomedical Sciences Graduate Program, UCSF, San Francisco, CA; 2) Department of Bioengineering and Therapeutic Sciences, UCSF, San Francisco, CA; 3) Department of Pediatrics, UCSF, San Francisco, CA; 4) Institute for Human Genetics, UCSF, San Francisco, CA; 5) Institute for Quantitative Biosciences, UCSF, San Francisco, CA; 6) Institute for Computational Health Sciences, UCSF, San Francisco, CA.

Understanding the genetic basis of complex phenotypes is a critical problem in medical and evolutionary genetics. The evolutionary forces of natural selection and demographic history, including migration, have shaped patterns of worldwide genetic variation, which in turn have shaped the genetic architecture of human phenotypic variation. The population of the United States has changed immensely over the 20th century as a result of immigration, and this will continue to be one of the primary modes of population growth as the US approaches a "minority-majority" country. Hispanics/Latinos, the largest and fastest growing group, are a genetically heterogeneous population as result of admixture between African, European, and Native American populations. The effect of these large-scale migrations in contributing to shaping genetic variation and subsequently phenotypic variation is unknown. For 9,646 Hispanics/Latinos from the Hispanic Community Health Study/Study of Latinos (HCHS/SOL) we tested the hypothesis that the admixture process in the US has been dynamic over the 20th century. Using ADMIXTURE and RFMix we estimated global and local ancestry and investigated how ancestry has changed over time as a function of birth year for Mexican Americans, Puerto Ricans, Dominican Republicans, Cubans, Central Americans, and South Americans. Our results have shown a significant ($p=3.3 \times 10^{-21}$) increase in Native American ancestry in both US born and foreign born Mexican Americans over time. These shifting ancestry proportions could be due to an increase in longer Native American ancestral tracts, variation in the origin of subcontinental Native American ancestry or assortative mating. Our results suggest that the admixture process in the US has been dynamic in the last century, which could have further implications on the genetic architecture of complex traits. This demonstrates a crucial need to improve our understanding of the role of demographic history in shaping the genetic architecture of phenotypic variation in Hispanic/Latino populations.

2782T

The ChileGenomico Project creates a public resource to test association between socioeconomic and epidemiologic variation among communes of Chile and ancestry. *R.A. Verdugo^{1,10}, P. Araneda¹, A. Di Genova², L. Herrera¹, M. Moraga¹, M. Acuña¹, S. Berríos¹, E. Llop¹, C.Y. Valenzuela¹, M.L. Bustamante^{1,3}, D. Digman¹, A. Symon¹, S. Asenjo¹, P. Pezo-Valderrama¹, P. López¹, A. Blanco¹, J. Suazo¹, F. Caba⁵, E. Barozet⁶, K. Salgado⁷, A.M. Naranjo⁷, P. Portales⁸, N. Loira⁹, A. Maass^{2,9}, L. Cifuentes¹*

1) Human Genetics Program ICBM, University of Chile, Santiago, R.M., Chile; 2) Mathomics, Centro de Modelamiento Matemático y Centro para la Regulación del Genoma, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, Santiago, Chile; 3) Departamento de Psiquiatría y Salud Mental Norte, Facultad de Medicina, Universidad de Chile, Santiago, Chile; 4) Instituto de Investigación en Ciencias Odontológicas, Facultad de Odontología, Universidad de Chile, Santiago, Chile; 5) Instituto de Salud Pública, Facultad de Medicina, Universidad Andrés Bello, Santiago, Chile; 6) Departamento de Sociología, Facultad de Ciencias Sociales, Universidad de Chile, Centro de Estudios de Conflicto y Cohesión, Social, Santiago, Chile; 7) Facultad de Ciencias de la Salud, Universidad de Tarapacá, Arica, Chile; 8) Corporación Municipal de Desarrollo Social, Iquique, Chile; 9) Departamento de Ingeniería Matemática, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, Santiago, Chile; 10) Departamento de Oncología Básico Clínica, Facultad de Medicina, Universidad de Chile, Santiago, Chile.

Introduction and objectives. The Chilean population is the result of a process of admixture between Europeans, Africans and Amerindians since the XXVI century. The genetic variation between them allows estimating the proportion of ancestors of each of these origins in current individuals. Estimates of ancestry show significant differences between subpopulations of Chile, related to latitude and social stratum. However, a systematic study of the differences in the ancestry components of Chileans has not been carried out. **Material and methods** The ChileGenomico project carried out the largest sampling to date of mixed Chilean population ($n = 3200$), recruiting healthy individuals in nine urban centers from Arica to Puerto Montt. The genome of 48 individuals with Amerindian ancestry was typified by complete genomic sequencing (3.6x) or microarrays. Using a genetic algorithm to maximize genomic coverage and ancestry normativity, we selected 147 autosomal SNPs, then genotyped by KASP in 2,843 individuals. **Results.** The average ancestry was 51% European, 5% African and 44% Amerindian (20% Aymara and 25% Mapuche). The Amerindian ancestry is lower in the central zone, the Aymara decreases with latitude and the Mapuche increases. The Amerindian component rises 5-10% per level of decline in socioeconomic stratification. Average ancestry estimates for 39 communes of Chile were estimated and made available through an online that allows identifying association to socio-economic and epidemiological data (<http://genoma.med.uchile.cl/ancestry>). **Conclusions.** We have created a public resources that will be allow controlling for ancestry heterogeneity of the Chilean population in public health studies and scientific research.

2783F

Interpreting variants of uncertain significance using common variation in non-human primates. H. Gao, L. Sundaram, S. Reddy, J. McRae, S. Batzoglou, K. Farh. Illumina Artificial Intelligence Laboratory, Illumina Inc., Foster City, CA.

Chimpanzees share 99.4% amino acid sequence identity with human, suggesting that purifying selection operating on naturally occurring chimpanzee variation might also model the consequences on fitness of human mutations that are identical-by-state. Using population sequencing data from human and six species of non-human primates, we show that the presence of an identical-by-state variant at common allele frequencies in a non-human primate species is strongly indicative that the variant is clinically benign in human. We show that, on average, each non-human primate species contributes more likely benign missense variants than the whole of the ClinVar database. Because common variation in each primate species has evolved independently, we estimate that cataloging common variation from each of the 504 extant primate species would be sufficient to saturate most of the possible benign positions in the human genome, substantially improving interpretation for millions of variants of uncertain significance.

2784W

Investigating race, ethnicity, and ancestry in clinical genomics. A.B.

Popejoy¹, D.I. Ritter², J. Berg³, V.L. Bonham¹, E. Currey⁴, S. Dwight¹, S.M. Fullerton⁵, C.R. Gignoux⁶, L.A. Hindorf⁷, E. Kenny⁷, B. Koenig⁸, S.S. Lee¹, S. Oh⁹, K.E. Ormond¹, S.E. Plon², S. Prabhu¹, E.M. Ramos⁴, M. Rivas¹, A.E. Shields^{9,10}, E.P. Sorokin¹, T.A. Thornton⁵, D. Williams⁹, M.W. Wright¹, J. Zou¹, C.D. Bustamante¹, The Clinical Genome Resource (ClinGen) Ancestry and Diversity Working Group (ADWG). 1) Stanford University, Palo Alto, CA; 2) Baylor College of Medicine; 3) University of North Carolina Chapel Hill; 4) National Human Genome Research Institute; 5) University of Washington; 6) University of Colorado, Anschutz Medical Campus; 7) Icahn School of Medicine at Mount Sinai; 8) University of California at San Francisco; 9) Harvard University; 10) Massachusetts General Hospital.

The question of how best to use ancestry information in clinical genomics remains unanswered, leaving patients from groups that are under-represented in clinical genomic databases less likely to benefit from advances in genomic medicine. For example, variant curation guidelines recommend review of allele frequency data from publicly available population databases, but these resources lack robust representation from individuals of non-European ancestry. Lack of population diversity could significantly impact the reliability and comprehensiveness of variant interpretations. The Clinical Genome Resource (ClinGen, clinicalgenome.org), a National Institutes of Health (NIH)-funded effort, is dedicated to defining the clinical relevance of genes and variants, and its mission requires the resulting clinical knowledgebase reflects the diversity of human genomic variation. ClinGen's Ancestry and Diversity Working Group (ADWG) was established to ensure clinical genomics professionals have the necessary tools such as robust databases, variant classifications and guidelines for the accurate application of genomics to individuals from traditionally under-represented ancestral populations, many of which already bear a disproportionate burden of illness. The ADWG has begun this work by conducting three primary analyses: 1) characterizing the current questions about patient race, ethnicity, and ancestry (REA) on requisition forms from clinical genetic testing laboratories that frequently submit annotations to ClinVar; 2) quantifying information about different ancestral populations in public databases used for variant interpretation and curation, e.g. gnomAD and ClinVar; and 3) surveying clinical genomics professionals about perceptions and general use practices of REA data in clinical settings. Preliminary results of our analyses show that different clinical labs use highly variable questions to collect REA data, and the amount and quality of information available for clinically relevant variants is highest in European ancestry populations. Clinical genomics professionals may vary in their perceptions about patient REA information and the appropriateness of its use in a clinical setting, which is among the hypotheses we are testing with our survey of clinicians. The ADWG is using these data in the development of recommendations for the use of REA information in clinical genomics.

2785T

Available protein 3D structures do not reflect human genetic and functional diversity.

G. Sliwoski^{1,2}, N. Patel³, R.M. Sivley¹, C.R. Sanders^{2,4,5}, J. Meiler^{2,6}, W.S. Bush³, J.A. Capra^{1,7,8}. 1) Biomedical Informatics, Vanderbilt University, Nashville, TN; 2) Center for Structural Biology, Vanderbilt University, Nashville, TN; 3) Inst for Computational Biology, Dept of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH; 4) Biochemistry, Vanderbilt University, Nashville, TN; 5) Dept of Medicine, Vanderbilt University Medical Center, Nashville, TN; 6) Chemistry, Vanderbilt University, Nashville TN; 7) Biological Sciences, Vanderbilt University, Nashville, TN; 8) Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN.

Genomic databases and clinical trials are substantially biased towards European ancestry populations, and this bias significantly contributes to health disparities. Structural biology has an essential role in investigating protein function and clinical variant interpretation, providing powerful tools for investigating the impact of genetic variants on protein structure and function. However, studies that analyze the 3D structure of proteins typically consider a single "canonical" amino acid sequence as representative of the protein. Here, we evaluate the potential for this simplification to bias results toward different populations by evaluating how well all 66,971 experimentally characterized human protein 3D structures represent the sequence diversity of the proteins they model. Thousands of protein structures have unrepresented alternative sequences commonly found in human populations, and African ancestry individuals' sequences are the least likely to be represented by available structures. Because sequence variability is often limited to a few positions within a protein, we evaluate the likelihood of these small changes to influence protein function. Combining existing annotations and computational modeling, we identify thousands of proteins for which use of a single structure as representative of "wild type" may bias results against certain populations or individuals. Variants segregating in human populations, but unrepresented in structures, are observed across functional sites involved in stability (134 disulfide bond cysteines), regulation (94 phosphorylation sites), DNA binding (322 residues), small molecule binding (1,463 residues with 362 within drug binding sites), and protein-protein interfaces (6,144 residues). We computationally model a subset of unrepresented variants' effects on protein stability and protein-protein interaction. Changes in predicted protein stability are found for 28% (156) of the 556 variants, with stabilizing (41) and destabilizing (115) effects predicted. Of 161 protein-interface variants modeled, 25% (41) are predicted to impact protein-protein binding. These variants in human populations have potential to impact the study of their protein's structure and function. With the widespread use of protein structures in basic science and clinical variant interpretation, human protein sequence and structural diversity must be considered to enable accurate and reproducible conclusions from structural analyses.

2786F

Mining patterns of singleton and loss-of-function variants in deeply sequenced genomes. D. Taliun¹, S. Gogarten², S. Zöllner¹, R. Hernandez², M. Boehnke¹, G. Abecasis¹, The NHLBI Trans-Omics for Precision Medicine (TOPMed) Program. 1) Department of Biostatistics, University of Michigan, Ann Arbor; 2) Department of Biostatistics, University of Washington, Seattle; 3) Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco.

The Trans-Omics for Precision Medicine Program (TOPMed) aims to discover mechanisms underlying heart, lung, blood, and sleep disorders by generating whole genome high depth sequencing data in large numbers of well-phenotyped individuals of different ancestries. The current TOPMed release of 463 million high quality genetic variants from >62,000 genomes is already available through both dbSNP and the BRAVO variant browser. The large number of deeply sequenced (38X) individuals makes TOPMed the best source to date for identifying biologically important regulatory and protein-coding regions through patterns of singletons and very rare loss-of-function (LoF) variants. In this analysis, we focused on a subset of >53,000 genomes with permission for broad sharing through dbGaP and where quality control had been completed. These genomes include 410 million variants (46% singletons), including 4,959,646 protein coding variants (47% singletons) and 242,713 predicted protein truncating (or loss-of-function, LoF) variants (58% singletons). The average individual carried 249 LoF variants. Only 997 (5.1%) out of 19,479 protein coding genes had no LoF variants across all individuals. We searched for gene sets with lower numbers of LoF variants per coding base-pair and lower proportions of common variants than expected. These gene sets are less tolerant to protein loss of function. Gene sets depleted in loss of function variants and in common variants included genes encoding for DNA and RNA binding proteins, spliceosomal complexes, translation initiation, and RNA splicing and processing proteins. As expected, genes involved in olfactory receptor activity had higher numbers of loss of function variants and large proportions of common variants. Genes associated with human disease in the OMIM, GWAS Catalog, ClinVar, and COSMIC (cancer) databases showed evidence of fewer LoF variants per base-pair than expected by chance. Among regulatory regions, only promoters and transcription factor binding sites were depleted in common variation after adjustment for mutability. Nonsense mediated decay and immunoglobulin genes had an elevated proportion of singletons in 3' UTR regions, which was not observed among other protein coding genes. In summary, we demonstrate the value of a large collection of high-quality genomes to identify biologically important regions through singleton and LoF variants.

1341W

Synthetic RNA-seq data for intervention and drug discovery by conditional generative adversarial networks. *M. Xiong¹, Y. Liu¹, Q. Ge^{2,1}*. 1) University of Texas School of Public Health, Houston, TX., US; 2) School of Mathematical Science, Fudan University, China.

RNA-seq data provide a powerful tool to investigate complex traits and common disorders' pathogenesis. Therefore, whole-transcriptome analysis is increasingly acquiring a key role in unravelling mechanisms underlying complex diseases. However, RNA-seq data are scarce, expensive, and concerned with privacy of individuals. The majority of RNA-seq data is inaccessible for general public research. There are increasingly interests in inexpensively simulating the RNA-seq data under various conditions, filling up the missing data, and creating a public and synthetic RNA-seq dataset to uncover the mechanism of complex diseases, explore artificial interventional operations and open a novel avenue for designing drug targets. To achieve these goals, we propose a novel, generative pipe line that consists of (1) generating ultra-realistic original RNA-seq data by conditional generative adversarial networks via primal-dual subgradient training methods which is the simplified mathematical models of the human brain and one of the most promising advances in artificial intelligence in the past decade, (2) learning functional causal models from observational RNA-seq data to identify causal paths from genes to diseases via gene expressions and phenotypes with deep causal generative neural networks and (3) learning causal implicit generative models from true observational and interventional distribution of RNA-seq with dual adversarial training which can perform interventions and drug tests on the resulting causal implicit generative models. The Generative adversarial networks can integrate diverse RNA-seq data sets, generate RNA-seq datasets under normal and disease status, infer gene regulatory networks, and predict effect of external perturbations on gene expressions. The proposed generative pipe line was applied to a RNA-seq dataset of type 2 diabetes and Alzheimer's disease (AD) with 448 individuals, 19 phenotypes and environments, six type of diseases and 299 pathways with RNA-Seq. We have identified 20 causal gene expressions and 5 causal pathways including FoxO, MAPK and PI3K-AKT pathways shared by type 2 diabetes and AD. The generative adversarial networks are emerging as a powerful tool for understanding the underlying complex structure of the genomic, epigenomic and molecular world which we measure and explore. We are entering a new machine learning era for integrated genomic, epigenomic and imaging data analysis, and biomedical and public health research.

1342T

Discovery of obesity-associated noncoding causal variants based on convolutional neural networks. *J. Oh, K. Jang, S. Lee, T. Lee, J. Kwon, M. Bae, J.K. Choi*. Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Daejeon, Korea, Republic of.

Obesity is a metabolic disorder resulting from behavioral and heritable causes. A number of genome-wide association studies (GWASs) have been conducted on obesity, but in most cases, the underlying mechanisms by which the variants contribute to obesity remain unresolved. GWASs can only report large clusters of single nucleotide polymorphisms (SNPs), or linkage disequilibrium blocks, including not only causal variants but also many linked neutral SNPs. Moreover, a majority of the obesity-associated GWAS SNPs are found in non-coding regions. Our incomplete knowledge of noncoding regions limits the functional interpretation of underlying DNA variants. To overcome these limitations, we constructed a convolutional neural networks framework for combinational, nonlinear modeling of complex patterns shared by risk variants scattered among associated multiple loci. We applied the model to the GWAS data for the body mass index adjusted by the waist-hip ration provided by The Genetic Investigation of ANthropometric Traits (GIANT) consortium. We evaluated the performance of the model and discovered several putative causal noncoding SNPs. Our results will provide insight into the mechanisms by which noncoding variants contribute to the obese phenotypes.

1343F

An Omics Analysis, Search and Information System (OASIS) for enabling discovery in the TOPMed diabetes working group. *J.A. Perry, J.R. O'Connell on behalf of the TOPMed Diabetes Working Group.* Program for Personalized and Genomic Medicine and Division of Endocrinology, Diabetes and Nutrition, University of Maryland School of Medicine, Baltimore, MD, 21201, USA.

Introduction: Members of the TOPMed Diabetes Working Group have computed over 200 million single-variant associations between TOPMed whole genomes and harmonized phenotypes for T2D, Fasting Glucose, Fasting Insulin and HbA1c. This massive quantity of results is typical of TOPMed research and transforming this “raw information” into “biological discovery” can be challenging. Scientist-friendly tools are needed to provide an integrated approach that includes data visualization, broad annotation, and fine-mapping techniques. **Methods:** An Omics Analysis, Search and Information System (OASIS) was constructed for the TOPMed Diabetes Working Group by making major enhancements to an existing web-based application developed at the University of Maryland. The enhancements, which support the large TOPMed datasets, required redesign of the underlying database architecture and use of highly efficient data compression and data handling made available with the MMAP software (<https://mmap.github.io/>). The OASIS webserver is an approved “TOPMed Cloud Computing Platform” and thus allows Working Group analysts to directly upload association results for sharing and comparing alternative analyses. Multiple datasets can be created and access control features allow datasets to be selectively shared with other registered TOPMed OASIS users. **Results:** OASIS datasets of up to 70 million associations from TOPMed freeze4 (219 million variants) and freeze5b (470 million) have been created for diabetes phenotypes. Boxplots split by genotype, study and ancestry are used to track populations for rare variants. Significant signals driven by a single study cohort have been identified. Variants are annotated by OASIS as they are reported during a user-initiated query and those lying in regulatory regions are easily spotted. Known-loci lists can be applied to query reports to identify variants in windows around known loci. Conditional and multi-covariate analysis is available for user-selected variants and integrated LocusZoom and Haploview plots provide linkage visualizations. Queries can be filtered by p-value, effect size, gene, variant type and/or function, rsid lists, genomic positions, and allele frequency. **Conclusion:** Transforming massive volumes of TOPMed association results into “biological discovery” has been made dramatically easier. As a web-based tool, OASIS allows both analyst and non-analyst easy access to a wealth of annotation, visualization and fine-mapping techniques.

1344W

AMP-T2D: Federated analysis and discovery of genetic variants with association to type-2 diabetes. *J.D. Spalding, C.Y. González, J.I. Ciges, S. Kamatchinathan, D. Bennett, T. Keane.* European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, United Kingdom.

Novel therapeutic targets and functional pathways can be uncovered by discovering genetic variants which have an association to type-2 diabetes (T2D). The Accelerating Medicines Partnership in T2D (AMP-T2D) is a multi-sector, pre-competitive partnership among government, industry, and nonprofit organizations. AMP-T2D will produce and aggregate data, build infrastructure for data storage and presentation (knowledge base), and develop the tools and analytical methods to run on this infrastructure. The resulting open-access resource is the AMP-T2D Knowledge Portal (<http://www.type2diabetesgenetics.org>). This resource provides a wide variety of interactive tools, including custom online association analysis with dynamic sample filtering, covariate specification, and conditional analysis, while enabling sophisticated analytics of sensitive genetic data and ensuring patient privacy. Data are stored in a Data Coordinating Center at the Broad Institute and also at the European Bioinformatics Institute (EBI), in the first of several planned federated nodes. Federation allows integrated analysis across data sets, while ensuring that individual level data can remain at the site of origin where required to comply with ethical, legal, and society issues (ELSI). As the rate at which genetic data is generated increases, federation increases the capacity to store and analyse these data and reduces the requirement to distribute the whole dataset (analysis-to-data instead of data-to-analysis). As of June 2018 the EBI knowledge base (KB) was hosting 18031 samples in 4 datasets, with an additional 3404 samples, which include longitudinal and metabolite data, undergoing analysis and quality control. Federated data storage and subsequent analysis creates challenges in data harmonization, duplication, and management. All data at both knowledge bases must undergo the same analysis, imputation and quality control processes, which require that the same or equivalent pipeline is run at all KBs. To facilitate data harmonization, a registry and accessioning service have been developed. The registry defines the phenotypes (both attributes and terminologies), while the accessioning service either assigns new accessions or returns current accessions for existing object helping to prevent data duplication across a federated network. Where possible we are developing these federated processes to be compatible with standards developed by the Global Alliance for Genomics and Health (GA4GH).

1345T

Retrospective electronic medical record analysis identifies patients at risk of hypophosphatasia and illustrates paucity of family history information. C. Peroutka¹, M. Marzinke⁴, N. Parikh², A. Alade^{1,2}, J. McGready³, K. Schulze^{2,3}, J. Hoover-Fong^{1,2}. 1) McKusick Nathans Institute of Genetic Medicine, Johns Hopkins Hospital, Baltimore, MD; 2) Greenberg Center for Skeletal Dysplasias, Johns Hopkins Hospital, Baltimore, MD; 3) Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD; 4) Clinical Pharmacology Analytical Laboratory, Johns Hopkins Hospital, Baltimore, MD.

Introduction: Hypophosphatasia (HPP) is a pleiotropic condition caused by variants in tissue-nonspecific alkaline phosphatase (*TNSALP/ALPL*), manifesting as perinatal lethal through adulthood with skeletal fragility, respiratory compromise, neurologic and constitutional symptoms. Often unrecognized as pathologic, low serum alkaline phosphatase (AP) is a key indicator of HPP.

Thus, HPP is underdiagnosed. Misdiagnosis as osteoporosis may involve bisphosphonate treatment, which counteracts available enzyme replacement for HPP. We conducted a risk prevalence study for HPP, including family history data, by query of a large hospital electronic medical record (EMR), demonstrating this methodology can identify individuals at-risk of rare disease.

Methods: The Johns Hopkins (JH) IRB approved this study with waiver of consent. JH patients with ≥ 1 AP level from Jan 2013-Dec 2015 were compiled by medical record number (MRN). Site and ICD-9 codes were merged to MRNs with ≥ 1 low AP using age- and sex-specific norms. MRNs with hematologic, oncologic, drug and transplant codes associated with low AP, or liver (direct bilirubin ≥ 5) or renal (creatinine ≥ 2.6) disease were excluded. Patients with 5-20 AP levels, $\geq 80\%$ abnormal, were prioritized for chart review. EMR data extracted included medical and family history, and physical features suggestive of HPP. A REDCap database and STATA stored and generated results. **Results:** 45,233 (49% in patients ≤ 19 yo) low AP levels from 11,730 patients yielded 7,129 patients for review after exclusion by ICD-9/site codes. 351 had 5-20 AP levels, $\geq 80\%$ abnormal. Data from 200 charts is presented. Medical history suggests HPP in 31%. Signs/symptoms include: 3% premature tooth loss, 9% excessive caries, 8% nephrolithiasis, 9% muscle pain, and 33% chronic pain. 8% have osteoporosis; 3% received bisphosphonates and 1 subject received a PTH analog. Multiple specialists provide care: 14% endocrinology, 19% orthopedics, 9% genetics, 9% rheumatology, 4% pain management, and 5% nephrology. FH suggests HPP in 2%, but only 48% have FH recorded for ≥ 3 relatives. HPP was not on the differential for any patient, though 15% had other molecular testing. **Conclusions:** EMR query of a large diverse population can identify individuals at-risk of rare, treatable conditions like HPP. With IRB approval, we are now contacting at-risk subjects to obtain histories in-person, and offer genetic counseling, molecular testing, and referral for treatment to those affected.

1346F

A comparison of classical integration methods applied on multi-omics data. Y. Zhang¹, J. Cho², Q.L. Duan^{1,2}. 1) School of Computing, Queen's University, Kingston, Ontario, Canada; 2) Department of Biomedical and Molecular Sciences, Queen's University, Kingston, Ontario, Canada.

Recent technological advancements have made it feasible to ascertain multiple types of genome-scale data. In order to gain a more comprehensive assessment of these datasets and infer knowledge about the dynamics of molecular systems, various methods have been developed for integrating multiple levels of biological datasets. However, these are limited by a priori knowledge of biological systems, algorithmic limits, and the variability of information captured at a genome-scale. Previously described methods for integrating biological datasets address three aims: 1) to cluster samples/subjects; 2) to understand molecular mechanisms; 3) to predict clinical outcomes. In this study, we applied two existing methods to integrate transcriptomics and epigenomics data from a childhood asthma cohort (GSE40732 and GSE40576) to identify genomic factors underlying asthma. These methods include: Similarity Network Fusion (SNF) tool, which is designed to cluster subjects and predict clinical outcomes by constructing a fused similarity network based on the different types of omics data; Data Integration Analysis for Biomarker discovery using a Latent component method for Omics studies (DIABLO), which identifies molecular mechanisms and clusters subjects using sparse generalized canonical correlation analysis. DIABLO shows that 35 genes from the transcriptomics data and 46 methylated sites were significantly correlated with asthma. Integration of these two levels of data discovered that a methylation probe correlated with the expression of the *GZMB* (Granzyme B) gene may underlie asthma in this cohort. Ongoing work in our lab aims to address some of the limitations of DIABLO to develop a novel integrative method that combines the network structures efficiently and allow for detection of molecular mechanisms.

1347W

Integration and visualization of phenotype-genotype data and drug response in diabetes mellitus patients using the Diabetome™ Knowledgebase. H. GH¹, R.M. Anjana², R. Praveen², J. Somasekar², S. Jebarani², V. Radha², B. Muthu Narayanan², R. Pradeepa², R. Unnikrishnan², R. Gupta², S. Santhosh¹, V. Srinivasan², V. Mohan². 1) MedGenome Inc, Foster City, CA; 2) MedGenome Labs, Bangalore, India; 3) Dr. Mohan's Diabetes Specialities Centre, Chennai India.

Diabetes research today suffers from a number of challenges due to the fragmentation and complexity of the data. The lack of access to cohorts large enough to provide sufficient statistical power hinders the study of disease complications. To alleviate these challenges, MedGenome Inc. had developed Diabetome™ Knowledgebase, a database of >300,000 diabetes phenotypes with longitudinal clinical, biochemical, therapeutic & familial data. In addition to the curated datasets, the solution has integrated analytic and visualization tools that can enable researchers to select specific diabetic cohorts, understand the biology, predict risk of diabetes and stratify patients into responders and non-responders for drugs of interest. It can also help to understand the molecular basis for the differences in response by utilizing the biochemical data captured in the database and facilitate studying the underlying mechanisms of disease via correlation to other phenotypic information present in the database. To demonstrate stratification of cohorts based on response to oral hypoglycemic agents, we considered HbA1c levels up to 6 months before drug treatment and levels till 18 months after drug treatment, and recorded reduction of HbA1c levels by 1 % at any point, compared to pre-drug treatment HbA1c levels as responsive to the drug. We then compared all clinical variables between respondent and non-respondent groups to find out the statistical significance among different variables between the groups. To then demonstrate the utility of integration of genotype-phenotype information of patients in finding association of genetic markers for a particular phenotype, we used whole exome data of 700 subjects from selected phenotype cohorts, identified variants and annotated them using Ensembl VEP (variant effect prediction) tool based on GRCh38 human genome build and further processed into relational database tables. We then performed association analysis on the data to determine association of SNPs with phenotypes using PLINK association test a whole genome association analysis toolset, and considered the variants for association analysis that were retained after cleaning the genotype data with assigned thresholds (MAF of 0.01, HWE of 0.0001), and a Chi-square test with a p-value threshold of 0.01. We will present the new features & studies utilizing the Diabetome™ that can enable early stage pharma discovery, clinical decision making and research towards precision medicine. .

1348T

Multi-resolution association analysis for exome-wide sequencing. E. Katsevich, S. Mantravadi, M. Sesia, C. Sabatti. Department of Statistics, Stanford University, Stanford, CA., USA.

In genetic association studies, the goal is to find a set of variants associated with a given phenotype. Candidate variants discovered by association studies are often interpreted through their membership in various biologically relevant units, like genes or genetic loci. In this sense, a set of genetic variants is associated with sets of scientific hypotheses defined at multiple resolutions. It is thus desirable to provide statistical guarantees on the Type I error not only for single genetic variants, but for all resolutions at which these variants are interpreted. The Multilayer Knockoff Filter (2018) was proposed to address this problem by controlling the false discovery rate (FDR) with respect to an arbitrary set of resolutions. For example, one could find a set of variants whose FDR is controlled at both the single-variant and gene levels. In this work, we (1) make the MKF methodology more flexible by allowing a gene to be discovered even if no "responsible" variant can be pinpointed and (2) apply this extended MKF methodology to an exome-wide sequencing data set. The extension (1) is particularly important in the case of sequencing studies since most variants are rare, and thus their effect can often be detected only when aggregating multiple variants (e.g. via a burden test).

1349F

Adaptive combination of Bayes factors method as a powerful polygenic test for gene-environment interactions when external information is unavailable. W. Lin^{1,2}, C. Huang¹, Y. Liu³, S. Tsai^{4,5}, P. Kuo^{1,2}. 1) Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan; 2) Department of Public Health, College of Public Health, National Taiwan University, Taipei, Taiwan; 3) Center for Neuropsychiatric Research, National Health Research Institutes, Miaoli County, Taiwan; 4) Department of Psychiatry, Taipei Veterans General Hospital, Taipei, Taiwan; 5) Division of Psychiatry, National Yang-Ming University, Taipei, Taiwan.

The exploration of "gene-environment interactions" (G x E) is important for disease prediction and prevention. However, G x E have been difficult to identify and replicate. Single marker analysis, a commonly used approach, suffers from low statistical power due to a harsh penalty from multiple testing. To address this difficulty, the scientific community usually uses external information to construct a genetic risk score (GRS), and then tests the interaction between this GRS and an environmental factor (E). However, external genome-wide association studies (GWAS) are not always available, especially for non-Caucasian ethnicity. Although GRS is a popular polygenic approach for detecting G x E in GWAS, its performance remains unclear when there is no external information. We here develop an "adaptive combination of Bayes factors method" (ADABF) as a powerful polygenic approach for G x E when external information is unavailable, and compare it with the GRS based on marginal effects of SNPs (GRS-M) and GRS based on SNP x E interactions (GRS-I). With comprehensive simulations, we find that ADABF is generally much more powerful than GRS-M and GRS-I. ADABF is not only a powerful polygenic test for detecting G x E, but it also can pinpoint individual SNPs that interact with E. A user-friendly R code to perform the ADABF method can be downloaded from: <http://homepage.ntu.edu.tw/~linwy/ADABFGEPoly.html>. We further apply these polygenic methods to Taiwan Biobank data containing 16,555 unrelated subjects. ADABF and GRS-M identified gene x alcohol and gene x smoking interactions on blood pressure (BP). BP-increasing alleles elevate more BP in drinkers (smokers) than in nondrinkers (nonsmokers). Moreover, we identified rs79990035 x smoking interaction on BP (FDR = 1.1%) through the resampling procedure in ADABF. SNP rs79990035 is located in the *acylphosphatase 2 (ACYP2)* gene. Our result replicates the finding that ACYP2 interacts with smoking, previously reported to be related to blood pressure in the Han Chinese population. In summary, this is the first study about polygenic methods to identify G x E in GWAS when external information is unavailable. Identification of G x E in the context of GWAS has been very difficult to date. Insufficient statistical power of existing methods may be the most important factor behind the limited successes in G x E findings. With the ADABF approach, many hidden G x E are expected to be identified in the near future.

1350W

Prioritizing environment-environment interactions using clarite. N.

Palmiero¹, A. Lucas², M.D. Ritchie², I. Vazquez-Vida³, P.J.H. Jones³, M.A. Hall¹. 1) Department of Veterinary and Biomedical Sciences, The Pennsylvania State University, University Park, PA; 2) Department of Genetics, Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, PA; 3) Richardson Centre for Functional Foods and Nutraceuticals, University of Manitoba, Winnipeg, MB, Canada.

While genome-wide association studies are an established method of identifying genetic variants associated with disease, environment-wide association studies (EWAS) highlight the role of non-genetic components in complex phenotypes. However, a lack of high-throughput quality control (QC) pipelines for EWAS data lends itself to analysis plans where the data are cleaned after a first-pass analysis, which can lead to bias, or are cleaned manually, which is arduous and susceptible to user error. We offer a novel R package, CLeaning to Analysis: Reproducibility-based Interface for Traits and Exposures (CLARITE), as a way to efficiently clean non-genetic data, perform regression analysis, and visualize results on a single platform through user-guided automation. Despite a focus on EWAS, CLARITE is intended to also improve the QC process for phenotypes and clinical lab measures for a variety of downstream analyses, including phenome-wide association studies and gene-environment interaction studies. Using CLARITE, we performed and sought replication for 4 separate EWAS in the National Health and Nutrition Examination Survey (N overall discovery=8583, N overall replication=9320) for: high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), triglycerides (TG), and total cholesterol (TC) and ~350 environment variables, adjusting for sex, age, body mass index, race, socioeconomic status, statin use, and survey year. Fifty-one HDL-C, 4 LDL-C, 17 TG, and 22 TC results replicated at a Bonferroni $p < 0.05$. We further interrogated these results for complex environment-environment interactions (ExE) by prioritizing an interaction search between replicating EWAS results and all other non-correlated variables in the analysis, adjusted for covariates. We identified 11, 20, 58, and 78 ExE models that replicated at a Bonferroni likelihood ratio test (LRT) $p < 0.05$ in HDL-C, LDL-C, TG, and TC, respectively. Among the top results was an interaction between Folate (ng/mL) and Vitamin A (ug/dL) associated with HDL-C (discovery Bonferroni LRT p : 1.08e-09, replication Bonferroni LRT p : 3.08e-12), both of which have demonstrated associations with HDL-C in studies of dietary intake; however, this is the first example of an ExE interaction between these features relating to HDL-C. These findings show the potential impact of ExE in complex phenotypes as well as the utility of CLARITE in the QC pipeline and as a method of prioritizing these interactions.

1351T

Mining publicly available gene expression data in hunt for novel biomarkers for rheumatoid arthritis. D. Rychkov¹, C. Lin², M. Sirota¹. 1) Institute for Computational Health Sciences, UCSF, San Francisco, CA; 2) Stanford University, Stanford, CA.

Rheumatoid arthritis (RA) is a chronic inflammatory disease in which the body's immune system targets synovial tissues of the joints. Diagnosis and monitoring the disease progression is challenging and there is currently no biochemical test for detection of the early-stage disease. In this study, we aimed to define a Rheumatoid Arthritis meta-profile and identify biomarkers by leveraging publicly available gene expression data and machine learning approaches. We carried out a comprehensive search for publicly available microarray data in the NCBI Gene Expression Omnibus database for whole blood and synovial tissue in RA and healthy controls. For the synovium, we were able to collect 13 datasets with 312 biopsy samples (276 RA and 36 controls). For whole blood, we were able to collect 11 datasets with 2,153 samples (1,394 RA and 759 controls). Using the Significance Analysis of Microarrays (SAM) approach with the cutoff of FDR < 0.05 and abs(FC) > 1.2 we were able to identify 882 genes (502 up- and 380 down-regulated) that were significantly differentially expressed in the synovium. In blood we identified 339 (166 up- and 173 down-regulated) significantly differentially expressed genes. Leveraging the gene list enrichment analysis tool ToppGene we performed pathway analysis that confirmed the gene regulation of the immune system process and response, type I interferon signaling, and cell activation and aggregation in both innate and adaptive immune system in RA. Then we filtered out genes that cumulatively contribute to the biological variance less than 5% and applied a Variable Selection Using Random Forests (VSURF) approach and found 5 significant genes (p-value = 0.001) in both tissues in the same direction with 3 up-regulated (*TAP1*, *MMP9*, *DRAM1*) and 2 down-regulated (*DDX3Y*, *MYC*) genes. Finally, we built a Random Forest classification model trained on the synovium data. We obtained Cohen's kappa equaled to 0.61 with sensitivity 0.86 and specificity 0.91. In the final step, we validated the prediction model on the blood data, resulting kappa of 0.57 with sensitivity 0.54 and specificity 0.98. The computational analysis allowed us to perform a comprehensive in-silico search for biomarkers for RA. We found three protein coding genes with the strongest association with RA. Identification of extensive proteins secretion in blood could help in disease detection on an early-stage and make a positive impact on monitoring disease progression and patient treatment.

1352F

Typing highly divergent genes (HLA/KIR) from large-scale next-generation sequencing projects. S. Tian¹, H. Yan¹, S. McDonnell¹, W. Ding², N. Ihrke³, M. Kalmbach³, E. Klee³, S. Baheti⁴, S. Johnson¹, J. Olson⁴, J. Cerhan⁴, L. Bergsage⁵, S. Slager¹. 1) Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, MN 55905, USA; 2) Division of Hematology/Medical Oncology, Department of Internal Medicine, Mayo Clinic, Rochester, MN 55905, USA; 3) Division of Research and Education Support Systems, Department of Information Technology, Mayo Clinic, Rochester, MN 55905, USA; 4) Division of Epidemiology, Department of Health Sciences Research, Mayo Clinic, Rochester, MN 55905, USA; 5) Division of Hematology/Oncology, Department of Internal Medicine, Mayo Clinic, Phoenix, AZ 85054, USA.

Whole-genome (WGS) and whole-exome sequencing (WES) have been widely used to identify variants implicated in human disease. However, current variant discovery approaches often start with reads mapping, whose performance deteriorates in hyperpolymorphic regions, such as human leukocyte antigen (HLA) and killer immunoglobulin-like receptors (KIR) that play pivotal roles in the immune system. To date, many large-scale sequencing projects, including the 1000 Genomes, have created deep catalogues of human genetic variation, representing valuable resources for identifying genetic basis of diseases. However, limited by the bioinformatics challenges, the genetic variation of highly divergent regions, such as HLA/KIR, are not well characterized. Furthermore, the existing HLA typing algorithms only type a few HLA class I (A/B/C) and/or class II genes (*DRB1/DQA1/DQB1*), missing over 50 highly divergent genes that may also have important roles in immune diseases. To maximize the use of information from large-scale sequencing projects, we developed a bioinformatics workflow for accurate typing of HLA/KIR genes in the IPD at four-digit resolution (accuracy: 92.3%, success rate: 94.7 for class I genes with WGS). Specifically, this pipeline selects informative reads for haplotype-based assembly, optimizes the inference of allele status, and uses well-performing HLA typing algorithm to minimize typing ambiguity. We typed 39 HLA and 17 KIR genes across 981 (53,988 alleles) and 956 (21,362 alleles) Mayo Clinic Biobank samples (EUR/WGS), respectively, and 2,328 (82,492 alleles) and 2,494 (43,862 alleles) samples from 5 populations (WES) in the 1000G Project. For the majority of the HLA genes, the levels of heterozygosity are largely comparable among populations. However, marked differences exist in several genes including *TAP1* (e.g., 62.5% heterozygosity in SAS vs. 28.6% in EUR (p<0.000001, OR=4.08), suggesting obvious ethnic diversity. Given that several HLA/KIR alleles are known to be associated with multiple myeloma (MM) risk, we demonstrated the potential application of these large collections of HLA/KIR alleles from multiple ethnic groups by typing nearly 600 MM patients (WES, CoMMpass Study in MMRF). All samples were typed for HLA/KIR, generating 32,228 and 10,480 alleles, respectively. Our bioinformatics workflow for accurate typing of HLA/KIR genes expands our ability to interrogate complex genomic regions through sequencing studies for discovery of disease genes.

1353W

CRISPResso2: Characterization of repair outcomes and allele specific analysis from CRISPR nuclease and base editor genome editing. K. Clement^{1,2,3}, H. Rees^{1,4,8}, M.C. Canver^{1,2,3}, J.M. Gehrke^{2,3}, R. Farouni^{1,2,3}, J.Y. Hsu^{1,2,3}, M. Cole^{5,6,7}, D.R. Liu^{1,4,8}, J.K. Joung^{2,3}, D.E. Bauer^{5,6,7}, L. Pinello^{1,2,3}. 1) Broad Institute of MIT and Harvard, Cambridge, MA; 2) Molecular Pathology Unit, Center for Cancer Research, and Center for Computational and Integrative Biology, Massachusetts General Hospital, Charlestown, MA; 3) Department of Pathology, Harvard Medical School, Boston, MA; 4) Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA; 5) Division of Hematology/Oncology, Boston Children's Hospital; Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, MA; 6) Department of Pediatrics, Harvard Medical School, Boston, MA; 7) Harvard Stem Cell Institute, Cambridge, MA; 8) Howard Hughes Medical Institute, Harvard University, Cambridge, MA.

The field of genome editing is rapidly advancing, and technologies to modify the genome are becoming increasingly more accurate, efficient and versatile. These technologies are accelerating discovery and interpretation of functional genetic variants from GWA and eQTL studies by enabling targeted perturbations, even of a single base pair. Amplicon sequencing is the gold standard for studying editing outcomes, but the technical expertise required to analyze amplicon sequencing data is a critical barrier to correct interpretation of editing outcomes. In addition, current software toolkits are not designed for modern genome editing tools such as base editors which are capable of inducing single-nucleotide changes. Here, we introduce CRISPResso2, a comprehensive analysis toolkit that introduces five key innovations for the analysis of genome editing data: (1) Comprehensive analysis of sequencing data from base editors; (2) Allele specific quantification of heterozygous reference sequences; (3) A novel biologically-informed alignment algorithm; (4) Ultra-fast processing time; and (5) Batch mode for analyzing and comparing multiple editing experiments. We generated experimental data for six modern genome editing agents: Cas9, four C→T base editor variants, and one A→G base editor using guides targeting five well-studied genomic loci. To demonstrate the utility of CRISPResso2, we analyzed this dataset and provide an in-depth characterization of the editing properties of these agents. Using the intuitive graphics produced by analyzing multiple samples simultaneously, the editing characteristics of each agent can be distinguished and compared. When comparing the four C→T base editor variants, our plots specifically designed for base editors highlight shifted activity windows and increased editing purity resulting from modification to the deaminase enzyme. We also used CRISPResso2 to recover alleles important for fetal hemoglobin induction by analyzing targeted perturbation of key enhancer elements harboring trait-associated variants. Together, these results emphasize the utility of CRISPResso2 in analyzing amplicon sequencing data from genome editing experiments, particularly in comparing mutational profiles and properties of different editing agents including base editors and enabling accurate and intuitive analysis of coding and non-coding elements in perturbation studies. CRISPResso2 is available online at: crispresso2.pinellolab.org.

1354T

Machine learning classifiers for chronic fatigue syndrome using immune cell specific RNA-seq. P. Comella, ND. Beckmann, GE. Hoffman, EE. Schadt. Genetic and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY.

Chronic Fatigue Syndrome (CFS) is a disease resulting in extreme fatigue without any known underlying medical condition. CFS often presents in patients following a viral infection and patients may appear to have slightly impaired immune systems. With no positive diagnostic, CFS is often misunderstood and misdiagnosed. A pilot clinical study was established to give further insight into the immunological role of CFS. This pilot study included 15 CFS patients and 15 age and sex matched controls undergoing cardiopulmonary exercise testing (CPET). Blood was extracted at four time points - immediately before, 1 day after, 2 days after and 3 days after. Five immune cell types were sorted from whole blood collected from each participant, and RNA was extracted and sequenced for each cell type, as well as whole blood, at each time point. Additionally, each participant underwent extensive clinical surveys and questionnaires. Our team has employed state of the art machine learning approaches to construct classifiers that can accurately distinguish CFS from non-disease states in immune cells using RNA-Seq data. This technique has generated non-linear gene associations that can be used to understand distinct gene expression differences between disease and non-disease states. We were able to identify B cell and Monocyte classifiers as being the highest performing classifiers in the immune cell populations, suggesting these cell types contain more interesting disease-state biology. Our classifier pipeline further illustrates the stability of the features/genes of the classifier, resulting in specific genes that are consistently chosen as important features to classify disease with. These classifier algorithms can also be used to predict unseen observations as a future diagnostic tool.

1355F**A computational approach to identify novel disease-causing genes**

from exome sequencing. S. Rana¹, U. Sunderam¹, K. Kundu¹, Y. Zou², A.N. Adhikari², A.G. Sharo², J.M. Puck³, S.E. Brenner², R. Srinivasan¹. 1) Innovation Labs, Tata Consultancy Services, Hyderabad, India; 2) University of California, Berkeley, CA 94720, USA; 3) Department of Pediatrics, University of California San Francisco, San Francisco, Box 3118, CA, USA; 4) Institute for Bioscience and Biotechnology Research, University of Maryland, Rockville, MD, USA; 5) School of Life Sciences, Fudan University, Shanghai, China.

Whole exome sequencing has advanced discovery of causal variants underlying rare genetic disorders. Many rare Mendelian diseases, such as severe combined immunodeficiency (SCID) are life threatening without treatment, yet can be effectively treated if diagnosed early. Whole exome sequencing can facilitate diagnosis by identifying the disease-causing variants. However, detecting the causative genomic variants is challenging, due to a number of difficulties, including variant prioritization, pathogenicity prediction, and limited knowledge of gene function and pathways involved in disease. To address this challenge, we have developed a variant prioritization protocol that is optimized for trio-exome sequencing of primary immunodeficiency cases. To yield high quality sets of possible causative variants, we used multi-sample calling and integrated variant annotation, variant filtering, and gene prioritization. We first focus on candidate variants from genes previously known to be associated with the disease. If there are no compelling findings in these genes, we expand the disease gene list by integrating information from patient phenotypes and gene pathways. This expanded list includes genes extracted from Human Phenotype Ontology (HPO), Mouse Phenotype Ontology (MPO) and various protein interaction databases. To prioritize important genes, we score each gene based on its associated phenotypes. We categorize patient phenotypes into 1) primary phenotypes, relevant to patient diagnosis and indicative of a functional defect, and 2) secondary phenotypes, derivative consequences of the primary functional defects. The genes associated with a patient's primary phenotype are given a higher score than the genes associated with the patient's secondary phenotype or generic HPO and MPO ontology terms. Our protocol has revealed causative variants in SCID and other immunodeficiencies including early diagnosis of Ataxia Telangiectasia, Nijmegen Breakage Syndrome, as well as several novel syndromes. These cases highlight unique features of the analysis framework that facilitate discovery of genetic variants leading to disease. These findings provide crucial information for avoidance of a diagnostic odyssey, to inform prompt and appropriate treatment and to guide family genetic counseling.

1356W**An application of *ChARMDiff* to discover epigenetic difference between CD4 memory and naive primary cells.** S. Park, S. Kim. Soongsil University, Seoul, South Korea.

Various chromatin modifications, identified in large-scale epigenomic analyses, are associated with distinct phenotypes of different cells and disease phases. To improve our understanding of these variations, we have developed *ChARMDiff*, a new computational approach based on association rule mining, which is pattern discovery of *de novo* differential chromatin modifications and characterize biological significance of globally occurred patterns for combinatorial chromatin state difference between multiple cell types and conditions. We report an application of *ChARMDiff* to two pairs of epigenomes in CD4 memory and naive primary cells from roadmap epigenomics project. *ChARMDiff* provides a scalable framework that can easily be applied to find various levels of combination patterns, which should reflect a range of globally common to locally rare chromatin modifications. Our approach provides new insights into the resolution of the histone code hypothesis to characterize epigenetic variations in distinct phenotypes of different cells, disease phase and experimental conditions.

1357T

The NIAID clinical genomics data repository, Genomic Research Integration System (GRIS): Streamlined data capture, standardization, integration, and analysis to enable identification of immune-related genetic variants for clinical reporting and genomics research. S. Xirasagar¹, A.J. Oler¹, Z. Li¹, X. Cheng¹, K. Huang¹, D. Veltri¹, M. Similuk², Y. Zhang², C. Hong², M. Walkiewicz², J. Johnson², J. Barnett¹, L. Zhang¹, Y. Fan¹, O. Juarez-Espinoza¹, M. Lenardo², J.D. Milner². 1) Office of Cyber Infrastructure and Computational Biology (OCICB), NIAID, National Institutes of Health, Rockville, MD; 2) NIAID Clinical Genomics Program, Division of Intramural Research, NIAID, National Institutes of Health, Bethesda, MD.

The NIAID Genomic Research Integration System (GRIS), a web-based collaboration-engendering web application facilitates identification of causal genetic variants of rare and/or novel immunological disorders. Users can access integrated and standardized phenotypic and genomic data that is analyzable via powerful genetic analysis tools. Systematic and automated capturing and linking of patient clinical and genomic data from disconnected systems and paper-based records accompanied by standardized annotations allow cross comparisons across data from different clinical studies. Multiple levels of access coupled with secure data storage and transfer facilitates quick access to relevant data for all aspects of the clinical study lifecycle. Data access and sharing workflows are optimized to promote maximum data usage while protecting patient privacy and confidentiality. The use of standardized vocabularies allows researchers to efficiently query and retrieve data from all the studies in GRIS, thus increasing the analysis power and making discoveries otherwise not possible from individual study datasets. The system leverages existing open source tools, ontologies, and genomic annotation databases. A customized version of PhenoTips®, with a user-friendly interface has been implemented to enable capture, searching, and display of patient annotations standardized with controlled vocabularies including the Human Phenotype Ontology (HPO) and Online Mendelian Inheritance in Man® (OMIM). An intuitive pedigree editor allows documentation, display and reporting of pedigrees for publication or in machine-readable formats for use within genetic analysis tools. The genetic variant analysis tool, seqr, has been integrated in conjunction with standardized analysis pipelines based on open source tools (BWA, PICARD, GATK/Queue, BCFtools, VEP) for sequence alignment, variant data processing, quality control, and annotation. seqr enables Mendelian and custom inheritance variant analysis at individual, family, or family group levels. Advanced filtering capabilities and high-quality genomic annotation databases including HGMD, ClinVar and OMIM facilitate meaningful biological interpretation. Importantly, GRIS enables both standardized variant analyses for clinical reporting and flexible analyses for research analyses. GRIS is thus a scalable platform to accelerate the pace of molecular diagnosis and treatment of diseases.

1358F

BREATH: An open-access database of normal lung development generated by the molecular atlas of lung development program (LungMAP). R.F. Clark¹, N.C. Gaddis¹, J.L. Levy¹, M. Duparc¹, M.C. Krzyzanowski¹, C. Ball¹, S.M. Palmer², the LungMAP Data Coordinating Center. 1) RTI International, Research Triangle Park, NC; 2) Duke Clinical Research Institute, Duke University School of Medicine, Durham, NC.

The lung is a complex organ with high cellular heterogeneity, and research is needed to define the interactive gene networks and dynamic crosstalk among multiple cell types that coordinate normal lung development. Significant knowledge gaps exist in the understanding of lung development from late fetal to perinatal stages and through early childhood, a critical period when the diverse lung cells go through terminal differentiation and maturation and when alveoli form. Coordination of information from disparate assays and experimental approaches is critically needed to help the pulmonary research community uncover the molecular determinants of lung development. The goal of LungMAP is to build an open-access resource providing a comprehensive molecular atlas of late-stage lung development in humans and mice, making otherwise dispersed data, reagents, and protocols freely available to the research community. We have created the Bioinformatics Resource Atlas for the Healthy lung (BREATH) database, applying novel data management and bioinformatics approaches to manage high-throughput multidimensional experimental data. BREATH is built on a triplestore database backbone, integrated with anatomical ontologies for lung development, and provides access to novel web-based tools for the analysis and visualization of data generated by the four Research Centers and Human Tissue Core of the LungMAP Consortium. The current version of BREATH contains confocal immunofluorescence images, *in situ* hybridization images, histological images, nano-DESI, and 3D Vibra-SSIM confocal and uCT images of developing mouse and human lungs at several time points, as well as single-cell RNAseq, multi-cell RNAseq, microRNA, MeDIP-seq, proteomic, lipidomic, and metabolomic data from mouse and human lung cells. The LungMAP website (www.lungmap.net) provides an entry portal to the BREATH database and tools for exploring and interacting with the lung images and omics data. A better understanding of the basic molecular pathways that regulate normal lung development will enable development of innovative therapies that advance treatment of lung injury repair and regeneration.

1359W

Deep recurrent neural networks for accurate variant calling in 21-hydroxylase-deficient congenital adrenal hyperplasia. A. Ganesh^{1,2}, S. Hong¹, D. Muzzey¹, K.R. Haas¹, D. Petcolas¹, K.A. Beauchamp¹. 1) Counsyl, South San Francisco, CA; 2) Stanford, CA.

Introduction Congenital adrenal hyperplasia (CAH) is an autosomal recessive disease that impairs steroidogenesis. CAH affects approximately 1 in 10,000 persons but is often excluded from expanded carrier screening (ECS) due to variant-calling challenges posed by high homology and frequent gene rearrangements between the *CYP21A2* gene and its pseudogene *CYP21A1P*. We previously described—and have used extensively in a clinical setting—a molecular and computational strategy that identifies 11 common CAH-causing variants but requires nuanced manual review. Here we show that a deep learning model trained on human-annotated samples calls CAH variants with high sensitivity and specificity, critically lessening the burden of human review. Methods An anonymized cohort of 118,476 ECS patients was split into training and validation sets of 94,781 and 23,695 patients, respectively. For each patient, we aggregated two classes of next generation sequencing features. First, we collected allele-balance measurements summarizing the fraction and count of genic and pseudogenic alleles of interest. Second, we obtained copy-number measurements consisting of normalized sequencing depths at important loci on *CYP21A2* and *CYP21A1P*. The observations for training were the human-reviewed (and possibly human-relabelled) output of our classical variant-calling algorithm, stored as 11 binary genotypes per patient. These data were used as input to a five-layer recurrent neural network comprised of long short-term memory (LSTM) cells and a weighted cross entropy loss function. The model was implemented in TensorFlow and trained using the Adam optimizer. Results Considering the human-reviewed calls as ground truth, the LSTM showed a validation error rate of approximately 1 in 1000. For example, the 30kb deletion variant was called with 99.58% sensitivity (1,406 true positives / 1,412 total positives), and 99.97% specificity (22,276 true negatives / 22,283 total negatives). We also evaluated performance on a held-out subset of 243 difficult calls for which human relabeling had been necessary. Model accuracy was 95% on this challenging set, suggesting that the LSTM model is capturing domain knowledge of expert human reviewers. Conclusions Deep recurrent neural networks achieved high accuracy for a technically challenging CAH variant calling task. When large training datasets are available, deep learning can streamline complex, human-aided processes in clinical sequencing workflows.

1360T

Predicting the time-to-event outcome of cystic fibrosis related diabetes using genotype data. Y. Lin¹, L.J. Strug^{1,2}. 1) Department of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada; 2) Genetics and Genome Biology, The Hospital for Sick Children, Toronto, ON, Canada.

Cystic Fibrosis (CF) is a life-limiting genetic disease caused by mutations in the Cystic Fibrosis Transmembrane Conductance Regulator (*CFTR*). CF affects multiple organs including the exocrine pancreas which contributes to Cystic Fibrosis Related Diabetes (CFRD), seen in over 40% of patients by age 40 with an estimated incidence of 3.4% per year. Untreated CFRD is associated with increased CF-related mortality whereas early detection can improve outcomes. Given the variation in age-of-onset of CFRD, we aim to develop a predictive model to identify individuals at risk of CFRD at different ages based on genetic and clinical measures early in life. Early-age clinical measurements that are robust CFRD biomarkers but prone to missingness further complicate model generation. Using 2,347 individuals participating in the Canadian CF Gene Modifier Study, we separate 1,600 and 747 participants into training and test sets, respectively. Regression models are used to estimate early-age biomarkers including Immunoreactive Trypsinogen (IRT) at birth in individuals for whom these measures are not available. Initial model generation has used penalized Cox Proportional Hazard models with 3,984 SNPs annotated to previously associated gene sets and early clinical variables including gender, estimated IRT and Meconium Ileus (MI). Component-wise gradient boosting and LASSO were employed for model-fitting, and a 10-fold cross validation procedure was used to select the parameters. Area Under the Curve (AUC) for time-dependent ROC curves were computed between ages 8 and 40, with emphasis on improved CFRD prediction at younger ages. The final predictive model included MI, IRT and 22 SNPs including 4 *CFTR* variants. Adequate predictive accuracies are achieved when assessing risks at older ages (AUC=0.708, Age=40); however, improved performance is required for predictive risk at younger ages (AUC=0.608, Age=12). Ongoing work involves the inclusion of genome-wide genotype data and longitudinal BMI measurements to improve predictive accuracy among younger CF patients. Moreover, significant variability in the final model when conducting variable selection on high-dimensional data highlights the limitation of AUC and other discrimination measures alone for validation. Future work will incorporate calibration and methods with optimized statistical properties into model generation.

1361F

Development of the BiomedScape data-mining suite for deep semantic structure-based drug repurposing for DMD: A pilot feasibility study. *J.W. Ulm, S.F. Nelson.* UCLA Medical Center, Department of Genetics, Los Angeles, CA., Select a Country.

A text-mining-based pilot feasibility study was initiated to focus on identification and quantification of deep semantic structures (DSS's) so as to determine and amplify a telltale "information scent" linking DSS's of disease-related literature (for e.g. CF or DMD) with those of literature pertinent to potential drug repurposing candidates, derived from published abstracts or comprehensive pharmaceutical databases such as Drugbank. To accomplish this, it was necessary to first create a large-scale "informatic infrastructure" to provide the tens of millions of documents in the form of lists of strings or nested lists—within the biomedical field but unrelated to the specific topic of interest—to train the AI-like text-mining models. This was accomplished through development of the comprehensive BiomedScape tool, written chiefly in the Python programming language, which has provided these document corpora (semantic spaces) for the first time, in the form of carefully delimited abstracts scraped from 1911 that can be randomly recombined into training corpora, alongside dozens of modular code blocks facilitating rapid and highly versatile text-mining studies. Subsequent application of the tool has provided indications of a DSS-based information scent that has partially recapitulated previous serendipitous drug repurposing findings, with the scent amplified by a number of data-mining parameters such as a specific focus on drug and disease molecular mechanism reports, chronological narrowing, and word bigram or trigram comparisons versus traditional bag-of-words. These studies have in turn produced preliminary hints of potential drug repurposing candidates for DMD, though as yet no test statistic or index that can be used to firmly predict a single high-probability candidate. As the BiomedScape tool is still early in development, future efforts will focus on elucidating indices and summary statistics that can more robustly identify a "smoking gun" that decisively suggests a repurposing candidate indicated by previous findings but otherwise overlooked.

1362W

A 29-gene signature predicts progression to active tuberculosis in exposed household contacts. *Y. Zhao^{1,2}, S. Leong³, W.E. Johnson^{1,2}, P. Salgame².* 1) Section of Computational Biomedicine, Boston University School of Medicine, Boston, MA; 2) Bioinformatics Program, Boston University, Boston, MA; 3) Rutgers New Jersey Medical School, Newark, NJ.

Background: One-third of the world's population is latently infected with *Mycobacterium tuberculosis* (Mtb). Clinical management of latent tuberculosis (TB) infection (LTBI) remains a challenge due to lack of diagnostic biomarkers that can predict which individuals will progress to active TB disease. **Methods:** RNA-sequencing was performed on peripheral blood mononuclear cell samples from a prospective Brazilian household contact (HHC) cohort consisting of individuals with LTBI that either progressed to TB disease (progressors) or not during long-term follow-up (nonprogressors). We used the RNA-Seq data from GSE79362 dataset to train a baseline biomarker, using an ensemble feature selection pipeline, and then validated this biomarker on the RNA-seq data from our Brazilian cohort. **Results:** The published 16-gene correlate of risk signature derived from an adolescent cohort (ACS-COR) performed poorly in predicting progressors from non-progressors in the Brazilian dataset. Since the performance of the ACS-COR signature improved closer to the time of TB diagnosis, we used the earliest time-point samples from progressors and non-progressors from the African dataset to derive a novel 29-gene signature that was significantly better at predicting progressors from non-progressors in the Brazilian dataset with a sensitivity of 74.2% and a specificity of 84.8%. **Conclusions:** PREDICT29 signature offers strong predictive performance in distinguishing nonprogressors from progressors at early time-points, several years prior to TB diagnosis, and across geographical regions and despite ethnically distinct host and bacterial genetics. Therefore, this signature is a potential tool for risk stratification of infected individuals for targeted anti-TB preventive therapy.

1363T

From GWAS to the cross-phenotype neighborhood of diseases in the interactome: Network-based pathway prioritization identifies concordant and discordant pathways between COPD and IPF. A. Halu^{1,2}, S. Liu^{1,3}, S.H. Baek¹, B.D. Hobbs^{1,4}, M.H. Cho^{1,4}, E.K. Silverman^{1,4}, A. Sharma¹. 1) Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 2) Center for Interdisciplinary Cardiovascular Sciences, Cardiovascular Division, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 3) Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN; 4) Division of Pulmonary and Critical Care, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA.

Chronic obstructive pulmonary disease (COPD) and idiopathic pulmonary fibrosis (IPF) are two pathologically distinct diseases that result from similar exposures, show common symptoms, and have high comorbidity. Genome-wide association studies (GWAS) have identified shared loci for COPD and IPF, including ones with opposite directions of effect. The existence of further such loci, as well as shared mechanisms between the two diseases at the molecular level remains to be explored. Taking a network-based approach, we studied the shared molecular neighborhood of COPD and IPF in the interactome, with the aim of discovering genes and pathways involved in both diseases. We used GWAS- and literature-derived genes as seeds to prioritize disease-associated candidates by a random walk-based method that corrects for degree bias. We built disease modules using an approach that maximizes connectivity significance. We validated the biological relevance of the resulting COPD and IPF disease modules using gene expression data from lung tissue, where the module genes displayed strong disease signals at the transcriptional level. The two disease modules showed statistically significant overlap and network proximity, sharing 19 genes. To uncover pathways at the intersection of COPD and IPF, we developed a metric, NetPathScore, which prioritizes the pathways of a disease by their shared interactome neighborhood with another disease. Applying NetPathScore to the pathways enriched in the up- and down-regulated portions of diseases modules enabled the determination of concordant and discordant pathways between the COPD and IPF. These concordant and discordant pathways offer important clues about the pathobiology of the two diseases.

1364F

An application of genetic algorithm based Naive Bayesian classifier in metabolomics profiling. K. Su^{1,2}, Q. Zhao³, H. Deng^{1,2}. 1) Department of Global Biostatistics and Data Science, Tulane University, New Orleans, LA; 2) Center for Bioinformatics and Genomics, Tulane University, New Orleans, LA; 3) Department of Preventive Medicine, The University of Tennessee Health Science Center, Memphis, TN.

Metabolomics is an analysis of small molecules and biochemical intermediates applied on various phenotypes. Although metabolomics may explain only a small proportion of the total phenotypic variability on its own, metabolites have been widely used as biomarkers to identify the onset of many complex diseases. The present work demonstrated a hybrid approach combining genetic algorithm (GA) and Naive Bayesian (NB) method to identify the potential osteoporosis-associated metabolites. We performed a liquid chromatography-mass spectrometry (LC-MS) untargeted metabolomics analysis on 136 serum samples collected from Louisiana Osteoporosis Study. The known metabolites from LC-MS were included in a classical GA which used three main properties of rules ("selection," "crossover," and "mutation") to conduct multiple Random Forests via the evolutionary process to identify the potential osteoporosis-associated metabolites. The top five GA-selected metabolites were Stachydrine, Threonine, Pipecolate/L-Pipecolic Acid, Urate, and Sarcosine/Beta-Alanine. We then applied the NB classifier to the GA-selected metabolites and found that the NB model considering the top 15 GA-selected metabolites had the highest prediction accuracy rate of 73%. It is noteworthy that 4 of the top 5 GA-selected metabolites have previously been reported to have an association with osteoporosis. We conclude that the GA-selected metabolites considered in the NB classifier with highest accuracy rate may potentially be reliable biomarkers for assisting identification of osteoporosis, although further functional validation experiments are needed. Overall, this study shows that GA can efficiently detect osteoporosis-related metabolites with a limited number of subjects.

1365W

Integrating regulatory features for large-scale transcriptome-wide association study from GWAS summary statistics identified novel genes for osteoporosis. S. Yao, S.S. Dong, Y.J. Zhang, J.B. Chen, Y. Guo, T.L. Yang. Key Laboratory of Biomedical Information Engineering of Ministry of Education, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, China.

Genome-wide association studies (GWASs) have been demonstrated as an effective strategy to identify susceptibility loci for human complex traits. However, the biological mechanisms between genetic variations and traits are not fully understood. Previous studies have demonstrated that many genetic variants could influence complex traits by regulating gene expressions. Here we introduce a strategy to identify novel susceptibility genes for human complex traits based on the interpretation of regulatory features, genotype, gene expression and phenotype. We estimated gene expression effect sizes based on classified single nucleotide polymorphisms (SNPs) with functional features as a prior, and then estimated the associations between predicted expressions and traits with the combination of SNP-trait effect sizes while accounting for linkage disequilibrium among SNPs. We applied our strategy in GWAS summary data of bone mineral density with 185,817 individuals from the Genetic Factors for Osteoporosis (GEFOS) Consortium with expression data from blood across 696 individuals. Finally, we identified 130 bone mineral density-associated genes after multiple test correction ($P < 5 \times 10^{-6}$) after excluding the major histocompatibility complex region. The most significant gene, *FAM3C*, have been reported to be associated with a genome-wide significant increased risk of forearm fracture. Among these identified genes, 104 genes were located near at least one GWAS SNP (± 500 kb), and the remaining 26 genes implicated novel targets. Seven of the new identified genes are associated with relevant phenotypes in the Mouse Genome Informatics, which are *FKBP8*, *EPN1*, *PPARG*, *CERS1*, *ATP2B2*, *IDUA*, and *PNPLA1*. Especially, *IDUA* locus was identified to be associated with bone mineral density, and osteoblast cell number in a mouse knockout of *Idua*. Our results revealed the power of integrating regulatory features, genotype, gene expression and phenotype to identify novel genes for human complex traits.

1366T

Integrative analysis of clinical and molecular data on a well-phenotyped lupus cohort identifies distinct molecular subtypes. I. Paranjpe¹, C. Lanata¹, J. Nititham¹, K. Taylor¹, M. Gianfrancesco¹, D. Rychkov¹, L. Barcellos², P. Katz¹, M. Dall'Era¹, J. Yazdany¹, L. Criswell¹, M. Sirota¹. 1) University of California, San Francisco, San Francisco, CA; 2) University of California, Berkeley, Berkeley, CA.

Systemic Lupus Erythematosus (SLE) is a chronic autoimmune disease with heterogeneous manifestations. We aimed to define how molecular differences underlie this clinical heterogeneity through an integrative approach leveraging methylation, genetic, and phenotypic data from a well characterized multi-ethnic cohort of 330 SLE patients. We first stratified patients based on the 18 American College of Rheumatology (ACR) criteria which include hematologic manifestations, serum autoantibody status, cutaneous rash, and neurologic manifestations. Unsupervised K-means clustering on these 18 input features revealed 3 stable clusters, with stability evaluated by a bootstrap resampling method. We found 2 severe clusters, one characterized by serum autoantibodies and lupus nephritis (LN), and the other by serum autoantibodies, LN as well as hematologic manifestations and 1 mild cluster characterized by malar rash, arthritis, photosensitivity, and oral ulcers. Interestingly, these clusters reflect known clinical associations with race, with the mild cluster enriched for White patients and the severe clusters for non-White. We then identified individual CpGs associated with clinical clusters adjusted for age, sex, genetic ancestry, and cell composition. This analysis revealed 241 cluster-associated CpGs (FDR<0.1) with significant enrichment of genes relating to Type 1 Interferon signaling and IFN-gamma. Overall Type 1 interferon genes were hyper-methylated in the mild cluster compared to the severe clusters, suggesting a role for epigenetic control of this pathway in SLE outcome. By mapping CpGs to regulatory regions defined by the Epigenome Roadmap 15 chromatin state model, we observed significant enrichment (hypergeometric test $q < 0.01$; odds ratio > 2) of cluster-associated CpGs in enhancers and transcriptional start sites in peripheral T cells, B cells, monocytes, neutrophils, and NK cells, indicating DNA methylation mediated epigenetic remodeling in peripheral blood cells may contribute to the heterogeneity observed in SLE. MeQTL analysis revealed that from the 241 cluster-associated CpGs, 38 were under cis genetic control (FDR<0.01) and 24 under trans control (FDR<0.1), suggesting a functional link between genetic variation and methylation in the Type 1 Interferon pathway. In conclusion, by applying an integrative approach, we reveal several clinically relevant SLE subtypes characterized by multiple phenotypes and supported by a strong epigenetic and genetic signature.

1367F

Development and validation of algorithms using electronic health records to identify azathioprine use and azathioprine-induced leukopenia for pharmacogenetic studies. A. Dickson, B. Stanley, P. Anandi, V. Kawai, K. Birdwell, A. Hung, C. Stein, Q. Feng, C. Chung. Vanderbilt University Medical Center, Nashville, TN.

Background: Use of the immunosuppressant azathioprine (AZA) presents significant challenges: a narrow therapeutic index, marked inter-individual variability, and limited ability to predict frequent side effects. A novel approach to improve prediction of drug response is through the linkage of electronic health records (EHRs) with biobanks in pharmacogenetic studies. Because medical records review may be a limiting factor in these studies' feasibility, we developed and validated algorithms first to identify AZA users and then to determine cases of AZA-induced leukopenia, one of AZA's most serious side effects. **Methods:** Using a database of de-identified EHRs at Vanderbilt University Medical Center, we found potential users of AZA by mining structured (e.g., prescriptions) and unstructured (e.g., clinical notes) data for any mention of AZA or its brand names. Our first algorithm identified periods of current AZA use as 90-day exposure windows following an indicated use of AZA with at least one white blood cell count (WBC) during the exposure window. We estimated the algorithm's ability to predict AZA use through medical records review of a statistically significant random sample of 134 potential AZA users. Our second algorithm defined cases of AZA-induced leukopenia as at least one episode of WBC <4,000/ μ l within those 90-day exposure windows. We estimated the algorithm's ability to detect leukopenia occurring during AZA exposure in 750 cases and controls (375 each), previously confirmed by medical records review. **Results:** From the 6,625 potential users of AZA we initially found, the first algorithm identified 3,212 as AZA users and excluded 3,413 as non-users or for limited data. Overall accuracy for determining AZA users was 89.55%, with a PPV of 81.54% and a NPV of 97.10% among potential users. Among the 3,212 identified users, the second algorithm categorized 898 as cases of leukopenia during AZA exposure and 2,314 as controls. The agreement between medical records review and the second algorithm's determination of cases and controls was 94.67%, with a Cohen's kappa coefficient of 89.56%. **Conclusion:** Computerized definitions of AZA use and AZA-induced leukopenia, linked with genetic data, can facilitate pharmacogenetic studies of this commonly prescribed immunosuppressant. Its adaptability offers potential for studying multiple side effects of this and other immunosuppressants.

1368W

Computational prediction and molecular validation of novel therapeutic targets for potent splicing modulators. D. Gao^{1,2,3}, E. Morini^{1,2}, M. Salani^{1,2}, C. Montgomery^{1,2}, S. Erdin^{1,3}, M.E. Talkowski^{1,2,3,4}, S.A. Slaugenhaupt^{1,2}. 1) Center for Genomic Medicine, Massachusetts General Hospital, 185 Cambridge St, Boston, MA 02114, USA; 2) Department of Neurology, Harvard Medical School, 25 Shattuck St, Boston, MA 02115, USA; 3) Program in Medical and Population Genetics and Stanley Center for Psychiatric Research, Broad Institute, Cambridge, MA 02142, USA; 4) Departments of Psychiatry and Pathology, Massachusetts General Hospital, Boston, MA 02114, USA.

While Mendelian genetics has seen remarkable advances in recent years in the discovery of novel genes and pathogenic mechanisms, therapeutic development remains incremental. One strategy for speeding development is to discover novel targets for known compounds. Here we applied a combination of *in silico* and molecular approaches to identify targets of splicing modulator compounds (SMCs), which target splicing in various diseases. We treated six independent human fibroblast cell lines from healthy individuals with BPN-15477, one of the SMC compounds originally designed to target *ELP1* (*IKBKAP*) splicing in familial dysautonomia. We investigated splicing using RNA-seq by assessing all plausible exon 'triplets' (three consecutive exons) transcriptome-wide. Percentage-of-spliced-in (ψ) analysis of the middle exon from each triplet revealed 73 exon inclusion (EI) and 130 exon skipping (ES) outcomes in treated vs. untreated cells. We trained gradient boosting machines for EI and ES respectively, to learn sequence features that correlate with treatment consequences and to predict transcriptome-wide splicing influence of the drug. Our EI machine with sixteen features achieved an ROC of 0.89 while our ES machine with four features gave an ROC of 0.68. Applying the two machines across the transcriptome enabled us to predict potential gene targets for the drug. In 11,150 triplets (3.0% of 330,523 total), exon inclusion is predicted after treatment while skipping was predicted in 9,244 triplets (2.8%). Next, we explored the potential applications of the drug under two functional models: 1) That the EI effect of the treatment can ameliorate the impact of exon skipping caused by pathogenic mutations in human disease and 2) That ES might rescue gene expression in diseases where mutations create premature termination codons that result in nonsense-mediated decay. By comparing the overlap of these predicted triplet candidates with causal mutations in ClinVar, we found that 61 triplets were associated with mutations that could be rescued by EI (e.g. cystic fibrosis, Lynch syndrome and phenylketonuria) and 66 triplets associated with mutations rescuable by ES (e.g. Duchenne muscular dystrophy, retinoblastoma and Gorlin syndrome). While confirmation of these predictions using *in vitro* models is ongoing, our method demonstrates the potential impact of coupling machine learning strategies with transcriptional signatures to identify novel targets for known splice modulator drugs.

1369T

Meta-analysis of gene expression variability in Alzheimer's disease.

L.R.K. Brooks^{1,3}, G.I. Mias^{2,3}. 1) Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI; 2) Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI; 3) Institute for Quantitative Health Science and Engineering, Michigan State University, East Lansing, MI.

The Centers for Disease Control and Prevention rank Alzheimer's disease (AD) as the 6th leading cause of death within the United States (US). AD is a significant health-care burden, given its increased occurrence (particularly in the aging US population), ineffective treatments and a lack of preventive methods. AD and other dementias, will account for an estimated \$277 billion US society costs in 2018. AD targets neuronal function and can cause neuronal loss due to the buildup of amyloid-beta plaques and intracellular neurofibrillary tangles. Increased risk of developing AD has been previously associated with aging. However, it is still unclear how to differentiate between normal- and AD- age-related memory loss (and associated brain degeneration). We will present our investigation of temporal changes within gene expression using a meta-analysis of AD and healthy cohort gene expression, utilizing publicly available microarray expression datasets, that we collected and curated from the Gene Expression Omnibus (GEO) database by National Center for Biotechnology (NCBI). We utilized any datasets that reported donor age and gender, irrespective of platform (8 datasets, 2,089 samples). We preprocessed the microarray raw data, including background correction, normalization across arrays and correction of batch effects. To identify temporal trends in the data across ages, brain region and varying tissue, and disease state we performed a linear model analysis, including age as a linear time series. Statistically significant data ($p < 0.05$) were further analyzed using our MathlOmica framework to identify overrepresented Reactome pathways and gene ontology terms. Our results include gene expression differences (2,309 genes) between disease and healthy individuals, and also with respect to varying age. We identified differences in more than 300 genes associated to the Immune System, as well as 29 genes related to Neuronal System and transmission across Chemical Synapses, and pathways such as Regulation of TP53 Expression. Our approach allowed us to effectively combine multiple microarray datasets and identify gene expression differences between AD and healthy individuals including age considerations. Our findings provide potential gene and pathway associations that can be targeted to improve AD diagnostics and potentially treatment or prevention.

1370F

VCPA: Genomic variant calling pipeline and data management tool for Alzheimer's disease sequencing project.

Y. Chou¹, Y.Y. Leung¹, O. Val-ladares¹, H.-J. Lin¹, A.B. Kuzma¹, L. Cantwell¹, L. Qu¹, P. Gangadharan¹, A.D. Sequencing Project (ADSP)¹, W.J. Salerno², G.D. Schellenberg¹, L.-S. Wang¹.

1) Penn Neurodegeneration Genomics Center, Department of Pathology and Laboratory Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX.

The Alzheimer's Disease Sequencing Project (ADSP) is an integral component of the National Alzheimer's Project Act (NAPA) towards a cure of Alzheimer's Disease (AD). ADSP will eventually analyze whole-genome sequencing (WGS) and whole-exome sequencing (WES) data from >20,000 late-onset AD patients and cognitively normal elderly to find new genetic variants associated with disease risk. To ensure all sequencing data are processed consistently and efficiently according to best practices, a common workflow called "Variant Calling Pipeline and data management tool" (VCPA) was developed by the Genome Center for Alzheimer's Disease (GCAD) in collaboration with ADSP. VCPA 1) is optimized for large-scale production of WGS/WES data, 2) includes a tracking database with web frontend for users to track production process and review quality metrics; 3) is implemented using the Workflow Description Language for better deployment and maintenance, 4) is designed for the latest human reference genome build (GRCh38/hg38, version GRCh38DH) and follows best practices for WGS analysis with input from TOPMed (Trans-Omics for Precision Medicine) and CCDG (Centers for Common Disease Genomics). VCPA consists of two independent but interoperable components: a tracking database with a web frontend and a SNP/indel calling pipeline. The pipeline (available as an Amazon Machine Images (AMI)) was optimized for automatic processing WGS/WES data in various file formats (either WGS/WES pair-end reads in FASTQ, BAM, or CRAM (compressed BAM) with flow cell information), from mapping sequence reads to the latest human reference genome (GRCh38) and variant calling. The tracking database (available as another AMI) was designed for monitoring the job status and recording quality metrics for each processed sample. With a dynamic web interface of the database, researchers can easily compare, share and visualize all these individual level quality metrics. The workflow is modularized and consists of four stages. Users can configure the workflow to skip individual stages to reduce time and cost. We evaluated our pipeline using the NA12878 sample from the Genome In A Bottle project using the hg38 high confident set. Sensitivity/precision of VCPA calls were 0.999/0.994 for SNPs and 0.985/0.987 for indels respectively. The pipeline source code and step-by-step instructions are available from the National Institute on Aging Genetics of Alzheimer's Disease Data Storage Site (<http://www.niagads.org/VCPA>).

1371W

NIA Genetics of Alzheimer's Disease Data Storage Site (NIAGADS)

Alzheimer's genomics database: 2018 update. E. Greenfest-Allen^{2,4}, P. Gangadharan^{1,2}, A. Kuzma^{1,2}, L. Qu^{1,2}, O. Valladares^{1,2}, Y.Y. Leung^{1,2,3}, W. Bush⁵, A. Naj⁶, G.D. Schellenberg^{1,3}, C.J. Stoeckert, Jr.^{2,4}, L. Wang^{1,2,3}. 1) Penn Neurodegeneration Genomics Center, Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; 2) Institute for Biomedical Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; 3) Institute on Aging, University of Pennsylvania, Philadelphia Perelman School of Medicine, PA; 4) Department of Genetics, University of Pennsylvania, Philadelphia Perelman School of Medicine, PA; 5) Department of Epidemiology, University of Pennsylvania, Philadelphia Perelman School of Medicine, PA; 6) Institute for Computational Biology, Department of Population and Quantitative Health Services, Case Western Reserve University, Cleveland OH.

Statement of Purpose: NIAGADS is a national genetics data repository for genotypic data concerning the study of late-onset Alzheimer's disease (AD) and related neuropathologies. The Alzheimer's Genomics Database (DB) is a searchable annotation resource provided by NIAGADS that allows researchers to interactively explore, interpret, and analyze public datasets from the NIAGADS repository, including both published GWAS summary statistics and AD-specific variants and annotations compiled by the Alzheimer's Disease Sequencing Project (ADSP). **Methods:** The NIAGADS Alzheimer's Genomics DB (<https://www.niagads.org/genomics>) makes available to the public 43 GWAS summary statistics datasets from 11 NIAGADS accessions for interactive exploration and data-mining. The GWAS results are linked to dbSNP variants, NHGRI GWAS Catalog annotations, and allele frequencies and linkage disequilibrium inferred from 1000 Genomes data, as well as variant annotations compiled by ADSP, including predicted variant effect and effect impacts, transcription factor occupancy (CATO), deleteriousness (CADD), and ExAC allele frequencies. Using the website's graphical search interface, users can perform sophisticated data-mining of the NIAGADS datasets in real time and use analysis tools (e.g., co-location) to perform data transformations or find overlap between variants and genes lists, functional genomics datasets from ENCODE or FANTOM5, or uploaded user data. Search results can be downloaded, or saved and shared with other researchers, both privately and publicly. The NIAGADS Alzheimer's Genomics DB also provides tools for interactively browsing genomic regions of interest and downloading region- or feature-specific subsets of the available datasets or annotations, including detailed variant, gene, and genomic region record pages. **Results:** With a powerful search interface, analysis tools, and comprehensive record pages linking GWAS summary statistics and feature annotations to variants and genes, the NIAGADS Alzheimer's Genomics Database is a rich resource and valuable tool for AD research.

1372T

Integration of deep learning and causal discovery for medical image semantic segmentation and their application to longitudinal neuroimage data for Alzheimer's disease. Y. Liu¹, Q. Ge², R. Jiao¹, N. Lin¹, Z. Li³, M. Xiong¹. 1) Department of Biostatistics and Data Science, The University of Texas Health Science Center at Houston, Houston, TX; 2) Fudan University, Shanghai, China; 3) Chengdu Jinjiang District Maternity and Child Health Hospital, Chengdu, China.

Semantic segmentation and image feature selection play an important role in prediction, diagnosis, treatment and imaging-genomics data analysis. Convolutional neural networks (CNN) have achieved great success for image classification in medical research, but it is also well known as a 'black box' due to its low interpretability to humans. Understanding and visualizing the image features involved in CNN classification will provide us valuable information about brain regions and disease pathology. Causal inference has been an emerging topic in various domains, with great potentials to speed up biomedical discoveries. We have developed a novel general framework that integrate neural networks and causal inference for semantic segmentation and feature discovery. The new paradigm consists of two stages: (1) use occlusion map to find image regions that are most distinctive for disease status and (2) the state-of-the-art causal inference tools to determine if the selected image regions are causal for AD under the additive noise model framework. This two stage platforms for performing image semantic segmentation and searching image features were applied to the Alzheimer's Disease Neuroimaging Initiative (ADNI) longitudinal dataset. Our CNN models achieved prediction accuracies of 0.9371, 0.9290, 0.9275, and 0.9272 at baseline, 6 months, 12 months and 24 months, respectively. In addition, we were able to predict disease status from baseline to 6 months, from 6 months to 12 months, and from 12 months to 24 months with accuracies of 0.9284, 0.8957, and 0.8137, respectively. We have successfully identified several brain regions including temporal lobe which were important for the progression of Alzheimer's disease as indicated by CNNs and our causal detection tools.

1373F

NIA Genetics of Alzheimer's Disease Data Storage Site (NIAGADS):

Update 2018. O. Valladares^{1,2}, A. Kuzma^{1,2}, E. Greenfest-Allen^{2,4}, P. Gan-gadharan^{1,2}, B. Vogel^{1,2}, C. Zhong^{1,2}, Z. Katanic^{1,2}, L. Qu^{1,2}, M.K. Moon^{1,2}, Y.K. Leung^{1,2,3}, A. Naj^{1,5}, C.J. Stoeckert Jr.^{1,2,4}, G.D. Schellenberg^{1,3}, L.S. Wang^{1,2,3}.

1) Penn Neurodegeneration Genomics Center, Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA; 2) Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, PA; 3) Institute on Aging, University of Pennsylvania, Philadelphia, PA; 4) Department of Genetics, University of Pennsylvania, Philadelphia, PA; 5) Department of Epidemiology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA.

Background: NIAGADS is a national genetics data repository that facilitates access of genotypic data to qualified investigators for the study of the genetics of late-onset Alzheimer's disease (AD) and related dementias (ADRD). Collaborations with large consortia and centers such as the Alzheimer's Disease Genetics Consortium (ADGC), Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium, the Alzheimer's Disease Genome Sequencing Project (ADSP), and the Genome Center for Alzheimer's Disease (GCAD) allow NIAGADS to lead the effort in managing large AD datasets that can be easily accessed by the research community and utilized to their fullest extent. In 2018, NIAGADS began hosting and distributing GRCh38 mapped and variant called data from the ADSP and other NIA-funded studies. **Methods:** Since 2012, NIAGADS has been supported by National Institute on Aging (NIA) under a cooperative agreement (U24 AG041689). All data derived from NIA funded AD genetics studies are expected to be deposited in NIAGADS or another NIA approved site. NIAGADS has developed a Data Sharing Service (DSS) that facilitates the deposition and sharing of genomic data and association results with approved users in the ADRD research community. In addition, researchers are able to freely use the NIAGADS Genomics Database to search annotation resources that link published AD studies to AD-relevant sequence features and genome-wide annotations. **Results:** As of June 2018, NIAGADS houses 37 datasets with >37,000 samples and over 24 billion genotypes, of which 11 have summary statistics loaded into the Genomics Database for interactive exploration and data-mining. The first release of data in the DSS will include 3,985 ADSP whole-genomes from the Discovery + Extension phases and 809 ADNI whole-genomes, all mapped to GRCh38 and processed by GCAD. Qualified Investigators can submit an application for the sequencing data through the newly developed DSS Data Access Request Management system and can download the data through the DSS Portal. Raw data on GRCh38 and quality controlled VCF files are available as of summer 2018. **Conclusions:** NIAGADS is a rich resource for AD researchers, with the goal of facilitating advances in AD research. By housing datasets from multiple projects and institutions, NIAGADS enables AD researchers to be more efficient in meeting their research goals. Datasets, guidelines, and new features are available on our website at www.niagads.org.

1374W

Genotyping of variable number tandem repeats with adVNTR. M. Bakh-tiari¹, S. Shleizer-Burko², M. Gymrek^{1,2}, V. Bansal³, V. Bafna¹.

1) Department of Computer Science and Engineering, University of California San Diego, San Diego, CA; 2) Department of Medicine, University of California San Diego, San Diego, CA; 3) Department of Pediatrics, University of California San Diego, San Diego, CA.

Motivation. Whole Genome Sequencing is increasingly used to identify Mendelian variants in clinical pipelines. These pipelines focus on single nucleotide variants (SNVs) and also structural variants. Repeat mediated sequence variants are relatively less explored with the exception of Short Tandem Repeats (STRs) which are exact repeats of a pattern 1-6bp. We consider the problem of genotyping Variable Number Tandem Repeats (VNTRs), composed of inexact tandem duplications of short (6-100bp) repeating units. VNTRs span 3% of the human genome, are frequently present in coding regions, and have been implicated in multiple Mendelian disorders (e.g., Medullary cystic kidney disease and Myoclonus epilepsy) as well as complex disorders such as Alzheimer's disease and Bipolar disorder. Despite their importance in resolving the genetic basis of these disorders, VNTR genotyping (requiring identification of variation in the number of tandem repeats as well as mutations in the VNTR sequence) remains a challenging problem. Here, we describe a method, adVNTR, that can be used to reliably genotype VNTRs including determination of Repeat Unit (RU) counts and mutation detection. **Method.** adVNTR uses Hidden Markov Models (HMMs) to model the repeat units allowing for sensitive detection of VNTR containing reads, specialized counter-states for RU counts, and a Bayesian model for genotyping diploid counts. Separate models for each VNTR were generated for short-read (Illumina) and single molecule (PacBio) whole genome and exome sequencing. We evaluated adVNTR using long range PCR experiments and assessed Mendelian inheritance consistency in two trios. **Result.** We found that more than 130,000 of the 160,000 VNTRs in exonic region, introns, and promoters of human genes in human genome could be reliably genotyped using adVNTR with Illumina data. While STR genotyping tools such as ExpansionHunter and HipSTR were unable to make calls on cases with non-identical repeating patterns, adVNTR's estimates were perfect in 90.5% of the cases (97% accuracy for one of the two haplotypes). We also used adVNTR for larger population-scale studies of VNTR genotypes and identified many allelic differences between sub-populations. While repeat frequencies were consistent with HW equilibrium in most sub-populations, their violation in a few cases suggested a functional role for those VNTRs.

1375T

A minimally-supervised case-control classifier based on integrative deep-learning. B. Henning^{1,2}, T. Araújo^{1,2}, M. Molck^{1,2}, F. Torres^{1,2}, M. Viana³, I. Lopes-Cendes^{1,2}, F. Cendes^{2,4}, C. Yasuda^{2,4}, B. Carvalho^{2,5}. 1) Department of Medical Genetics, University of Campinas, Campinas, São Paulo, Brazil; 2) Brazilian Institute of Neuroscience and Neurotechnology / BRAINN, Campinas, Brazil; 3) IBM Research, São Paulo, Brazil; 4) Department of Neurology, University of Campinas; 5) Department of Statistics, University of Campinas.

Integrative medicine can improve diagnosis and treatment efficiency by fusing data from heterogeneous sources, such as modalities including, molecular data, imaging exams, environmental variables, and personal lifestyle information. Currently the data integration is carried out manually due to the lack of appropriate statistical and bioinformatics tools. Here we aim to start fulfilling this lack giving the first step towards a truly integrative analysis. We developed a deep learning-based framework to classify subjects as case or control using two modalities of data annotated for epilepsy by experts, brain MRI (3T-Philips, T1-weighted images, isotropic voxels 1x1x1mm) and the corresponding genetic data that was obtained with genome-wide human SNP array 6.0 (Affymetrix). We trained a single mode classifier with MRI (106 case and 71 control subjects) and compared its performance with a multimodal classifier that uses both MRI and the genetic data obtained by the CRLMM algorithm. Both classifiers are composed by 2D convolutional layers that are suitable for learning high-level features of input images. The genetic data was encoded as a three-channels image, where each channel represents a genotype (AA, AB, BB). The output of the convolutional layers in both classifiers is used as input for a fully-connected layer that predicts the subject class. Our findings indicate that both classifiers easily overfit the training data and do not generalize well for the testing set. This behavior may be improved by adding extra information on the subjects. In particular, the genetic information we have currently available is comprised of genotypes obtained at SNPs commonly observed in populations worldwide. Common SNPs may not be sufficient to express predisposition for complex phenotypes like epilepsy and events like these may be better explained with more detailed information, like copy number status. Another source of difficulties is the small number of subjects available and the high dimension of the genetic data, that spans many times the number of subjects in our dataset. Additionally, the minimalist case-control annotation of our data does not contain any information about regions of the MRI or the genetic data relevant for the disease diagnosis. To overcome the current challenges we plan to use 3D convolutional layers, reduce the genetic data dimension and incorporate other types of data.

1376F

Methods for genotype-based risk prediction for complex neuropsychiatric disorders. Z. Yang¹, A. Bose², P. Drineas², P. Paschou¹. 1) Biological Science, Purdue University, West Lafayette, IN; 2) Computer Science, Purdue University, West Lafayette, IN.

Genome-wide association study (GWAS), as a widely-used approach to investigate the genetic basis for complex neuropsychiatric disorders, detects the association between common single nucleotide polymorphisms (SNPs) and phenotypic traits. However, the results from GWAS are always sensitive to change of sample size or even minor genotyping errors, which makes some of the findings extremely difficult to be replicated. Moreover, due to the lack of robustness, variants returned by GWAS do not necessarily yield a satisfying accuracy when being used for trait prediction. To great extent, this deficiency in robustness is resulted from the presence of noise in dataset. Here, we apply denoising techniques over genotyping datasets in order to increase the reliability of trait-related predictor selection. Subsequently, machine learning approaches are implemented on the denoised dataset, with the expectation of a better performance in disorder trait prediction. This framework has been applied over synthetic genotype data as well as genotype of neuropsychiatric and neurological disorders such as Parkinson disease and schizophrenia. A successful application of this framework will lead to an increased power in predictive disorder-related genetic marker detection and higher accuracy in risk prediction.

1377W

Variation among intact tissue samples reveals the core transcriptional features of human CNS cell classes. *K.W. Kelley, H. Inoue, A.V. Molofsky, M.C. Oldham.* UCSF, San Francisco, CA.

It is widely assumed that cells must be physically isolated to study their molecular profiles. However, intact tissue samples naturally exhibit variation in cellular composition, which drives covariation of cell-class-specific molecular features. By analyzing transcriptional covariation in 7221 intact CNS samples from 840 individuals representing billions of cells, we reveal the core transcriptional identities of major CNS cell classes in humans. By modeling intact CNS transcriptomes as a function of variation in cellular composition, we identify cell-class-specific transcriptional differences in Alzheimer's disease, among brain regions, and between species. Among these, we show that *PMP2* is expressed by human but not mouse astrocytes and significantly increases mouse astrocyte size upon ectopic expression *in vivo*, causing them to more closely resemble their human counterparts. Our work is available as an online resource (<http://oldhamlab.ctec.ucsf.edu>) and provides a generalizable strategy for determining the core molecular features of cellular identity in intact biological systems.

1378T

Comparative analysis of gene prioritization tools for neuromuscular disorders. *P.S. Lai, S. Tomar, R. Sethi, P.S. Low, S.K.H. Tay.* National University of Singapore, Singapore, Singapore.

Identifying true causal genes for rare Mendelian disorders from whole genome sequencing (WGS) data remains a challenging and time-consuming effort. Gene prioritization is the process of filtering and ranking annotated variants to facilitate gene-disease associations for clinical diagnostics. In this study, a cohort of 30 patients with neuromuscular disorders (NMDs) carrying previously identified variants in genes such as *LAMA2*, *LMNA*, *CAPN*, *NEB*, *DMD*, *FKRP*, *FHL1*, *MYH7*, *TTN* and *RYR1* were analysed using four popular web-based gene prioritization tools (VarElect, Phenolyzer, OMIMExplorer and Endeavour). The aim was to investigate how well each previously identified gene could be identified as the top candidate or among the top candidates by the four tools upon blinded re-analysis of annotated variants generated from WGS. **METHODOLOGY:** Patient specific HPO terms were used as queries and performance assessed using four well-established measures, namely Median Rank Ratio (MedRR), Normalized Discounted cumulative gain (NDCG), Area under the curve (AUC) and Receiver Operating Characteristic (ROC) curve. **RESULTS:** When candidate genes from the top 5%, 10%, 30% and 50% of the prioritized gene lists were compared, VarElect showed lowest MedRR scores (0.76 – 4.17) and highest NDCG value = 1 across all 30 cases showing that it ranked each real causal gene higher up in the list than the other tools. However, OMIMExplorer was able to correctly identify the causal gene with a 100% true positive rate across all ranked gene lists. The area under the ROC curve gave the probability whether a tool would correctly identify a causal gene in top 1% of gene list across all gene prioritizations and OMIMExplorer showed the highest performance (0.823 AUC; 100% sensitivity) followed by VarElect (0.745 AUC; 84.8% sensitivity), Endeavour (0.736 AUC; 85.3% sensitivity) and Phenolyzer (0.731 AUC; 87.90% sensitivity). Significant correlation was observed between VarElect and Phenolyzer rank ratios. In this NMD cohort, the use of gene prioritization tools was a useful decision-support strategy in identifying causal variants. We recommend incorporating a comprehensive phenotype-based gene prioritization step into the computational pipeline for efficient clinical variant identification and interpretation.

1379F

Molecular modeling, informatics, and pathway prediction of varying complex biological systems for drug repositioning, repurposing, and rescue.

K. Nguyen^{1,2,4}, L. Tian^{1,2}, A. Wenocur^{1,2}, X. Wang^{1,2}, P. Sleiman^{1,2,3}, H. Hakonarson^{1,2,3}. 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Division of Genetics, Children's Hospital of Philadelphia, Philadelphia, PA; 3) Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 4) Center for Dynamic Data Analytics, Rutgers University, Piscataway, NJ.

Precision medicine refers to the tailoring of medical treatment to the individual characteristics of each patient based on harvesting information from multiple methods establishing an information pipeline. In this case, the first and last (for validating purposes) series of steps traditionally consist of a high performance and throughput *in silico* approach using state-of-the-art supercomputers, computer clusters, or graphical processing units (in single workstations). If a gene and its encoded protein are well-characterized and directly linked to the disease in question, then Molecular Docking and Dynamics Simulations through Structure-Activity Relationships, its derivatives, and Structure-Based Virtual Screening can be performed using well-resolved experimental structures. These methods have been successful in elucidating interactions that are not tractable with the most advanced experimental methods. However, there have been instances where this is not the case. For a therapeutic targeting a molecular pathway, Ligand-Based VS and Computational Pharmacological methods can provide valuable information to determine if modifying the structure will provide a better solution. These chemical/molecular informatics methods were created to be delocalized and dynamic with traditional biological informatics and systems/network biology methods. As a result, repurposing therapeutics to target genes and proteins with respect to their disease have been made increasingly transparent. Here, the examples analyzed from this pipeline are associated with amyloidosis and characteristically neurodegenerative. An Artificial Intelligence approach, derived from Deep Machine Learning, has also been in development for benchmarking purposes with respect to each traditional method in the pipeline. In this case, VS using Molecular Descriptors in Computer-Aided Drug Design has been well-established where their property (structure and energetics) and activity (dynamics and function) can be generated using artificial neural networks through layers of nonlinear processing units. Current progress has been validating shallow data sets before increasing to a more proper architecture. Drug Repurposing is already a fraction of the time and cost of traditional *de novo* efforts. Nevertheless, the purpose of exploring this approach is to execute this pipeline with greater speed, effectiveness, and affordability.

1380W

Automated extraction of multiple sclerosis treatment timelines: Preparing free-text electronic health record data for use in pharmacogenetic analyses. J.T. Beales¹, J.C. Denny², M.F. Davis¹.

1) Department of Microbiology and Molecular Biology, Brigham Young University, Provo, UT; 2) Departments of Biomedical Informatics and Medicine, Vanderbilt University Medical Center, Nashville, TN.

Disease-modifying treatments for multiple sclerosis (MS) are associated with a number of adverse drug reactions (ADRs), including hepatic and cardiac damage. Attempts to use pharmacogenetic analysis to identify genetic factors associated with such ADRs are complicated by the relative rarity of some ADRs. Using large databases of electronic health records (EHRs) represents one potential approach to overcoming this challenge. However, this approach necessitates the ability to reliably extract treatment timelines from these records. We developed a rule-based algorithm able to extract MS treatment timelines for glatiramer acetate, a common MS treatment, to aid in pharmacogenetic studies. Records referencing glatiramer acetate from 846 MS patients were retrieved from the Vanderbilt University Medical Center Synthetic Derivative, a deidentified version of the EHR, and used to develop the algorithm. Regular expressions were developed to match language patterns used by clinicians to indicate a beginning, continuation, or end of a treatment period. The algorithm employs these regular expressions to classify EHR medication references, generating a timeline of treatment events for patients. This timeline can guide ADR extraction in future studies. A training dataset of 100 EHR notes was manually reviewed and classified. For the medication references in the training set, the algorithm reliably identified treatment initiation events (accuracy, sensitivity, specificity, and precision were > 95%) and treatment discontinuance events (accuracy, specificity, and precision were > 95%, sensitivity 88.9%). More rigorous testing of this particular algorithm is necessary to assess its performance on larger datasets. However, this study demonstrates that automated, rule-based extraction of MS patient treatment timelines from free-text EHRs is possible. These timelines will be used in pharmacogenetic studies to investigate variants associated with ADRs to glatiramer acetate. .

1381T

Target gene notebook: Putting genetics to work for drug discovery. *M.P. Reeve, J. Wierzbowski, A. Kirby, J. Hutz.* Human Biology and Data Science, Eisai, Andover, MA.

Many thousands of associations between genetic variants and disease/phenotype provide clues to the causal mechanisms of nearly all common human diseases. Very few GWAS hits, however, have yielded actionable therapeutic hypotheses due to challenges in connecting variants to genes and function. While many browsers and databases provide access to genome annotations and different types of genetic results, linking and maintaining group curation of all relevant information from these sources and integration with proprietary experimental data remains challenging. We created freely-available software, Target Gene Notebook (TGN), to facilitate the process of therapeutic target evaluation and creation of durable institutional or public knowledgebases. With a gene or locus as the focal point, TGN gathers published associations from sources including GWAS Catalog, UK Biobank, and GTEx eQTLs and allows the user to review and store an assessment of each finding. Published LD patterns derived from reference panels are used to establish a 'credible set' of variants from each association. Interactive visualization of credible sets allows rapid assessment of colocalization across these diverse results and helps highlight variants of known functional consequence. Along with automated access to large public resources, new associations can be added, either as single associations or as additional datasets. Additional public resources of relevance to target evaluation (e.g., variation patterns in ExAC/Gnomad, tissue expression patterns in GTEx) are also presented. TGN also provides a bibliography – initially seeded with publications from the above resources – permitting manual expansion, curation/annotation and capture of key figures and tables made available in expandable thumbnails. TGN exists to support group annotation with multiple users having write access to specific gene databases and all users having immediate access to new annotations and references added by any user. Implementation is such that private instances can be shared across groups or sites as needed – potential users include groups within pharma (for tracking knowledge of targets across biology, computational and chemistry teams) and disease consortia for collating results and interpretations of their conclusive findings.

1382F

Using model predictions as quantitative traits to expand cohort size for heritability estimation and genetic studies. *P.M. Thangaraj^{1,4,6,7}, A.O. Basile^{1,4,6,7}, N. Shang¹, F. Polubriagino^{1,4,6,7}, B.R. Kummer^{1,3}, R. Vanguri^{1,4,6,7}, T. Lorberbaum^{1,4,6,7}, J.D. Romano^{1,4,6,7}, K. Quinnes^{1,4,6,7}, A. Yahji^{1,4,6,7}, K. Kiryluk², A. Gharavi^{2,6}, C. Weng¹, M.S. Elkind^{2,5}, G.M. Hripcsak¹, N.P. Tatonetti^{1,4,6,7}.*

1) Department of Biomedical Informatics, Columbia University Irving Medical Center, New York, NY; 2) Division of Nephrology, Department of Medicine, Columbia University Irving Medical Center, New York, NY; 3) Department of Neurology, Columbia University Irving Medical Center, New York, NY; 4) Department of Systems Biology, Columbia University Irving Medical Center, New York, NY; 5) Department of Epidemiology, Columbia University Irving Medical Center, New York, NY; 6) Institute for Genomic Medicine, Columbia University Irving Medical Center, New York, NY; 7) Department of Medicine, Columbia University Irving Medical Center, New York, NY.

Large genetic repositories connected to electronic health records (EHR) form the basis of local and national precision medicine initiatives. These linked repositories promise the ability to perform thousands of simultaneous genetic studies using diagnoses, procedures, and treatment responses that are routinely captured by the EHR. Not all patients have genetic data available, however, limiting what actually can be studied. There is an opportunity to use machine learning methods to derive quantitative proxy variables and expand the sample population. For example, genetic studies of stroke, a complex disease and a leading cause of death, can require hundreds of thousands of patients. In the 6.2 million patient records of New York-Presbyterian Hospital from 1986-2018, we have 79,000 coded cases of stroke. We hypothesized that the output of a supervised machine learning classifier can be used as a proxy probability for a previous stroke diagnosis for cohort expansion. This study investigates the feasibility of converting a dichotomous phenotype to a quantitative phenotype. We applied our model to estimate observational heritability of acute ischemic stroke. We identified families from the EHR using RIFTEHR, an algorithm that uses emergency contact information to infer familial relationships between patients. We then trained classifiers on stroke cases and controls using EHR-extracted features and tested every patient in the EHR for a previous stroke diagnosis. This expanded the cohort 20-fold without the requirement of manual phenotyping. The area under the receiver operative characteristic curve was 0.75-0.99 for the random forest (RF) models. Mean RF classifier heritability, 0.379 ± 0.052 , was within margin of error of the literature heritability estimate, showing successful estimation of heritability using a quantitative trait proxy. To demonstrate the utility of using this proxy in a genetics study, we simulated a complex trait under an additive model, varying the effect sizes, penetrance, and allele frequencies. Finally, we validated our approach by running a GWAS in EHR-linked genotypes containing algorithmically phenotyped cases ($n=289$) and controls ($n=1721$) of venous thromboembolism. We found that conversion from a dichotomous to quantitative trait resulted in minimal changes to the GWAS estimates for established causal variants. We demonstrate that our quantitative proxy trait can improve performance by increasing the effective sample size.

1383W

Identifying splice alterations using RNAseq to improve genetic diagnosis for rare Mendelian disorders. H. Lee¹, A. Eskin², S.F. Nelson^{1,2}, *Undiagnosed Diseases Network Members*. 1) Department of Pathology and Laboratory Medicine, University of California, Los Angeles, Los Angeles, CA; 2) Department of Human Genetics, University of California, Los Angeles, Los Angeles, CA.

Whole exome and genome sequencing (WES/WGS) are becoming more accessible as sequencing and computational cost decrease. However, >70% of WES/WGS analyses do not reveal a Mendelian mutation and diagnosis given rare disease case, and the vast majority of DNA variants, both *de novo* and inherited, are difficult to interpret including variants within promoters, introns and exons. Here we describe an analytical approach to identify rare splicing abnormalities linked to specific DNA variants from RNAseq data. We demonstrate how these data can be intersected with WES/WGS data to identify which rare genomic variants are pathogenic and improve genetic diagnosis for individuals within the Undiagnosed Diseases Network and exome negative individuals. RNA was extracted from blood, fibroblast or muscle following standard protocols. Illumina TruSeq Globin-Zero gold (blood) or Ribo-zero gold (other tissues) were used for library preparation and sequencing was performed to generate ~50 million 69 base paired-end reads on the HiSeq 4000. FASTQ files were aligned to the genome using STAR-2.5.2b twopassMode, and splice junctions from STAR output were exported to an internal splice junction database. The splice junction database is populated with count data for each detected splice junction across all RNAseq samples available where there was at least one read that overlapped a given junction. Second, for each start or end of the junction, the junctions are flagged to have “alternative splicing” if there are more than one junction at the same start or end of the junction and categorized into one of the 4 types of alterations that we consider here: full exon skipping, inclusion of intronic pseudoexon, exon extension or exon reduction. Third, a Z score of the ratio of the coverage for the alternate junction(s) in an individual relative to the coverage for the canonical junction is calculated. Commonly observed alternate junctions are filtered out if observed in more than 1% in our internal database. Finally, WES/WGS variants are annotated to correlate DNA variant with effect on mRNA. We describe how this analytical tool identifies rare DNA mutations that lead to unanticipated splice alterations and changes to the mature mRNA, which improves the diagnostic rate for WES/WGS negative cases but can also be widely used to rule out genomic variants that are challenging to interpret, making transcriptome analysis an essential tool to augment WES/WGS data interpretation.

1384T

Repeat expansion detection with whole exome sequencing data. M. Bahlo^{1,2}, M.F. Bennett^{1,2,3}, R.M. Tankard¹, P. Degorski¹, P. Perucca^{4,5,6}, T. O'Brien^{4,5,6}, S.F. Berkovic¹, P. Kwan^{4,5,6}, I.E. Scheffer^{3,10}, M.B. Delatycki^{3,9,10}, P.L. Lockhart^{8,10}. 1) Population Health and Immunity Division, The Walter and Eliza Hall Institute of Medical Research, 1G Royal Parade, 3052 Parkville, Victoria, Australia; 2) Department of Medical Biology, University of Melbourne, 1G Royal Parade, 3052 Parkville, Victoria, Australia; 3) Epilepsy Research Centre, Department of Medicine, The University of Melbourne, Austin Health, Heidelberg, Victoria, Australia; 4) Department of Neuroscience, Central Clinical School, Monash University, Melbourne, Victoria, Australia; 5) Departments of Medicine and Neurology, The Royal Melbourne Hospital, The University of Melbourne, Parkville, Victoria, Australia; 6) Department of Neurology, Alfred Health, Melbourne, VIC, Australia; 7) Mathematics and Statistics, Murdoch University, Murdoch, WA, Australia; 8) Bruce Lefroy Centre for Genetic Health Research, Murdoch Children's Research Institute, Royal Children's Hospital, Parkville, Victoria, Australia; 9) Victorian Clinical Genetics Services, Parkville, Victoria, Australia; 10) Department of Paediatrics, University of Melbourne, Parkville, Victoria, Australia.

Repeat expansions cause over 20 neurogenetic disorders that can present with overlapping clinical phenotypes, making molecular diagnosis challenging. Single gene or small panel PCR-based methods are often employed to identify the precise genetic cause, but can be slow and costly, and often yield no result. Genomic analysis via whole exome and whole genome sequencing (WES and WGS) is being increasingly performed to diagnose genetic disorders. However, until recently analysis protocols could not identify repeat expansions in these datasets. Several methods have now been developed to specifically identify repeat expansions in short-read sequencing data, focusing on PCR-free whole genome sequencing application, where PCR stutter error is much reduced. We have developed exSTRA (expanded STR algorithm), an outlier detection method for the detection of repeat expansions in cohorts of sequenced individuals. It can be applied to existing WES or WGS data to identify likely repeat expansions. We demonstrate that exSTRA can be effectively utilized as a screening tool to interrogate WES and WGS sequencing data generated with PCR-based library preparations which can then be followed up with specific diagnostic tests. exSTRA produces informative and unique graphical summaries of the STR motif distributions which aid interpretation of the results. The exSTRA package is available from <https://github.com/bahlo-lab/exSTRA>. exSTRA demonstrated very high predictive capabilities, achieving a median area under the curve (AUC) of 0.9 when applied to WES and WGS (PCR-based library protocols) cohorts comprised of both normal and known repeat expansion disorder patients. An extensive simulation study has been conducted over 20 known STR loci, comparing the repeat expansion detection algorithms ExpansionHunter (Dolzhenko et al, Genom Res, 2017), exSTRA (Tankard et al, bioRxiv, 2017), TREDPARSE (Tang et al, AJHG, 2017) and STRetch (Dashnow et al, bioRxiv, 2017). The simulation study demonstrated consistent performance by all four algorithms in the context of repeat length, with longer repeats being more significant. Furthermore results were robust, even when only modest cohort sizes (controls) were available. We have applied these four repeat expansion detection methods to a cohort of 40 focal epilepsy patients that had previously undergone clinical whole exome sequencing (Perucca et al, Epilepsy Research, 2017). We detected a previously undiscovered SCA8 expansion in this cohort.

1385F**Tackling heterogeneity using network based stratification in autism spectrum disorders.**

Y. Kim^{1,2,3,4}, F. Cliquet^{1,2,3,4}, T. Rolland^{1,2,3,4}, R. Delorme^{1,2,3,4,5}, T. Bourgeron^{1,2,3,4}, G. Dumas^{1,2,3,4}. 1) Institut Pasteur, Human Genetics and Cognitive Functions Unit, Paris, France; 2) CNRS UMR 3571 Genes, Synapses and Cognition, Institut Pasteur, Paris, France; 3) University Paris Diderot, Sorbonne Paris Cité, Paris, France; 4) C3BI, USR 3756 Institut Pasteur and CNRS, Paris, France; 5) Department of Child and Adolescent Psychiatry, Robert Debré Hospital, APHP, Paris, France.

Autism Spectrum Disorders (ASD) are heterogeneous neurodevelopmental conditions diagnosed in more than 1% of the population. For better prognosis and treatment, genetic stratification is suggested which aims to stratify patients into more homogeneous subgroups. In the field of cancer research, the method of Network Based Stratification (NBS) has been proposed (Hofree et al. 2013). The NBS integrates the mutation profile of each patient mapped on a Protein-Protein Interactions (PPI) network. We validated the reproducibility of the NBS with a part of the original cancer data and STRING PPI network. We also applied the NBS on simulated data which confirmed efficiency of a clustering qualifying score: Proportion of Ambiguously Clustering (PAC). Then the NBS using Agile Protein Interactomes DataSever (APID) PPI network was applied to a large WES cohort of 2753 ASD patients and 2272 unaffected siblings from the Simons Simplex Collection (SSC). Our preliminary results show that the diffusion of the score of the mutated genes across the PPI network is one of the most important factors on stratification. The investigation of parameters such as minor allele frequency or predicted deleteriousness on the stratification is in progress.

1386W**Detecting inherited deletions in whole genome sequencing data using family structure.** *K.M. Paskov, D.P. Wall.* Biomedical Informatics, Stanford University, Stanford, CA.

Copy number variants (CNVs) have been linked to neurodevelopmental disorders such as autism, schizophrenia, and developmental delay. CNVs occur more frequently in individuals with developmental disabilities than in their neurotypical counterparts with increased CNV burden associated with more severe phenotypes [Cooper 2011]. CNV size and type both influence pathogenicity with deletions shown to be pathogenic at smaller size ranges (30-500kbp) compared to duplications (>500kbp) [Pinto 2010]. Detecting deletions as small as 30kbp in length is challenging. We introduce a novel method for the detection of inherited deletions from whole genome sequencing data using family structure. Inherited deletions result in distinctive inheritance signatures as variants are transmitted from parents to children. Since variant calling algorithms do not consider the presence of deletions, hemizygous sites (0/- and 1/-) are called as either homozygous reference (0/0) or homozygous alternate (1/1). This results in seemingly non-Mendelian transmission for variants spanning the deletion. Hidden Markov models (HMM) have previously been used to identify crossover sites in nuclear families with a median resolution of 2.6kbp [Roach 2010]. We adapt this approach to identify inherited deletions. Our HMM models the crossover sites and parental chromosomes inherited by each child, and then uses this information to identify regions corresponding to deletion-specific inheritance patterns, supporting each deletion with genotype calls in multiple family members. The resulting model can detect deletions as small as 0.1kbp in length and can identify deletion start and endpoints with a median resolution of 0.2kbp. We applied our method to 30x whole genome sequence data for 869 multiplex families with idiopathic autism, composed of 1747 affected children, 389 unaffected siblings and their parents. We discovered thousands of inherited deletions per family, with 82 regions showing a significant association between presence of a deletion and affected status even after Bonferroni correction. Of these regions, 91% are expressed in infant or fetal brain [Jaffe 2014]. Finally, we compared the sensitivity of our method to Lumpy [Layer 2014] and Breakdancer [Chen 2009], two structural variant detection algorithms that use split reads and read depth to identify deletions. These results demonstrate the value of sequencing whole families and suggest that inherited deletions contribute to autism risk.

1387T

Comparison of tools for RNA-Seq differential gene expression analysis of small, heterogeneous effects requiring large sample sizes. C. Armoskus¹, O.V. Evgrafov², D. Conti¹, J.A. Knowles². 1) University of Southern California, Los Angeles, CA; 2) SUNY Downstate Medical Center, Brooklyn, NY.

Transcriptome analysis via short read massively parallel sequencing (RNA-Seq) presents great opportunities for understanding the biology of many diseases. However, the vast majority of evaluations of differential expression tools are done examining situations with large effect sizes and consistent effects in small sample sizes; the results of such evaluations may not be applicable to examining differential expression in other settings. Common complex diseases which feature small, heterogeneous differences in gene expression that require large sample sizes to identify, such as many psychiatric disorders, present a substantially different situation that must be examined to identify optimal methodologies for this application. To examine relevant gene expression differences for this type of condition, we simulated data based on studies such as the CommonMind and CNON projects, which use hundreds of samples to examine schizophrenia. Simulated data was based on a number of statistical distributions, such as the negative binomial, log-normal, and multinomial, all parameterized to match characteristics observed in real data. A simulation approach allows the creation of data matching a wide variety of specifications as well as knowledge of truth; this offers a clearer examination of correct results compared to consistency based approaches that compare results between different tools or subsets of a dataset. A variety of analysis programs (e.g., DESeq2, edgeR, and voom-limma) with different settings were examined, such as the use of different normalization techniques. Batch effects and other sources of technical variance and methods for removing them (e.g., using covariates such as batch and SVA) were tested and model selection techniques were employed. Methodologies were evaluated primarily by ROC and precision-recall curves, and secondly by power and control of the FPR/FDR. Work is ongoing, with early results showing the superiority of newer normalization techniques such as edgeR's TMM and DESeq2's median-of-ratios methods compared to quantile normalization and scaling to the total number of reads. Following this the greatest variance between programs is based on what distribution they expect, with voom-limma performing better on log-normal data and edgeR/DESeq2 performing significantly better on negative binomial data. This highlights the need for a greater examination of the true/most common distribution of RNA-Seq data both in general and in a specific dataset.

1388F

DISTMIX2: Direct imputation of summary statistics from mixed ethnicity cohort in large reference panel. C. Chatzinakos, V. Vladimirov, T. Webb, B. Riley, K. Kendler, S. Bacanu. Virginia Commonwealth University (VCU), Richmond, VA.

To increase the signal resolution for large-scale meta-analyses of genome-wide association studies, genotypes at unmeasured Single Nucleotide Polymorphisms (SNPs) are commonly imputed using large multi-ethnic reference panels. SNPs imputation is expected to increase the statistical power because it increases the accuracy of genomic predictions. Direct imputation of summary statistics was shown to be practically as accurate as genotype imputation method, while incurring orders of magnitude lower computational burden. However, to be precise, direct imputation methods need accurate estimates of LD between SNPs and larger reference panels. A previous version of our software, DISTMIX, achieved such estimates by using in-cohort allele frequency (AF) estimates. As genome wide association studies (GWAS) stopped providing AFs, we now use a novel method to estimate ethnic composition of the cohort using only the GWAS summary statistics, e.g. Z-scores, and reference panel data. Simulations under the null hypothesis of no association, show that the proposed method adequately controls the Type I error rates, while using the pre-estimated weights fails to control them. DISTMIX and its main direct imputation competitors still use a small reference panel of 1,000 subject coming from the 1000 Genome (1KG) project. Because the success in GWAS studies depends on good phenotypic and genotypic data (reference panel), the new version of the software, DISTMIX2, uses a much larger reference panel. For the larger reference panel, we use the publicly and privately available data to obtain a 33,000 (33K) panel which includes 11K Han Chinese from the CONVERGE study. We apply the method to impute summary statistics from the ethnically diverse (cosmopolitan) Psychiatric Genetic Consortium (PGC) data sets. Results of these analyses clearly show that the method detects new signals that were not available when using the 1KG panel. Among them, the most interesting is the analysis of PGC Schizophrenia Phase 2, which provides several novel signals in chromosome 13 that seem to be specific to East-Asians.

1389W

An interactome perturbation framework prioritizes damaging missense mutations for developmental disorders. S. Chen^{1,2,3}, R. Fragoza^{1,2,3}, L. Klei⁴, Y. Liu^{1,2}, J. Wang⁵, K. Roeder⁶, B. Devlin⁴, H. Yu^{1,2}. 1) Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY 14853, USA; 2) Weill Institute for Cell and Molecular Biology, Cornell University, Ithaca, New York 14853, USA; 3) Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York 14853, USA; 4) Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15213, USA; 5) Department of Statistics and Data Science, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA; 6) Computational Biology Department, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA.

Identifying disease-associated missense mutations remains a challenge, especially in large-scale sequencing studies. Here we establish an experimentally and computationally integrated approach to investigate the functional impact of missense mutations in the context of the human interactome network and test our approach by analyzing ~2,000 de novo missense (dnMis) mutations found in autism subjects and their unaffected siblings. Interaction-disrupting dnMis mutations are more common in autism probands, these mutations principally affect hub proteins, and they disrupt a significantly higher fraction of hub interactions than in unaffected siblings. Additionally, they tend to disrupt interactions involving genes previously implicated in autism, providing complementary evidence that strengthens previously identified associations and enhances discovery of new ones. Importantly, by analyzing dnMis data from six disorders, we demonstrate that our interactome perturbation approach offers a generalizable framework for identifying and prioritizing missense mutations that contribute risk to human disease.

1390T

Systems biology approach to evaluate genetic factors of antidepressant treatment outcome in major depressive disorder. A.C. Hauschild^{1,2}, M. Maciukiewicz³, V.S. Marshe^{3,4}, D.J. Müller^{3,4,5}. 1) Krembil Research Institute, Toronto, Canada; 2) University Health Network, Toronto, Canada; 3) Pharmacogenetic Research Clinic, Center for Addiction and Mental Health, Toronto, ON, Canada; 4) Institute of Medical Science, Faculty of Medicine, University of Toronto, Toronto, ON, Canada; 5) Department of Psychiatry, University of Toronto, Toronto, ON, Canada.

Major depressive disorder (MDD) is a costly, debilitating and among the most prevalent psychiatric disorders. Affecting over 350 million people worldwide, it is rated the fourth leading factor of disability globally (WHO, 2016). While different classes of antidepressant are available, they frequently result in solely moderate rates of remission. Hitherto, we are missing knowledge about the underlying genetic and molecular mechanisms that influence treatment outcome. In a previous study by Maciukiewicz et al. 2018, we analyzed a data set of 186 MDD patients treated with duloxetine and subsequently categorized as “responders/non-responders” and “remitters/non-remitters” based on the MADRS score. We used a sophisticated nested cross-validation schema including lasso regression, classification-regression trees and support vector machines to classify responders and remitters. The evaluation identified 19 single nucleotide polymorphisms (SNPs) that were most robust among the cross-validation predictors of remission. Following up this research, we will present a systematic approach to validate those findings in two independent samples and uncover the underlying molecular mechanisms leading to different treatment outcomes. At first, we query the SNPs of interest and corresponding genes for their protein-protein-interactions (PPI) and their interacting neighbors. Subsequently, we identify structurally connected components like subnetworks and cliques in the resulting PPI network that can be annotated with functional information such as gene ontology and metabolic pathways using NAViGaTOR, pathDIP, and g:Profiler. Finally, we show how this evaluation can highlight biological functions and mechanisms that are influencing the effectiveness of antidepressants and pave the way for genotype-driven personalized MDD treatment.

1391F

DVINDEL: Towards an accurate de novo indels caller. A. Lian^{1,2,3,4}, D.

Antak^{2,3,4,5}, P. Tandon^{2,3,4}, O. Shanta^{2,3,4}, M. Gujral^{2,3,4}, J. Sebat¹. 1) Center for Medical Genetics, School of Life Sciences, Central South University, China; 2) Beyster Center for Genomics of Psychiatric Diseases, University of California San Diego, La Jolla, CA 92093, USA; 3) Department of Psychiatry, University of California San Diego, La Jolla, CA 92093, USA; 4) Department of Cellular and Molecular Medicine and Pediatrics, University of California San Diego, La Jolla, CA 92093, USA; 5) Biomedical Sciences Graduate Program, University of California San Diego, La Jolla, CA 92093, USA.

De novo mutations (DNMs) affect pathogenic risk for autism. Despite continued success in identifying de novo single nucleotide variants that confer risk, de novo indels (insertions/deletions) have been poorly characterized. Current software tools for detecting de novo indels have limited success and carry a high false discovery rate (FDR), affecting the analysis of genetic associations. Here we develop an accurate and convenient software tool that leverages statistical machine learning to predict de novo indels in next generation sequencing (NGS) with high accuracy. We implemented a Random Forest model trained on 11 pairs of monozygotic twins and their parents, with features obtained from GATK such as Variant Quality Score Recalibration (VQSR). To increase the number of positive examples for training, we created synthetic de novo mutations by genotyping private inherited variants in the offspring with unrelated parents. False positives for training were discordant putative de novo indels. We then optimized model parameters through 10-fold cross validation (FDR: 0.0024, F1 score: 0.941). We tested classifier performance on whole genome sequences from 94 offspring in the REACH cohort. Similar to training, we created synthetic de novo mutations by genotyping inherited variants using unrelated parents and defined false positives as any putative de novo call. We obtained an accuracy of 0.88, FDR of 0.005, and F1 score of 0.926. Finally, we tested the performance of our classifier on previously validated de novo indels from the Simons Simplex Collection (SSC) data (22 offspring) (Lossifov et al., Nature 2014). Since the de novo variants were discovered using whole exome sequencing, we only tested indel calls in exons and defined false positives as variants that were not previously annotated. Our classifier was able to distinguish between true and false positive de novo indels (accuracy: 0.9862, FDR: 0.1363, F1 score: 0.8837). Due to the high genotyping error rates of indels relative to SNPs, researchers may hesitate to analyze indels. We have shown that it's possible to rapidly detect and filter de novo indels from NGS data without compromising sensitivity and accuracy.

1392W

Identifying biomarkers to build a predictive model for PTSD using machine learning approaches. S. Woo, H. David. Axio Research, Seattle, WA.

Post-trauma stress disorder (PTSD) is one of the mental diseases which can be developed by traumatic events such as deployment in battlefields or natural disasters. Because traumatic events are more likely associated with mental health, building a predictive model with biomarkers based on machine learning approaches was not developed. In this study, we present workflow of building a predictive model using a publically available PTSD data (RNA-seq, GSE63878). For the task, we adapted machine learning approaches and penalized regression models to identify biomarkers for building a predictive model for PTSD and compared their performances for identifying true PTSD patients. They include Random Forest, Support Vector Machine (SVM), Elastic net, LASSO and Deep learning. Their performances are compared by AUC, Sensitivity, and Specificity as accuracy measures and misclassification rate as an error measure. Because of relatively small sample size, we utilized 0.632 bootstrapping method for simulating replicates (n=1000) and for estimating these accuracy measures and an error measure. Also, we show how those identified biomarkers are truly associated with a biological function or biological signal which have known for mental diseases.

1393T

Meta-analysis of Janssen DiseaseLand omics database for disease signatures and treatments. C. Wu¹, T. Lovenberg², B. Huang¹, X. Yao¹. 1) Computational Sciences, Discovery Sciences, Janssen R&D LLC; 2) Neuroscience Therapeutic Area, Janssen R&D LLC.

Background: Transcriptomics technologies such as high-throughput next-generation sequencing and microarray platforms provide exciting opportunities for improving diagnosis and treatment of complex diseases by facilitating the development of pharmacotherapies able to modulate gene expression. Despite their great promise, individual studies may generate results that are not reproducible due to a small number of samples, potential confounding factors, different platforms and bioinformatics pipelines. A clear understanding and unified picture of many complex diseases is still elusive, highlighting an urgent need to effectively integrate multiple transcriptomic studies for disease signatures. In this study, we provide an illustrative example using multiple meta-analysis algorithms to combine a total of 30 genome-wide gene expression studies for insight into bipolar disorder. **Methods:** We have collected comprehensive high-quality transcriptomic datasets from both public domains and internal sources. In the present study, a total of 30 RNA Sequencing (RNA-Seq) or Microarray-based datasets on bipolar disorders were included. Raw FASTQ or CEL files were extracted from Gene Expression Omnibus (GEO) and our internal Janssen DiseaseLand database. We applied the same QC processes and normalization approaches to datasets from multiple platforms. Both p-value and effect size based meta-analysis algorithms were used to identify differentially expressed (DE) genes. We then conducted pathway enrichment analysis and prioritized candidate compounds via anti-genomic similarity between compound profiles within our Janssen L1000 database and the bipolar-associated gene signature. **Results:** We created a shared DE gene list between the r-th ordered p-value (rOP) and random effect model (REM). Among >10,000 genes, 328 were identified as significantly differentially expressed (FDR < 0.05) in any region of the brain (24 studies) and 204 in the prefrontal cortex (PFC, 15 studies). Among these were *BDNF*, *VGF*, *WFS1*, *DUSP6*, *CRHBP*, *MAOA* and *RELN*, which have been implicated in bipolar disorders. Pathway enrichment analysis revealed a role for GPCR, MAPK, immune system and Reelin pathways. **Conclusions:** We established systematic meta-analysis approaches based on the Janssen DiseaseLand Omics Database. This pipeline can be applied in multiple disease areas to create a unified picture of the disease signature and prioritize drug targets, pathways, and compounds.

1394F

Association of addiction and psychiatric phenotypes to predicted transcript levels. D.M. Bost^{1,2}, C. Bizon^{1,2}, J. Tilson², X. Yin², K. Wilhelmsen^{1,2}. 1) Genetics (Bioinformatics and Computational Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Renaissance Computing Institute, Chapel Hill, NC.

GWAS studies unveil a wide range of variants associated with complex disease traits. Interpretation of these variants, however, is limited as it does not address the functional mechanisms involved. This deficit in identifying biological mechanisms, along with the penalty for multiple testing, hinders our ability for attributing causality from our variants to our traits. PrediXcan, a gene-based association method that prioritizes genes that are likely causal for phenotypes, has been introduced to circumvent the aforementioned hurdles in single marker analysis. PrediXcan utilizes gene regulation as a means to identify gene to trait associations. By quantifying the association between genetically regulated expression and phenotype PrediXcan allows the user to gather more biological inference from their subset of variants. PrediXcan uses reference transcriptome data to train additive models of gene expression levels. Weights were derived from the estimates of the genetically regulated expression. In regression testing which is used by Hae Kyung Im et al. there is an assumption of independence between data. This does not lend itself to studies with family data. To account for identity-by-descent we implement the use of a linear mixed model (LLM) that will produce a relatedness matrix. We implement this LLM with a Genome-wide Efficient Mixed Model Association (GEMMA) algorithm. Using the model weights estimated by PrediXcan, we predict transcript levels in brain tissues that we then perform association test on with GEMMA. Our association is conducted with addiction phenotypes, from the University of San Francisco Family Study, which include, but aren't limited to, alcohol use, marijuana use, and nicotine dependence. We also performed associations on psychiatric phenotypes obtained from the Minnesota Multiphasic Personality Inventory – 2nd edition which include, but aren't limited to, schizophrenia, hypomania, paranoia, and depression.

1395W**Psychiatric disorders related gene co-expression modules in brain development.**

Y. Dai¹, T. O'Brien¹, G. Pei¹, Z. Zhao^{1,2,3}, P. Jia¹. 1) Center for Precision Health, School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA; 2) Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA; 3) Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN 37203, USA.

Psychiatric disorders such as schizophrenia (SCZ), bipolar disorder (BIP), major depressive disorder (MDD), attention deficit-hyperactivity disorder (ADHD), and autism spectrum disorder (ASD) are often related to brain development. Both shared and unique biological and neurodevelopmental processes have been reported to be involved in these disorders. In this work, we developed an integrative analysis framework to seek for the sensitive spatiotemporal point during brain development underlying each disorder. Specifically, we overlaid Genome-Wide Association Study (GWAS) summary statistics of five psychiatric disorders to a reference protein-protein interaction (PPI) network, which was dynamically weighted by spatiotemporal co-expression data featuring four brain regions, including frontal cortex (FC), sub-cortical regions (SC), sensory motor regions (SM), and the temporal-parietal cortex (TP), among three developmental stages (embryonic, birth to 11 years old and older than 13 years). We adapted our previous method, EW_dmGWAS, to identify highly co-expressed modules in each spatiotemporal sites and ranked them by the combined effect of component genes weighted by GWAS summary statistics. As a result, we found that SCZ and BIP were closely clustered in 10 out of the 12 investigated spatiotemporal sites, while ASD appeared to be distantly related to the other four disorders. Surprisingly, we only observed very limited number of modules that were shared in multiple disorders in each sites, implying a much more complicated relationship among these disorders at the pathway/network level. At the gene level, we identified several genes that were shared among the most significant modules. Example genes included *CTNNB1*, a gene from the Wnt signaling pathway and involved in neurogenesis maintenance, which was found in top modules in SC in stage 2. An immune related gene *LNX1* was found to be in all top disease specific modules in TP in stage 3. A further function analysis of the module genes revealed the enrichment in immune and psychiatric disease related pathways. In summary, our study demonstrated the changes of disease related modules in different spatiotemporal point and discovered that disease related genes could contribute to the development of disorders in different spatiotemporal point.

1396T**A simple bioinformatics tool to facilitate implementation of pharmacogenomics in the clinical practice of psychiatry.** *A. Baskys.* Genomic Medicine Clinic, Newport Beach, CA.

Despite the improved side effect profiles of pharmacological agents, patients continue to experience significant side effects, symptom relapse, lack of treatment response or undesirable drug-drug and drug-gene interactions (DDIs and DGIs). Pharmacogenetic testing (PGx) based on identification of mutations in genes encoding cytochrome P450 peroxidase enzymes (CYP) that define Extensive, Intermediate, Poor and Ultra-rapid metabolizer phenotypes (EM, IM, PM and UM respectively) has recently become commercially available for prescribers. However, relatively large PGx data sets produced by genetic testing are nearly impossible for a human mind to meaningfully integrate into a clinical decision during a short patient visit (7 to 15 minutes), which contributes to a prescriber reluctance to implement PGx in clinical practice. Further, tabular PGx reports may be lengthy, static, and typically do not integrate both the DDI and DGI aspects of interactions. To address these problems, we developed MEDpicker, a simple bioinformatics tool that calculates drug metabolic rate deviations from expected values caused by DDIs and DGIs, and presents this information in a visual format. It factors in *CYP1A1*, *1A2*, *2B6*, *2C9*, *2C19*, *2D6*, *3A4* and *CYP3A5* gene variants that result in EM, IM or PM phenotypes as well as *CYP2D6* and *CYP2C19* Ultra-Rapid Metabolizer (UM) phenotypes to identify non-responders. The reference database includes over 1000 prescription drug-CYP enzyme interaction entries derived from published reports and publicly accessible databases. The graphic presentation of DDIs and DGIs makes medication selection fast and easy regardless of the number of interactions. Medication selections may be modified dynamically by a virtual substitution of one medication with another to obtain an optimal CYP enzyme induction or inhibition level and, consequently, the optimization of medication side effects and their effectiveness. This software tool is transferable from one clinical setting to another and can be used by the patient when selecting over the counter drugs or nutraceuticals. An evaluation version of this bioinformatics tool is freely available for prescribers, and educators with the expectation that it will facilitate the successful implementation of PGx technologies into clinical practice.

1397F

ANCO genes: Annotations and comprehensive analysis of candidate genes for alcohol, nicotine, cocaine, and opiate (ANCO) dependence. R. Hu¹, Z. Zhao^{1,2,3}, P. Jia¹. 1) Center for Precision Health, School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA; 2) Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA; 3) Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN 37203, USA.

Substance dependence has become a major health, socioeconomical, and behavioral issue across the world. Studies have found that genetic factors might account for approximately half of the likelihood of an individual developing substance dependence. So far, there have been thousands of studies reporting various genetic evidence (e.g. genome-wide association studies (GWAS), genome sequencing, gene expression, noncoding RNA expression, methylation, proteomics, metabolomics, iPSCs, and drug studies). These rich, but highly heterogeneous data, provide us an unprecedented opportunity to systematically collect, curate, and assess the genetic evidence from published studies and to perform comprehensive analysis of their features, functional roles, and druggability. This is especially important to compare the features among major substance dependence. In this study, we performed a comprehensive data collection of four related phenotypes: alcohol dependence (AD), nicotine dependence (ND), cocaine dependence (CD), and opiate dependence (OD). These datasets include resources from GWAS and candidate gene-based studies, transcriptomic studies, methylation studies, literature mining, and drug-target data, as well as derived data such as spatial-temporal gene expression, promoters, enhancers, expression QTL (eQTL). For GWAS, we curated a total of 6,337 SNPs with association *p-value* < 1×10^{-4} . There were 28 SNPs common in AD and ND. In addition, 787 SNPs have been identified to be related with DNA methylation in fetal brain, and 609 SNPs are associated with eQTLs significantly in frontal cortex. There were 580,000 enhancers and promoters mapped to these SNPs (± 5 kb). Our PubMed literature mining identified ~60,000 studies with genes related to the four phenotypes. Among them, *ALDH2*, *EGFR*, *BDNF*, and *CAMP* were the most frequently studied genes for AD, ND, CD, and OD, respectively. There were 369 genes shared by three or more phenotypes, among which, 161 genes were commonly reported to all four phenotypes. We further studied drug information for these candidate genes using drug-gene information from pharmacological databases. Finally, these genes can be examined by the tissue-specific expression and eQTLs from our website. We also studied the common or specific mechanisms underlying the development of these phenotypes. All the curated data will be publicly accessed, that allows the user to retrieve, and analyze the genes and their features for AD, ND, CD, and OD.

1398W

Clinical impact of cryptic splice mutations. J. McRae¹, K. Jaganathan¹, S. Kyriazopoulou Panagiotopoulou¹, S. Fazel Darbandi², S. Sanders², K. Farh¹. 1) Illumina, Inc, San Francisco, CA; 2) University of California, San Francisco.

Many patients with rare diseases who undergo genetic testing do not receive a molecular diagnosis, despite exhaustive searches of variants in protein-coding regions. We investigated whether cryptic splice variants may contribute to this deficit. We created a model that can accurately predict whether genetic variants outside essential splice sites alter splicing. We assessed the impact of cryptic splice variants in *de novo* mutations from two rare disease cohorts. Cases with autism spectrum disorder were 1.38-fold enriched for cryptic splice mutations compared to healthy controls, whereas cases with intellectual disability were 1.54-fold enriched. We estimate that ~5% of patients with rare diseases have a pathogenic cryptic splice mutation. We also identified seven additional genes at FDR < 0.01 when these variants were included in gene enrichment tests. We attempted to validate predicted cryptic splice mutations from 36 patient samples. We performed high-depth sequencing of RNA from blood-derived cell lines and examined splice events around the predicted variant. The predictions had a high rate of validation, as 75% of the predicted variants showed altered splicing. These results demonstrate the clinical utility of cryptic splice mutations in rare disease.

1399T

Dissecting heterogeneity of neurodevelopmental disorders using seed centric genetic modules. J.C. Chow¹, M. Jensen², H. Amini³, F. Hormozdian⁴, O. Penn⁵, S. Girirajan^{2,6}, F. Hormozdian^{1,7,8}. 1) Genome Center, University of California, Davis, CA; 2) Bioinformatics and Genomics Graduate Program, The Huck Institute of Life Sciences, Pennsylvania State University, University Park, PA; 3) Department of Plant Biology, University of California, Davis, CA; 4) T.H. Chan School of Public Health, Harvard University, Boston, MA; 5) Allen Institute for Brain Science, Seattle, WA; 6) Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA; 7) MIND Institute, University of California, Davis, CA; 8) Biochemistry and Molecular Medicine, University of California, Davis, CA.

Integration of protein-protein interaction and co-expression networks permits the discovery of highly connected genes that form functional subnetworks or modules. We extended our tool MAGI for the discovery of genetic modules based on high connectivity to an input "seed" gene. We constructed modules from 21 seed genes associated with various neurodevelopmental disorders to reveal unique modules for each of these seed genes. We observed that these modules are significantly associated with disease terms corresponding to their respective seed genes. Modules seeded with epilepsy genes (e.g., *SCN1A*, *GABRA1* and *KCNB1*) are associated with long-term potentiation, synaptic vesicle cycle, and calcium signaling pathways, and modules produced by autism seed genes (e.g., *SCN2A*, *CHD8*, and *POGZ*) are associated with axon guidance, synaptic transmission, and vocal learning. We found significant enrichment of *de novo* non-synonymous mutations and overlap with coding copy number variants within these modules from large cohorts of probands with autism, intellectual disability, developmental disability, or epilepsy. Additionally, the existence of differentially acetylated peaks and differentially edited RNA editing sites in the presence or absence of neurodevelopmental disorders suggests coordinated regulation within modules. We further identify phenotypic differences among probands with *de novo* mutations inside modules, finding lower average IQ, sex differences, and macrocephaly.

1400F

Predicting the clinical impact of human mutation with deep learning. L. Sundaram^{1,2}, J. McRae², SR. Padigepati², H. Gao³, Y. Li², S. Batzoglou², X. Li², KK. Farh². 1) Department of Computer Science, Stanford University, Stanford, CA; 2) Illumina Inc, San Diego, CA; 3) National Science Foundation Center for Big learning, University of Florida, Gainesville, CA.

Millions of human genomes and exomes have been sequenced, but their clinical interpretation remains challenging due to the difficulty of distinguishing disease-causing mutations from benign genetic variation. We have trained a semi-supervised deep neural network to classify benign versus pathogenic missense mutations, that substantially improves upon existing methods. When presented with a test set of *de novo* missense mutations from patients with neurodevelopmental disorders and healthy controls, our classifier successfully separates *de novo* mutations coming from cases versus controls with 88% accuracy, about half the error rate of other methods. The deep learning network uses only primary amino acid sequence as its input and learns to assign high pathogenicity scores to residues at critical protein functional domains from sequence, without additional human annotation. We show that the network assigns the highest pathogenicity scores to mutations at crucial amino acid positions within domains, such as the DNA-contacting residues of transcription factors and the catalytic residues. Applying our classifier to stratify damaging missense mutations increases the enrichment of *de novo* missense mutations in patients with intellectual disability from 1.5-fold to 2.2-fold, close to protein-truncating mutations (2.46-fold), while relinquishing less than one third of the total number of variants enriched above expectation. This substantially improves statistical power, enabling discovery of 14 additional candidate genes in intellectual disability, which had previously not reached the genome-wide significance threshold.

1401W

Identification of mitochondrial disease variants on whole exome

sequencing data. P. Garret^{1,2,3}, V. Procaccio⁴, P. Bonneau⁴, P. Vabres^{1,5,6}, N. Houcinat^{1,2,5,7}, E. Tisserant^{1,2}, A.L. Bruehl^{1,2}, V. Carmignac^{1,5}, P. Christophe^{1,2}, A. Sorlin^{1,2,5,8}, F. Tran Mau-Them^{1,2,7,8}, A. Vitobello^{1,2}, J.M. Costa³, D. Trost³, A. Boughalem³, L. Faivre^{1,8}, C. Thauvin-Robinet^{1,2,7,8}, Y. Duffourd^{1,2,8}, # these authors have equally contributed to the work. 1) Laboratoire GAD UMR1231 Inserm – Université Bourgogne-Franche Comté – Dijon – France; 2) Unité Fonctionnelle Innovation en Diagnostic génomique des maladies rares – FHU-TRANSLAD – CHU Dijon Bourgogne – Dijon – France; 3) Laboratoire CERBA – Saint-Ouen l'Aumône – France; 4) Institut MITOVASC – UMR CNRS 6015 – INSERM1083 – Université d'Angers – Angers – France; 5) Centre de Référence maladies rares « maladies dermatologiques en mosaïque » – service de dermatologie – FHU-TRANSLAD – CHU Dijon Bourgogne – Dijon – France; 6) Service Dermatologie – CHU Dijon Bourgogne – Dijon – France; 7) Centre de Référence maladies rares « déficience intellectuelle » – centre de génétique – FHU-TRANSLAD – CHU Dijon Bourgogne – Dijon – France; 8) Centre de Référence maladies rares « Anomalies du développement et syndromes malformatifs » – centre de génétique – FHU-TRANSLAD – CHU Dijon Bourgogne – Dijon – France.

Mitochondrial DNA mutations are responsible of various phenotypes, causing partial or entirely unexplained phenotypes. Clinical spectrum is huge and concerns one (e.g. internal ear in sensorineural deafness OMIM 580000) or several organs like muscles and brain (e.g. MERRF, myoclonic epilepsy associated with ragged-red fibers OMIM 545000). When suspected, mitochondrial DNA mutations identification requires targeted strategies on affected tissues. In the context of extending use of whole-exome sequencing (WES) in rare Mendelian disorders diagnosis, we developed a bioinformatics pipeline on WES data in order to detect mitochondrial variants in parallel with the routinely used “in-house” nuclear exome pipeline, in particular to detect mitochondrial disorders in patients with no initial clinical suspicion of this group of disease. FastQ data from the mitochondrial genome were extracted from WES data in 931 patients with developmental (90%) and/or neurological (10%) anomalies. After merging, the variants were filtered out based on database presence, cohort frequencies, haplogroups and protein consequences. In parallel, some patients with previously identified mitochondrial mutations were sequenced and tested as positive controls to validate the method. Two mitochondrial causal variants were identified in 2/931 patients: a homoplasmic m.9035T-C variant in a female with ataxia and axonal neuropathy, and a m.11778G-A variant in a male with Leber optic atrophy not suspected because of a complex phenotype. Two variants of unknown significance were discussed in 3 patients: an unreported homoplasmic variant (m.16114C-A) in a child with migraine headache and cyclic vomiting syndrome, very close to other suspected variants already described with the disease; and the homoplasmic m.5567T-C variant in two individuals with cerebellar ataxia. Mitochondrial genome studies based on WES data also highlighted secondary findings in 4 patients without mitochondrial phenotype at time of the test: the m.1494C-T (1 patient) and m.1555A-G (2 patients) variants described in association with aminoglycoside-induced deafness, and a m.14484T-C (1 patient) variant responsible for Leber optic atrophy. This study demonstrates the interest of including specific mtDNA variants detection in exome pipeline that improve diagnosis, in particular in patients with neurological impairments with no clinical orientation.

1402T

Mitochondrial variant calling in over 15,000 individuals in the genome aggregation database (gnomAD). K.M. Laricchia^{1,2}, G. Tiao^{1,2}, D.G. MacArthur^{1,2}, V.K. Mootha^{1,2,3}, L. Gauthier^{1,2}, M. Shand¹, J. Soto¹, J. Emery¹, M. Lek^{1,2}, S. Calvo^{1,2}, gnomAD Consortium. 1) Broad Institute of MIT and Harvard, Cambridge, MA, United States; 2) Massachusetts General Hospital, Boston, MA, United States; 3) Howard Hughes Medical Institute, Chevy Chase, MD, United States.

Mitochondria are maternally inherited organelles that each contain many copies of the 16,569bp mitochondrial genome (mtDNA). mtDNA encodes 13 proteins involved in oxidative phosphorylation, which is required for generating ~90% of all cellular ATP. Over 400 mutations in mtDNA have been associated with human disease, and a complete survey of mtDNA variants in tens of thousands of genomes would provide a powerful resource to help researchers assess the frequency and interpretation of mitochondrial variants. We explored mtDNA variant discovery using data from over 15,000 whole genomes from unrelated individuals, aggregated as part of the Genome Aggregation Database (gnomAD; <http://gnomad.broadinstitute.org>). To date, variants in these genomes have currently only been called for the 22 autosomes and chromosome X, omitting variants in the mtDNA. mtDNA is present in thousands of copies per cell, and the gnomAD genomes typically have >3000x mtDNA coverage. A specialized variant calling methodology is therefore required, as mtDNA variants can exist at any level of heteroplasmy (i.e., the percentage of mtDNA molecules with a given variant). We applied the tool mtDNA-Server, which is optimized for identifying mitochondrial variants, detecting contamination, and determining haplogroups (maternal-line ancestry). We identified low heteroplasmic, high heteroplasmic, and homoplasmic variants among gnomAD individuals. For quality control, we assessed variant concordance across duplicate samples, checked maternal transmission rates in mother-child pairs, and calculated quality metrics for variant detection in samples mixed *in silico* at predetermined heteroplasmic levels. All individuals were also classified into haplogroups, which aids in detecting contamination and provides additional information on ancestry to complement ancestry inference based on nuclear genome variation. This research provides data that can be used to benchmark other mitochondrial variant callers. Including mtDNA variants in future releases of the gnomAD database will provide critical population frequency information to help researchers understand the role of mitochondrial variation in health and disease states.

1403F

Mitochondrial variant detection, annotation and segregation from whole-exome sequencing data. J.C. Tsai^{1,2}, A.D. Laporte¹, D. Spiegelman¹, G. Rouleau¹, L. Xiong^{1,2}. 1) McGill University, Montreal Neurological Institute, Montreal, Quebec, Canada; 2) Centre de recherche, Institut universitaire en santé mentale de Montréal, Université de Montréal, Montreal, Quebec, Canada.

Introduction: Reads containing mitochondrial DNA are the most common off-target captures when performing Whole-Exome Sequencing (WES). Even without enriching panels with mitochondrial probes, it is possible to assemble the mitochondrial (MT) genome from off-target reads at relatively high depths; many open-source pipelines and tools have been developed and adapted to this means: MitoSeek, MToolBox, mtDNA-Server, and MitoSuite. **Methods:** Two alignments to the rCRS of the MT genome were generated with Burrows-Wheeler Alignment Tool (BWA) using cleaned fastq files from Illumina paired-end sequencing. Two alignment protocols for 243 WES samples from 15 Pakistani families containing neuropsychiatric disease subjects and controls were used as input into the MToolbox, mtDNA-Server and LoFreq pipelines. Calls made by the three pipelines were combined and positions present in all three pipeline's variant call files for the entire cohort were extracted, and variant annotation was performed using variants from these common positions. The annotation step was performed using Annovar with an Ensembl gene database, a dbSNP database, and custom-created MitoMap and HmtDB databases. Haplogroups were assigned using the MToolbox pipeline. MitoSuite was used for supplementary graph generation and haplogroup confirmation. **Preliminary Results:** Among the three tested variant callers, MToolbox has the highest concordance with the overlapped variants. A total of 718 MT variants were identified, 25 of which were multi-allelic variants, from 242 of the individuals. 97 variants were classified as non-coding control region variants, 286 were non-coding mt-RNA variants, and 360 were protein-coding variants. 8,955 genotypes were observed using mtDNA-server calls at these positions, of which 40% were homoplasmic variants (>95% heteroplasmy), 5% were heteroplasmic between 50-95% and 55% were heteroplasmic between 0-50%. **Conclusions:** Mitochondrial genome assembly from our WES data is possible with reasonable data quality. Variants with varying levels of heteroplasmy and a large number of homoplasmic variants were detected in our data-set, with some variants having potentially interesting biological significance. Variant segregation with disease phenotypes within the pedigrees will be investigated. Incorporation of MT-variant-calling with the selected pipelines and the Annovar annotation step into a larger WES pipeline is underway.

1404W

Methods for detecting the role of ncRNAs in rare mitochondrial diseases.

V.A. Yopez¹, M. Gusic², C. Mertes¹, P.F. Goldberg¹, L.S. Kremer², R. Kopajtich², H. Prokisch², J. Gagneur¹. 1) Department of Informatics, Technical University Munich, Munich, Bavaria, Germany; 2) Institute of Human Genetics, Helmholtz Zentrum Muenchen, Neuherberg, Bavaria, Germany.

Mitochondrial diseases represent one of the most frequent inborn errors of metabolism. They may result from pathogenic mutations in the nuclear or mtDNA. Causal defects have been identified in 300+ genes and this number continues to grow. DNA sequencing alone allows the diagnosis of ~50% of the cases, and subsequent use of RNA-Seq has improved the success rate by 10% [1]. Yet, so far focus has been on protein-coding genes. It has become increasingly apparent that the non-protein-coding genome is of crucial functional importance for normal development. Here, we present computational approaches we have developed in order to prioritize ncRNAs with a potential role in rare mitochondrial disorders, using our cohort of 230 patients with rare mitochondrial disorders and more than 9,000 transcriptomes from the GTEx Consortium. First, we applied OUTRIDER [2], a novel statistical method to detect aberrantly expressed ncRNAs. Second, we correlated the expression of ncRNAs with the expression of mito related genes. Third, we correlated the expression of ncRNAs with maximal cellular respiration obtained using OCR-Stats [3]. Finally, we developed a supervised learning method to associate ncRNAs with similar behavior as MitoCarta, mito disease causal and respiratory chain complexes genes, using GTEx data. Our preliminary results show that there are indeed some ncRNAs showing similar expression profiles as mito related genes and that correlate positively with maximal respiration. Our final aim is to identify and functionally characterize ncRNAs that are causal of mito diseases, which would be of great interest not only from the clinical and therapeutical aspects, but also to the better understanding of gene expression regulation. 1. Kremer LS et al. Genetic diagnosis of Mendelian disorders via RNA sequencing. Nat Comm, 2017 2. Brechtman F et al. OUTRIDER: A statistical method for detecting aberrantly expressed genes in RNA sequencing data. bioRxiv 322149; doi: <https://doi.org/10.1101/322149> 3. Yopez V et al. OCR-Stats: Robust estimation and statistical testing of mitochondrial respiration activities using Seahorse XF Analyzer. bioRxiv 231522; doi: <https://doi.org/10.1101/231522>.

1405T

Condition-dependent allele-specific expression in normal and diseased donor eyes reveals potential therapeutic targets for intermediate AMD. Q. Li¹, D. Morgan², L. Owen², D. Nix¹, M. DeAngelis². 1) Huntsman Cancer Institute, University of Utah, Salt Lake City, UT, USA; 2) Moran Eye Center, University of Utah, Salt Lake City, UT, USA.

While allele-specific expression (ASE) has been demonstrated in many human tissues, the ASE architectural landscape of the human eye is poorly understood. For discovery purposes, we have carried out a survey of well characterized fresh donor eye tissue (post mortem time of ≤ 6 hours old) of 14 individuals (6 intermediate stage age-related macular degeneration (AMD) patients and 8 healthy age-matched controls) using whole exome, RNASeq and SNP chip. In silico analyses of these data are of special challenge because no existing statistical method is available for non-paired differential ASE analysis eg., between a disease and a non-disease state. Another challenge is the availability of fresh human donor eyes, particularly the human neural retina and retinal pigment epithelium (RPE)-the tissues affected by AMD, for QPCR validation is very limited. Prioritization of the most promising candidate genes not just based on a significant p-value is very much needed as well. With these caveats in mind, we have developed a comprehensive pipeline to call ASE using rigorous quality control measures to minimize false ASE calling, customized DESeq2 design to deal with non-paired differential ASE analysis and an ensemble prioritization score to further prioritize the most promising candidate genes. In fact, this pipeline discovered significant 36 SNPs in 10 genes, specific to AMD RPE vs normal RPE and 79 SNPs in 11 genes to AMD neural retina vs. normal neural retina for differential ASE. No genes had evidence of differential ASE overlap in these tissue comparisons. This demonstrates the unique contributions of each of these tissue types to the pathophysiology of AMD. Additionally, no imprinted genes demonstrated differential ASE between disease state comparisons in either tissue type. The genes demonstrating differential ASE include those that function in the retina include folic acid metabolism (*FMN1*), *TDRP* (AMD GWAS associated gene) and those in the RPE include, *CAT*, *KDR* and *ITGB8*. Finally, we demonstrate the power of our approach by validating our findings from tissues which have developmentally similar origin as the neural retina and RPE from GTEx ASE data. Blinding diseases, including AMD are the leading cause of visual loss worldwide in the aging population for which there is no cure. Identifying disease mechanisms from our study would pinpoint appropriate therapeutic targets, especially that slow or stop disease progression.

1406F

Toward clinical-grade in silico prediction: A disease-specific machine learning model derived from manually curated dataset peaks accuracy of nonsynonymous variant impact prediction in hearing loss genes. X. Li¹, F. Bu^{2,3}, Y. Wang¹, J. Cheng², H. Azaiez², Y. Lu², R. Smith³, H. Yuan². 1) GeneDock Co. Ltd., Beijing, China; 2) Medical Genetics Center, Southwest Hospital, Chongqing, China; 3) Molecular Otolaryngology and Renal Research Laboratories, University of Iowa, Iowa, US.

Facilitated by high-throughput sequencing techniques, clinical genetic testing is considered to be a routine method for disorders. However, pathogenicity interpretation of novel variants may often be ambiguous due to lack of functional assessment. Many computational prediction tools have been introduced but only considered suggestive by the ACMG/AMP 2015 guideline because of inaccurate and inconsistent results. The discrepancies were likely caused by 1) inadequately curated training data, and 2) over-comprehensive testing scope that usually includes all genes and diseases. To overcome the issues, we developed a disease-specific model named **Deafness Variant function Prediction tool (DVPred)** focusing on hereditary hearing loss (HL), a well-studied Mendelian condition featured by high prevalence and genetic heterogeneity. We selected 66 genes which were reported as casual in HL and replicated in the China Deafness Genetics Consortium (CDGC) study, then manually curated 2604 pathogenic SNVs in testing genes from literatures, and retrieved 8149 curated benign SNVs from Deafness Variation Database (DVD) to generate the training dataset. Two testing datasets were composited by newly identified pathogenic SNVs from CDGC and MORL, Iowa. Matched benign SNVs were also retrieved from DVD. 35 gene specific features, 14 population frequency features, and 18 preexisting scores from published tools were included. Gradient boosting decision tree outperformed other machine learning algorithms in 5-fold cross validation. As shown in the table, the Area under the curve (AUC) of Receiver operating characteristic (ROC) is 0.9950 for training set, 0.9976 for testing set 1 (CDGC), and 0.9781 for testing set 2 (MORL). Comparing to commonly used tools, DVPred provides highly credible result that can dramatically improve interpretation of nonsynonymous SNVs in HL. This strategy also sheds light on developing in silico prediction tool in other genetic diseases.

			AUC ROC	AUC PRC
All SNVs	DVPred	Training (n = 10735)	0.995	0.9852
		Testing 1 (n = 808)	0.9976	0.9985
		Testing 2 (n = 420)	0.9781	0.9864
	REVEL	Training	0.9618	0.8847
		Testing 1	0.9672	0.9524
		Testing 2	0.94	0.9355
	MCAP	Training	0.9198	0.7214
		Testing 1	0.9256	0.9015
		Testing 2	0.8923	0.8838
Missense SNVs	DVPred	Training (n = 9830)	0.9925	0.9663
		Testing 1 (n = 543)	0.9931	0.9924
		Testing 2 (n = 349)	0.9669	0.9734
	REVEL	Training	0.965	0.889
		Testing 1	0.973	0.9583
		Testing 2	0.9486	0.9415
	MCAP	Training	0.9193	0.7177
		Testing 1	0.9251	0.9002
		Testing 2	0.89	0.8789

1407W

A geometric morphometrics and machine learning tool for facial shape analysis and classification based on non-invasive facial biomarkers: A Down's syndrome case study. X. Sevillano¹, A. González², R. González², J. Sharpe³, R. De La Torre³, M. Dierssen⁴, N. Martínez-Abadías². 1) Grup de Recerca en Tecnologies Mèdia (GTM), La Salle - Universitat Ramon Llull, Barcelona, Spain; 2) European Molecular Biology Laboratory (EMBL), Barcelona, Spain; 3) Integrative Pharmacology and Systems Neuroscience, IMIM-Hospital del Mar Medical Research Institute, Barcelona, Spain; 4) Center for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Barcelona, Spain.

Down syndrome (DS) is a genetic disorder caused by trisomy of chromosome 21, typically associated with cognitive impairment, immunodeficiency, congenital heart disease, and facial dysmorphologies. Focusing on this latter aspect, Geometric Morphometrics (GM), a sophisticated body of statistical tools to measure and compare shapes with enhanced precision and efficiency, has been used to find facial biomarkers to discriminate between Euploid (EU) and DS facial shapes. In this work, we combine GM with machine learning (ML) to develop a software application to classify subjects as EU or SD according to the features extracted from the GM-based face shape analysis. In particular, facial shapes are described upon a set of 21 anatomical landmarks which are subsequently represented in spaces in which statistical differences between the EU and SD groups are maximized: the Procrustes space (PS) and the Euclidean Distance Matrix Analysis space (ES). In PS, shapes are described according to spatial landmarks distribution, using Principal Component Analysis to reduce dimensionality and represent shapes as points on a two-dimensional space defined by the two first principal components, which usually explain around 25-35% of the data variance. In ES, interlandmark distances are employed as shape descriptors, identifying which interlandmark distances define the significant differences between groups. Taking into account that both spaces seek maximizing between-class variance, we consider them as suitable feature spaces for training ML models that learn to classify EU and SD subjects from their facial shapes. Thus, we propose training and testing supervised classification algorithms to create a phenotyping/diagnostic tool based on the PS (using a variable number of principal components) and on the ES (using a variable number of the most significantly different interlandmark distances). Using a sample of 177 EU and 43 SD subjects of ages between 0 and 18, we trained models both in PS and ES over different ranges of age (0 to 3, 4 to 12, and 13 to 18), obtaining classification accuracies over 95% (within the same age range) and 85% (when training and testing on different age ranges). To conclude, this tool combining GM and ML is highly valuable for diagnosis and prognosis based on non-invasive facial biomarkers, which can be applied to other disorders characterized by the presence of minor facial dysmorphologies like schizophrenia, bipolar disorder or autism.

1408T

nanotatoR: An annotation tool for next generation mapping. S. Bhattacharya, H. Barseghyan, E. Vilain, EC. Delot. Center for Genetic Medicine Research, Children's National Medical Center, Washington, DC.

Whole genome sequencing (WGS) has successfully been used to identify single-nucleotide variants (SNV), small insertions and deletions and, more recently, small copy number variants. However, due to utilization of short reads, it is not well suited for identification of structural variants (SV) and the majority of SV calling tools having high false positive and negative rates. Optical next-generation mapping (NGM) utilizes long fluorescently labeled native-state DNA molecules for de novo genome assembly to overcome the limitations of WGS. NGM allows for a significant increase in SV detection capability. However, accuracy of SV annotation is highly important for variant classification and filtration to determine pathogenicity. Here we create a new tool in R, for SV annotation, including population frequency and gene function and expression, using publicly available datasets. We use DGV (Database of Genomic Variants), which contains SVs from 7965 individuals, to calculate the population frequency of the SVs identified by the Bionano SVCaller in the NGM dataset of a cohort of 50 undiagnosed patients with a variety of phenotypes. The new annotation tool, nanotatoR, also calculates the internal frequency of SVs, which could be beneficial in identification of potential false positive or common calls. The software creates a primary gene list (PG) from NCBI databases based on patient phenotype and compares the list to the set of genes overlapping the patient's SVs generated by SVCaller, providing analysts with an easy way to identify variants affecting genes in the PG. The output is given in an Excel file format, which is subdivided into multiple sheets based on SV type. Users then have a choice to filter SVs using the provided annotation for identification of inherited, de novo or rare variants. Visualization options are available for distribution of the different SV types and size ranges. An option is also available to add the read counts from RNASeq data to the annotated SVs called by Bionano SVCaller, to determine whether the change in structure results in a change in expression. Additionally we provide the user with the median expression in Transcripts per Kilobase Millions (TPM) of the gene across different tissues from the Genotype-Tissue Expression (GTEx) database. nanotatoR provides integrated annotation and the expression patterns to enable users to identify potential pathogenic SVs with greater precision and faster turnaround times.

1409F

Calling trisomic variants from whole genome sequencing data using PEMapper and PEEaller. *H.R. Johnston, M.E. Zwick, S.L. Sherman, D.J. Cutler.* Department of Human Genetics, Emory University, Atlanta, GA.

The ability to characterize loci on a trisomic chromosome is of great interest to better understand molecular risk factors of nondisjunction and its consequences. Examining loci along the chromosome arms permits in-depth analysis of the location of recombination events, parent of origin effects, and even suggests likely microdeletion regions, when the allele count appears to be 50/50. We have developed an accurate method to differentiate between the two heterozygous states (AAa vs Aaa) using sequencing data, the most challenging aspect of this characterization. Using samples from the Emory University Down Syndrome Center cohort, approximately 200 whole genomes have been sequenced to 30x depth of coverage. These samples were mapped with PEMapper and called with PEEaller¹. Starting with ~84,000 variants on chromosome 21 with MAF (a) > .01 in gnomAD, the depth of each called allele was calculated for every position for every sample. A basic set of genotype likelihood models was fit to the data, with L(AAA) representing 99% reference alleles, L(AAa) 66% reference alleles, L(Aaa) 33% reference alleles and L(aaa) 1% reference alleles. We then took these likelihoods and calculated a probability that the underlying genotype is correct. From that, we calculate the probability of the minor allele at each site, as well as the uncertainty in the estimate. Using these estimates, combined with array-based genotypes on the parents of the individuals with trisomy 21, we will be able to construct chromosomal maps that can tell us from which parent the extra copy of chromosome 21 was inherited, the type of nondisjunction error (meiosis I or meiosis II) as well as hone in on the likely recombination breakpoint(s) in that parent in which the error occurred. Additionally, variants that appear to violate the model by having 50% reference alleles will be able to be investigated further for evidence of microdeletions in the region. All of these analyses will help us to better understand the roles that dosage and compensation play in trisomy-associated phenotypes, with specific application in this initial project to Down syndrome. 1. Johnston, H.R., Chopra, P., Wingo, T.S., Patel, V., Epstein, M.P., Mulle, J.G., Warren, S.T., Zwick, M.E., and Cutler, D.J. (2017). PEMapper and PEEaller provide a simplified approach to whole-genome sequencing. *Proceedings of the National Academy of Sciences* 114, E1923-E1932.

1410W

Phenotype enrichment and analysis of hundreds of EHR derived phenotypes in > 4,000 genetically reconstructed pedigrees from a large precision medicine cohort of 92,000 individuals. *J. Staples¹, C. Gonzaga Jauregui², R. Metpally², CV. Van Hout¹, C. O'Dushlaine¹, SE. McCarthy¹, DJ. Carey², AR. Shuldiner¹, A. Baras¹, L. Habegger¹, JG. Reid¹, Geisinger-Regeneron DiscovEHR.* 1) Regeneron Genetics Center, Regeneron Pharmaceuticals Inc, Tarrytown, NY; 2) Geisinger Health System, Danville, PA.

Rare, phenotypically impactful variants are valuable for better understanding human biology, identifying drug targets, and diagnosing patients. Family-based analyses are particularly useful for identifying these variants co-segregating with phenotypes of interest, whereas these variants may not be easily detected with a population-based approach. Unfortunately, many planned and ongoing large-scale sequencing initiatives are designed to ascertain few related individuals, precluding the use of these powerful family-based analyses. However, this is not the case for the 92K individuals sequenced in DiscovEHR. We have reconstructed 4,697 pedigrees of 3 or more individuals directly from exome sequence data, and we found over 300 electronic health record (EHR) derived phenotypes segregating in a Mendelian fashion amongst these pedigrees, providing over 2,000 potentially informative pedigree-phenotype pairings that enable traditional Mendelian analyses at a large scale. For each individual in these pedigrees, ICD10 three-digit codes were used to define affecteds of binary traits, and affecteds with extreme values for quantitative traits were defined as having a trait value more than two standard deviations from the cohort's mean. We used these sets of affected individuals to identify pedigrees amenable to family-based rare variant segregation analysis, and more than half of the segregating traits were segregating in two or more pedigrees. A preliminary rare variant segregation analysis identified likely causal mutations in known disease-causing genes for hereditary hemorrhagic telangiectasia and motor sensory neuropathy. We also found pedigrees segregating known, highly-penetrant, and actionable Mendelian disease-causing variants for diseases such as familial aortic aneurysms, thyroid cancer, pigmentary glaucoma, and familial hypercholesterolemia. Family-based rare variant segregation analyses are ongoing for these pedigrees to identify novel genes for disease-related traits; for example, we have preliminary evidence for a new candidate gene for bipolar disorder. While many precision medicine cohorts are not designed to directly ascertain pedigrees, we have demonstrated the value of obtaining informative pedigrees directly from the genetic data of a large precision medicine cohort to create a large dataset for traditional Mendelian analyses.

1411T**Scaling rare disease genomics to tens of thousands of samples.** *B.*

Weisburd^{1,2}, *H. Snow*^{1,2}, *H. Arachchi*^{1,2}, *S. Baxter*^{1,2}, *A. O'Donnell-Luria*^{1,2,3}, *H.L. Rehm*^{1,2}, *M.J. Daly*^{1,2}, *D.G. MacArthur*^{1,2}. 1) Broad Institute, Cambridge, MA; 2) Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA; 3) The Manton Center for Orphan Disease Research, Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA, USA.

Rapid advances in exome and whole-genome sequencing make it possible to identify the full spectrum of genetic variation in rare disease patients and family members. However, to make sense of this data, and discover the causal variants underlying rare diseases, powerful filtering and decision support tools are needed that can handle the very large data sets now being produced. We are addressing this challenge by developing seqr - an open source, interactive web-based tool that allows researchers to work together to search and annotate variant call-sets from exome and whole genome studies involving families of rare disease patients. The seqr platform has already been used by over 500 researchers to study potential causal variants across 10,000 families affected by rare disease, and is the core analysis platform for the Broad Institute's Center for Mendelian Genomics facilitating over 1,000 diagnoses. This presentation will describe the technology stack underlying seqr, including the database, server-side and client side technologies involved, with a focus on the design choices made to facilitate the scaling of the system to tens of thousands of exome and genome samples. The need to allow interactive searching of terabytes of genomic sequencing data requires custom solutions and thoughtful design trade-offs. We will present the technical details of how seqr dealt with these challenges. Additionally we will describe how interactive visualization of read-level data, copy number variation, detailed phenotype data, and other relevant data types were brought together in a single user-friendly interface. .

1412F**Construction of a structural variant detection pipeline for the Undiagnosed Diseases Program.** *B. Hospelhorn, D. Adams, W. Gahl.*

Undiagnosed Diseases Program, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD.

INTRODUCTION: The Undiagnosed Diseases Program (UDP) is celebrating its 10th year of seeking diagnoses for individuals with illnesses that remain undiagnosed despite extensive medical workup. Structural DNA variations (SVs) can have a significant impact on an organism's phenotype and are an important subset of candidates to consider when searching for causal variants in genetic disease. Genomic SVs that are copy number balanced or lie between the cytogenetic and short-read-sequencing size optima, have proven difficult to detect through computational means, even after the advent of Next Generation Sequencing. Overall, many software packages are available, but the field has not settled on an optimal analytic approach. We describe initial efforts to develop an SV detection pipeline for our cohort of undiagnosed UDP study participants. **METHODS:** In an effort to consider multiple lines of evidence, the current UDP SV pipeline utilizes multiple SV and CNV callers such as LUMPY and DELLY, and integrates CNV calls from SNP array data and indel calls from standard short-read sequence genotyping for cross-referencing. **RESULTS:** Preliminary results indicate progress toward the prioritization of deleterious variants. Genotyping alone can reduce the number of variants by as much as 75%, and forcing a call set consensus can further reduce that number. Additionally, initial efforts at prioritizing deletions, exonic variants, and segregating variants have proven effective at culling variant lists to almost manageable lengths. Encouragingly, the pipeline was able to detect the causal structural variant in a solved UDP case, a variant which was initially detected from SNP data, but was filtered during the quality control process. Even at this prototypical stage, the SV pipeline shows promise in its ability to pick up variants larger than the upper limit of standard NGS genotype calling (under 500bp), and resolve SVs in the 100-100000bp range far more reliably than can be done with SNP data alone. **CONCLUSION:** Exome and genome sequencing fail to provide a definitive diagnosis in a substantial number of individuals with characteristics of inherited disease. As the SV-detection field evolves, generation of new methods will need to be incorporated in the reanalysis of undiagnosed cohorts. The UDP SV detection pipeline is being positioned to fulfill that role for the NIH Undiagnosed Diseases Program.

1413W

Precision etiology for complex disease classification. *L.M. Schriml, C. Greene, C. Bearer, R. Lichenstein, R. Tauber, J. Munro, M. Giglio.* University of Maryland School of Medicine, Baltimore, MD.

Beyond monogenic diseases, clinical diagnosis is challenged by the complexity of etiologies for many genetic diseases. The complexity of etiologies for differential diagnosis challenges established bioinformatic knowledge capture in biomedical ontologies. Classification of complex disease necessitates an evolution towards precise disease etiology captured through clinically-defined semantic assertions. Here we describe how the Human Disease Ontology (DO) project (www.disease-ontology.org) is addressing the need to rigorously incorporate genetic etiology and environmental drivers through strictly defined semantic logical definitions (with OBO Foundry ontologies), precisely defining that a disease has exactly 1, a minimum (or maximum) of [x number] of (functional or structural) alteration(s); symptom(s); transmission method(s); infectious agent(s); environmental trigger(s); phenotype(s); anatomical location(s); cell type(s) of origin; food-based allergic trigger(s); or genetic susceptibility. As the OBO Foundry's domain ontology for human diseases, the DO represents the breadth of (~ 16,000) common and rare diseases in an etiology-based classification as both an axiomatized OWL and an OBO formatted ontology. The DO semantically integrates (> 46,000) disease and medical vocabulary terms through extensive logically-defined cross mapping (MeSH, ICD, NCI's thesaurus, SNOMED and OMIM). The DO provides standardized descriptions of human disease improves the capture and communication of health-related data across several hundred resources including biomedical databases, bioinformatics tools, genomic and cancer resources. Addressing these challenges the DO project has developed a complex disease model to drive the restructuring of DO knowledge for differential diagnosis. The DO's clinical team has assessed a set of complex and environmental diseases to expand DO's complex disease knowledgebase. Initially, examining the pleiotropy of a dozen diseases including: Prader-Willi syndrome, which can be a chromosomal deletion, a methylation defect or a single gene disorder; alpha 1-antitrypsin deficiency, which has variable expression and critical contributions from environmental factors; and cystic fibrosis, which involves multiple organ systems in a single disorder. The outlined enhancements are enabling precise incorporation of genetic and environmental components of disease etiology for complex diseases and thus the utility of the DO for clinical diagnosis.

1414T

The use of AI technologies in genomic interpretation, a pilot study. *R. Attali¹, S. Tzur¹, O. Farchy¹, T. Talmy¹, Y. Yang^{2,3}, C. Eng^{2,3}, C. Shaw², F. Xia^{2,3}, BG WES review group; BG Bioinformatics group.* 1) Emedgene Technologies, Tel Aviv, Israel; 2) Baylor Genetics, Houston, Texas, USA; 3) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, USA.

Introduction Due to the rapid growth in the use of genome sequencing in medicine, the interpretation workload for genome specialists has amplified dramatically. Hence, rises a need for novel solutions aimed at reducing analysis time, by implementing advanced technologies to provide autonomic findings and thereby support rapid clinical decision making. In this pilot study, we assessed the use of artificial intelligence (AI) technologies in diagnostic genomic analysis. **Methods** The pilot was performed by the experienced genomic specialists team of Baylor Genetics (BG) while applying technologies rooted upon machine learning and natural language processing (NLP), developed at Emedgene Technologies. The BG team performed paired analyses of 79 cases. The first analysis was performed using BG's standard operating procedure (SOP), whereas the second implemented the capabilities and additional findings provided by the AI pipeline. The auto-analysis output consisted of several dozen prioritized candidate variants, backed by a literature synopsis displaying relevant publications as part of a comprehensive graphical representation of supporting evidence, to be used for manual inspection. **Results** All notable variants reported for high impact or further investigation using the regular SOP procedure were successfully identified in candidate lists generated by the AI pipeline, barring one unsolved case, where a monoallelic nonsense variant in an AR gene was missed. The number of variants identified for review per case was significantly reduced when incorporating the AI pipeline as compared to regular pipeline (Mean 79%, range 59%-94%). Median analysis time was reduced by more than 47% per case for genomic specialists. Interestingly, the time saving effect looks to be more significant for less experienced genomic specialists. **Conclusions** Using novel AI technologies, genomic analysis was completed in a time efficient and evidence-based manner. Genomic specialists now can work more efficiently as a result of having a "virtual geneticist" at their side, performing a preliminary assessment of sequencing data and identifying suspicious variants. Moreover, the use of AI solutions provided a higher consistency in results between different team members. AI solutions could be an integral part of genomic diagnosis and are expected to enable full automatic interpretation and re-analysis of cases for the future era of precision medicine and mass population sequencing.

1415F

A novel machine learning algorithm for variant prioritization using genotype, phenotype and tissue-specific expression data. *L. Lange^{1,2}, A.T. Pagnamenta^{1,2}, D. Sims³, J.C. Taylor^{1,2}, Health Innovation Challenge Fund 2 (HICF2) Project Team 1, 2.* 1) National Institute for Health Research Biomedical Research Centre; 2) Wellcome Trust Centre for Human Genetics (WTCHG), University of Oxford, Oxford OX3 7BN, UK; 3) MRC Weatherall Institute of Molecular Medicine Centre for Computational Biology, John Radcliffe Hospital, University of Oxford, Oxford OX3 9DS, UK.

Background: Variant prioritization algorithms (VPA) are a widely used approach for clinical whole genome sequencing (WGS) analysis from rare disease (RD) patient samples. Existing algorithms combine genomic data and phenotypic terms in the Human Phenotype Ontology (HPO) format to rank variants. We developed a novel machine learning VPA that also leverages expression data for tissues assumed to be affected based on the patient's HPO profile. **Methods:** A dataset (N=50,514), consisting of 30% disease-causing variants from OMIM and 70% likely benign variants from the 1,000 Genomes project was used for training. Variants were annotated with Exomiser's pathogenicity and phenotype scores. HPO terms for each variant were used to create binary labels (BLs) for 25 GTEx tissues, capturing if a tissue was likely affected or unaffected by the phenotype. Cross-tissue and cross-gene scores, indicating relative gene expression across tissues and within tissues, were generated for each gene using GTEx data. To test if BLs improved performance, we trained a logistic regression classifier on the dataset, with and without binary vectors. Further, we trained the classifier on the entire dataset, benchmarking against Exomiser. We compared our algorithm's performance with Exomiser's on four RD patient cases from our WGS cohort. Strong candidate variants in novel genes, identified manually by human disease experts, were used as benchmarks. **Results:** With an accuracy of 0.84, BLs significantly improved the expression-based signal (accuracy = 0.70). Combining BLs, expression data and Exomiser scores, our algorithm outperformed Exomiser (accuracies of 0.96 and 0.91). Both algorithms ranked two out of four RD case benchmark variants in the top 10. Exomiser performed better for candidate genes that were novel for the patients' phenotype, but had previously been annotated with HPO terms. Our algorithm improved performance for candidate genes that had not been associated with disease phenotypes. **Conclusions:** Our results suggest that the HPO implicitly contains valuable information on tissue-specific expression of disease-related genes. Combining BLs with genomic and phenotypic data can improve VPA accuracy. Importantly, our algorithm performs well for novel genes with no known phenotypic annotations, demonstrating a powerful additional use of the HPO. Efforts are underway to apply our algorithm to more RD cases. Our pipeline will be made available on GitHub.

1416W

A statistically-based, heuristic for allele-specific, position-based error modeling of Illumina next-generation sequencing data for rare mosaic variant calling in unmatched samples. *J.N. Dudley, J.J. Johnston, L.G. Biesecker.* Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, NIH, Bethesda, MD.

Mosaic variant calling for single-gene somatic disorders (e.g., Proteus and PROS) is a fundamentally different problem than in cancer. In these disorders, clinical limitations in sampling and tissue heterogeneity can result in a wide range of variant allele fractions (VAF), from <0.1 to 50%, that are clinically relevant to diagnosis. In cancer, VAF detection of <10% is generally not relevant or useful. Most mosaic callers are designed for the cancer application and are therefore not sensitive or useful for VAF <10%. Next-generation sequencing (NGS) platforms are prone to misincorporations as the main source of error due to sequence context and motifs, including non-random errors. Thus, differentiating low level biological VAF from NGS error is critical. This requires evaluating the non-reference base calls not only within their sequence context, as related to their genomic position, but also the identities of the variant and reference alleles and their presence in a control genome and/or population. In this work, we provide clarity to this problem by using an allele-specific, position-based modeling approach for assessing VAF. To build our model, exome sequencing (ES) data from 1,017 patients in the ClinSeq® cohort was used to obtain total depth and total allele counts at every targeted position in a mosaic overgrowth gene panel. The depth and allele-specific VAF count of candidate test variants were then screened against the modeled position using a Fisher's Exact Test followed by multiple test correction. This approach addresses multiple concerns in regard to NGS analysis of rare somatic mosaic variation. First, low frequency variants are distinguished from background variation from sequencing. Second, variants that are common in the model population are filtered out. And finally, it eliminates the need to run matched normal samples that are required for many somatic variant callers. Using this approach to probe the most common causative variants in PROS with data produced by a MiSeq we identified low level variants with a positive predictive value greater than 60%. We found that at a depth of 1500, which is typical using standard flow cells with v2 chemistry, our limit of detection ranged from 0.5 to 0.7%.

1417T

Short tandem repeat expansions in undiagnosed rare genetic disease cases. P.A. Richmond^{1,2,3}, B.I. Drögemöller^{1,3}, I.S. Rajan-Babu^{1,3,7}, M. Tarailo-Graovac^{4,5}, A.B.P. van Kuilenburg⁶, C.J. Ross^{1,3}, C.D. van Karnebeek^{2,6}, W.W. Wasserman^{1,2,3}. 1) BC Children's Hospital Research Institute, Vancouver, BC, Canada; 2) Centre for Molecular Medicine and Therapeutics, University of British Columbia, Vancouver, BC, Canada; 3) Departments of Pharmaceutical Sciences, Paediatrics, and Medical Genetics, University of British Columbia, Vancouver, BC, Canada; 4) Alberta Children's Hospital Research Institute, University of Calgary, Calgary, AB, Canada; 5) Departments of Biochemistry, Molecular Biology and Medical Genetics, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada; 6) Departments of Clinical Chemistry, Pediatrics and Clinical Genetics; Laboratory Genetic Metabolic Diseases, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands; 7) BC Children's & Women's Hospital, Vancouver, BC, Canada.

Exome sequencing has been used to diagnose rare genetic diseases for the past few years by identifying genomic variants, single nucleotide variants (SNVs) as well as short insertions and deletions (indels), over exonic regions. While it is a powerful tool, current diagnosis rates of 25-45% lead to speculation that variants beyond the exome are responsible for undiagnosed disease cases. An alternative, PCR-free whole genome sequencing (WGS), is now being used in place of exome sequencing in some clinical research endeavours. WGS offers the distinct advantage of being able to identify variation beyond the exome, and can additionally detect other forms of genomic variation including microsatellite repeats, copy number variants, and complex structural variants. Short tandem repeats (STRs) have been implicated in genetic diseases, involving both expansion and contraction of repeat numbers. These repeats, herein defined as having unit lengths of 1-6bp, impact genes via both protein-coding and regulatory mechanisms. Trinucleotide STR loci have been implicated prominently in genetic disease, including coding expansions within *HTT* leading to Huntington's Disease, and 5' untranslated region expansions causing repression of *FMR1* in Fragile X Syndrome. Across the human genome are many potentially functional repeats that may contribute to disease risk. We investigated the potential for STRs to explain undiagnosed rare genetic disease cases. To do so, we first identified and annotated STRs throughout the genome, and tailored a bioinformatic pipeline that uses the ExpansionHunter tool to detect them from patient WGS data. Running this pipeline on 300 WGS samples allowed us to construct an in-house database to better establish the frequency of variation in the repeats for the experimental population. Next, we applied the informatics pipeline to undiagnosed cases of rare genetic disease WGS, including cases with extensive phenotypic data (gene expression and metabolite profiles) and strong candidate genes. This led to diagnosis of two cases with a STR expansion in the 5' UTR of a novel disease gene, which were further confirmed by triplet-primed PCR. As we continue to apply our pipeline to undiagnosed cases, we may identify new candidate STR expansions within genes that explain disease phenotypes. The bioinformatics pipeline is available on GitHub to enable the community to explore how genotyping STRs from WGS data can aid in the diagnosis of rare genetic diseases. .

1418F

Vegetarian diet-modulated epigenetic clock and longevity. X. Chen¹, W. Chen¹, K. Jaceldo-Siegl², V. Filippov², P. Duerksen-Hughes³, C. Wang¹. 1) Center for Genomics, Loma Linda University, Loma Linda, CA; 2) Center for Nutrition, Loma Linda University, Loma Linda, CA; 3) Basic Sciences, Loma Linda University, Loma Linda, CA.

Diet importantly impacts health, disease resistance, and longevity, but the molecular mechanisms mediating the effects of diet are poorly understood. The major reason for this profound knowledge gap is that we don't yet have a comprehensive understanding of how diet influences specific epigenomes. To address this knowledge gap, we exploit a powerful, well-characterized, long-standing cohort that has already made many significant advances: the Adventist Health Study 2 (AHS2), which constitutes a multi-ethnic group of participants whose detailed diets are known. Our CENTRAL MODEL is that long-term dietary patterns cause epigenomic reprogramming in ways that alter disease susceptibility, maintenance of health, and longevity. We performed DNA methylome studies using Illumina 450K BeadChips, TruSeq Epic next-gen sequencing (NGS) technologies and the frozen human white blood cells derived DNA from AHS2 participants consisting of three different groups: vegan, pescetarian and non-vegetarians. Age, gender, race, BMI, exercise, drinking and smoking status were balanced in each group. We determine the impact of dietary factors on both Horvath and Hannum epigenetic clocks, which are algorithms that use DNA methylation patterns to estimate biological age of specific cells and tissues. We also developed a new epigenetic clock, which, unlike Horvath and Hannum epigenetic clocks that only work for the array-based data, can measure the DNA methylation age using both array and NGS data. Our epigenetic clock shows consistent better prediction of biological age and performance metrics than Horvath and Hannum clocks using many benchmark samples. Overall, our results show that a lifetime vegetarian diet can reprogram the epigenomes, which contributes approximately 5 extra years of lifespan compared to non-vegetarians based on our AHS2 cohort. Our pilot DNA methylome study support our HYPOTHESIS that maintaining a specific diet causes epigenomic reprogramming leading to programming of health and longevity in humans.

1419W

Machine learning delivers enhanced genomic prediction of osteoporotic fracture risk. J. Keller-Baruch¹, M. Forest², V. Forgetta², A. Durand³, J. Kemp^{4,5}, J.A. Morris¹, J.A. Kanis⁶, D.P. Kiel⁷, E.V. McCloskey⁸, H. Johansson⁹, N. Harvey^{9,10}, D.M. Evans^{1,5}, J. Pineau³, W.D. Leslie^{11,12}, C.M.T. Greenwood^{1,2,13,14}, J.B. Richards^{1,2,15}. 1) Department of Human Genetics, McGill University, Montréal, Québec, Canada; 2) Centre for Clinical Epidemiology, Lady Davis Institute, Jewish General Hospital, McGill University, Montréal, Québec, Canada; 3) School of Computer Science, McGill University, Montréal, Québec, Canada; 4) University of Queensland Diamantina Institute, Translational Research Institute, Brisbane, Queensland, Australia; 5) MRC Integrative Epidemiology Unit, University of Bristol, UK; 6) Centre for Metabolic Bone Diseases, University of Sheffield, Sheffield, UK, and Australian Catholic University, Melbourne, Australia; 7) Institute for Aging Research, Hebrew SeniorLife, Harvard Medical School, Broad Institute of MIT & Harvard; 8) Mellanby Centre for Bone Research, Centre for Integrated Research in Musculoskeletal Ageing, University of Sheffield, and Sheffield Teaching Hospitals Foundation Trust, Sheffield, UK; 9) Medical Research Council Lifecourse Epidemiology Unit, University of Southampton, Southampton SO16 6YD, UK; 10) National Institute for Health Research Southampton Biomedical Research Centre, University Hospital Southampton, NHS Foundation Trust, Southampton SO16 6YD, UK; 11) Department of Medicine (Endocrinology), University of Manitoba, Winnipeg, Manitoba, Canada; 12) Department of Radiology (Nuclear Medicine), University of Manitoba, Winnipeg, Manitoba, Canada; 13) Gerald Bronfman Department of Oncology, McGill University, Montréal, Québec, Canada; 14) Department of Epidemiology, Biostatistics & Occupational Health, McGill University, Montréal, Québec, Canada; 15) Department of Twin Research and Genetic Epidemiology, King's College London, London, United Kingdom.

Background: Genomic prediction of disease risk factors might be helpful since genome-wide genotyping costs are lower than many clinical tests. We tested whether machine learning (ML) methods could provide clinically-relevant prediction of a highly-heritable trait, quantitative ultrasound measurement of bone speed of sound (SOS)—a risk factor for osteoporotic fracture. **Methods:** We used 298,768 individuals from UK Biobank with SOS measures at the calcaneus and genome-wide genotypes in lasso-regularized regression models to genomically predict SOS (gSOS), followed by model selection and model testing in 42,682 and 85,362 separate individuals also from UK Biobank, respectively. We then tested if gSOS predicted incident major osteoporotic fractures derived from questionnaire and ICD10 coding and whether genomic pre-screening could complement FRAX, a widely used osteoporotic fracture risk prediction tool. **Results:** Age, sex and body mass index explained 4.2% of variance in SOS. gSOS was correlated with measured SOS ($r = 0.49$, $P < 10^{-140}$) and explained 25% of its variance. gSOS was uncorrelated with clinical risk factors included in FRAX. In the test dataset, after adjusting for FRAX clinical risk factors (other than family history of hip fracture, which was not available), the association between gSOS and incident major osteoporotic fractures (2,030 fracture cases/83,332 controls, odds ratio = 1.33 [1.27–1.39]) was of slightly lower magnitude than for measured SOS (OR = 1.49 [1.41 – 1.57]). Individuals in the bottom 2.5% of gSOS had a 4-fold increase in the odds of fracture compared to the top 2.5%. A model combining gSOS with calculated FRAX score (without bone mineral density), yielded improved outcome discrimination compared with FRAX alone (area under the receiver operator curve = 0.65 [0.64–0.66] vs 0.63 [0.62–0.64] respectively) and improved the sensitivity 1.6-fold. Last, introducing a gSOS pre-screening step into the UK National Osteoporosis Guideline Group screening algorithm resulted in a reduction in the number of people requiring screening by 70% and the number of people requiring BMD testing by 50%, while only reducing the sensitivity of current guidelines to identify individuals eligible for therapy from 99% to 95%. **Summary:** ML-based genomic prediction of SOS resulted in a clinically-relevant prediction of fracture. This information could increase the efficiency of screening programs for osteoporotic fracture prevention.

1420T

Exonic variants in aging-related genes are predictive of biological aging status. M.E. Breitbach^{1,2}, S. Greenspan³, N.M. Resnick^{3,4}, S. Perera^{3,5}, A.U. Gurkar³, D. Absher⁶, A.S. Levine^{6,7}. 1) HudsonAlpha Institute for Biotechnology; 601 Genome Way, Huntsville, AL 35806; 2) University of Alabama in Huntsville; 301 Sparkman Drive, Huntsville, AL 35899; 3) Division of Geriatric Medicine, Department of Medicine, University of Pittsburgh School of Medicine; ; 3471 Fifth Avenue, Kaufmann Medical Building Suite 500, Pittsburgh, PA 15213; 4) Aging Institute of UPMC; 3600 Forbes Avenue at Meyran Forbes Tower Suite 10065 Pittsburgh, PA 15213; 5) Department of Biostatistics, University of Pittsburgh Graduate School of Public Health; 130 DeSoto Street, Pittsburgh, PA 15213; 6) Department of Microbiology and Molecular Genetics, University of Pittsburgh School of Medicine, 450 Technology Drive, Pittsburgh, PA 15219; 7) UPMC Hillman Cancer Center; 5117 Centre Avenue, Pittsburgh, PA 15213.

Recent studies investigating longevity have revealed very few convincing genetic associations with increased lifespan. This is, in part, due to the complexity of biological aging as well as the limited power of genome-wide association studies, which assay common single nucleotide polymorphisms (SNPs) and require several thousand subjects to achieve statistical significance. To overcome such barriers, we performed comprehensive DNA sequencing of a panel of 20 genes previously associated with aging, in a cohort of 200 individuals, half of whom were clinically defined by accelerated aging phenotypes ("early agers"), and half of whom were defined by reduced aging phenotypes ("late agers"). A validation cohort of 511 "late" agers was used to verify our results. We found early agers were not enriched for more total variants in these 20 genes than late agers. To identify variants or genes predictive of "late" aging, both logistic regression analysis and random forest-based machine learning were performed on the genetic variants identified across the 20 aging-associated genes. A variety of models were tested, and we found the most predictive model incorporated damaging exon variants (CADD > 15). Furthermore, the most heavily weighted variants were within genes involved in repair of DNA damage and maintenance of genomic stability. Overall, this study implemented a framework to apply machine learning to identify sequencing variants associated with aging and improves our understanding of how genetic variation contributes to the aging process.

1421F

Predicting splicing from primary sequence with deep learning. *K. Jagathan¹, S. Kyriazopoulou Panagiotopoulou¹, J.F. McRae¹, D. Knowles², Y. Li², W. Cui¹, J. Hakenberg¹, S. Batzoglou¹, K. Farh¹.* 1) Illumina Inc, San Diego, CA; 2) Stanford University, Stanford, CA.

At present, only 25-30% of rare disease patients receive a molecular diagnosis from examination of protein-coding sequence, suggesting that the remaining diagnostic yield may reside in the noncoding regions of the genome. Noncoding mutations that disrupt the normal pattern of pre-mRNA splicing despite lying outside the essential AG or GT splice dinucleotides, also known as cryptic splice variants, are recognized to play an important role in rare genetic diseases, but have been difficult to identify since the genomic space where they may occur is very large. To facilitate the interpretation of these variants, we developed a novel deep learning network that accurately predicts splice junctions from arbitrary pre-mRNA transcript sequence. Examination of the specificity determinants of our model reveals that it recapitulates well-studied mechanisms that confer spliceosome specificity, such as the interplay between motif strength, exon/intron lengths, and nucleosome positioning. Predicted cryptic splice variants validate on RNA-seq at nearly the same rate as essential splice disruptions, demonstrating that most of these predictions are functional. Using data from the ExAC and gnomAD cohorts, we show that synonymous and intronic variants with predicted splice-altering effects are negatively selected at a rate similar to stop-gain and frameshift variants, indicating that they are strongly deleterious. We estimate that each person carries an average of ~6 rare functional cryptic splice variants with less than 0.1% allele frequency, compared to ~9 rare protein-truncating variants, and that cryptic splice variants outnumber essential AG/GT splice disruptions roughly 3:1. These results point to a significant role of cryptic splice variation in human disease and show that our model is a powerful tool towards interpretation of the noncoding genome.

1422W

Bioinformatic analysis of TE-fusion internal exons and nonsense-mediated mRNA decay within human genome. *KH. Park.* Department of Core Facility Management Center, Korea Research Institute of Bios, Daejeon, South Korea.

To better understand the mechanism of genetic diversity and control of gene expression, we conducted a large scale analysis of reliable exonization of transposable elements (TE) in non-protein coding regions of the human genes. To study the connection between the TE-fusion internal exons and nonsense mediated mRNA decay, we used the publicly available mRNA and EST to genome alignment data from the UCSC genome browser. In this study, we systematically indentified TE which can lead NMD from TE-fusion internal exons. With this procedure, we identified NMD putative 519 transcripts that are derived from 321 known genes. The result shows that TEs can make premature termination codon (PTC) to nonsense-mediated decay (NMD) when they are analyzed with human expressed sequence tag and mRNA sequences dataset. In 519 cases (26.8%) of 1931 TE-fusion internal exons that are located within the protein coding region, the insertion of a TE-fusion internal exon results in a shortened protein, through an in frame stop codon within the TE-fusion internal exon itself. The vast majority of the TE-fusion internal exons that appeared within the coding region of mRNA caused a premature termination codon or a frame shift. We believe this discovery can provide biologists insight into a function of TEs as sources of nonsense-mediated decay, which are able to affect phenotypes including diseases. We suggest that TE to NMD process is a mechanism contributing to the evolution of gene regulation in human genome as well as having implications to diseases.

1423T

Conditional generating adversarial networks for synthetic MRI brain images, semantic segmentation and annotation. Q. Ge^{1,2}, W. Lin¹, M. Xiong², Y. Liu², R. Jiao², W. Pen², Z. Li². 1) School of Mathematical Sciences, Fudan University, Shanghai, CHINA; 2) Department of Biostatistics, The University of Texas School of Public Health, Houston, TX.

Deep learning has been successfully applied to image analysis by using large-scale annotated image datasets. However, obtaining a large number of medical images poses a great challenge. To meet the challenge facing us, we propose to use conditional generative adversarial networks for generating a large number of realistic, synthetic medical images with high quality. The synthesized images can be tuned by adjusting discrete and continuous phenotypes such as sex, blood pressure and cognitive scores. Further analysis for these synthetic brain images has been conducted. Since semantic segmentations of medical images can be considered a kind of image mapping, an image-to-image generative adversarial network is used to create semantic segmentation maps on images. The proposed methods were applied to ADNI dataset with MRI images. The results were very encouraging.

1424F

Challenges in calling CNVs from low read-depth genome sequencing data for products of conception. J.W. Ahn¹, B. Wild¹, C. Donaghue³, D.T. Bonthron², C.M. Watson², K. Mann³. 1) Genetics Laboratories, Guy's Hospital, London, UK; 2) MRC Medical Bioinformatics Centre, Leeds Institute for Data Analytics, University of Leeds, Leeds, UK; 3) Genetics Laboratories, Viapath, London, UK.

Products of conception (POC) and fetal tissue (FT) samples are routinely tested for CNV to investigate the cause of fetal abnormality and/or demise in order to inform future recurrence risk. Our laboratory currently provides a CNV detection service (1Mb resolution) for POC using cytogenomic microarrays (CMA). However, it is not always possible to reach this resolution due to the poor quality of DNA in these samples. Low read-depth genome sequencing (GS) has been demonstrated to detect CNV and has been proposed as a cost-effective, resilient alternative to CMA. We therefore assessed the performance of this technology in comparison to the current standard of care by sequencing samples with known CNVs. Samples were sequenced to an average read depth of 5X (2 x 100bp) and aligned to genome build hs37d5. The resulting BAM files were then processed with a number of CNV callers to evaluate performance. These included Canvas, Lumpy, Manta, Breakseq2, Breakdancer, Delly, and CNVnator. We also investigated a CNV caller developed by the Leeds MRC Medical Bioinformatics Centre. Overall, low read-depth GS performed similarly to CMA for CNVs >0.5Mb, however, issues surrounding mapping repetitive regions and breakpoint accuracy need to be resolved. We plan to test a refined pipeline on POC/FT samples with known CNVs to further develop this method. Our aim is to demonstrate whether the advantages of low read-depth GS for CNV detection offer an improved, low-cost service in comparison to CMA.

1425W

Rapid and reproducible data processing with CWL, Toil and the Sentieon Tools. R. Aldana, D.N. Freed, X. Liu, Z. Li, L. Chen. Sentieon Inc., Mountain View, CA.

Rapid whole-genome sequencing is essential in certain clinical contexts such as infectious disease and neonatal care, where rapid WGS has been shown to decrease infant morbidity in certain patient populations. Distributed data processing pipelines have the ability to shorten analysis times, but these can be complex to implement and maintain. Recently, several workflow tools and languages have been developed to improve the reproducibility and portability of scientific data processing pipelines, including the Common Workflow Language (CWL), the Workflow Description Language (WDL), Nextflow and Snakemake. In this work, we implement Sentieon's distributed DNaseq pipeline for rapid analysis of NGS data using Common Workflow Language and demonstrate the performance of our implementation on the Google Cloud Platform with the Toil workflow engine. Our pipeline is capable of processing a standard 30x Illumina WGS sample in 2.5 hours on a single 64-core machine, while distributed processing on ten machines can reduce the turn-around time to under 30 minutes. Our CWL pipelines are open-source and are available from <https://github.com/Sentieon/Sentieon-cwl>.

1426T

Findings from the Critical Assessment of Genome Interpretation, a community experiment to evaluate phenotype prediction. G. Andreoletti¹, R.A. Hoskins¹, J. Moulton^{2,3}, S.E. Brenner¹, CAGI Participants. 1) University of California, Berkeley, CA 94720, USA; 2) Institute for Bioscience and Biotechnology Research, University of Maryland, 9600 Gudelsky Drive, Rockville, MD 20850; 3) Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, MD 20742.

Interpretation of genetic variants plays an essential role in the analysis of cancer and monogenic disease, and increasingly also of complex trait disease, with applications ranging from basic research to clinical decisions. The field has lacked a clear consensus on the effectiveness of current methods for these different uses. The Critical Assessment of Genome Interpretation (CAGI, 'kā-jē) is a community experiment to objectively assess computational methods for predicting the phenotypic impacts of genomic variation. CAGI participants are provided genetic variants and make blind predictions of resulting phenotype. Independent assessors evaluate the predictions by comparing with experimental and clinical data. CAGI has completed four rounds of challenges with the goals of establishing the state of art in genome interpretation and of encouraging new methodological developments, with the last round having a special issue of *Human Mutation*. Predictions for the fifth round of CAGI, comprised of 14 challenges, have been submitted. Challenges for CAGI5 include nonsynonymous protein variants with activity, growth, stability, and phenotypic assays; cancer protein variants with integrative assays and case-control studies; variants impacting expression and splicing, and their molecular impact; complex disease exomes; clinical panel and genomes with associated diseases. Assessment for CAGI5 is underway and will be completed in July 2018. There have been notable discoveries throughout the CAGI experiments. Independent assessment has found that top missense prediction methods are highly statistically significant, but individual variant accuracy is limited. Missense methods tend to correlate better with each other than with experiment. Bespoke approaches often enhance performance. Interpretation of non-coding variants shows promise but is not at the level of missense. In challenges using clinical data predictors identified causal variants overlooked in the initial clinical pipeline analysis. CAGI results are increasingly used to inform clinical use of computational tools, including the ClinGen working group comments on the ACMG guidelines for appropriately considering evidence from computational approaches. For example, CAGI results suggest that running multiple uncalibrated methods and considering their consensus may result in undue confidence in a pathogenic assignment. Detailed information about CAGI may be found at <https://genomeinterpretation.org>.

1427F**Not by systems alone: Identifying functional outliers in rare disease**

pedigrees. S. Ballouz, M. Doerfel, J. Crain, M. Crow, G.J. Lyon, J. Gillis. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Studies of rare disease cohorts are inherently small and low-powered. Thus, for differential expression analyses it is common to take a “systems” approach to identify functional signals in aggregate through gene set or pathway enrichment. However, in some instances, the identified functions are not specific to the disease (e.g., protein binding) or are confounded with a biological signal stronger than the disease (e.g., population variation). What, then, is the true disease signal, and how does one look for it? Our focus is on a rare disease cohort, the *TAF1* syndrome, where we evaluate systems approaches by testing replicability across six pedigrees. The probands each have a unique variant in their *TAF1* transcription factor and share phenotypic attributes including intellectual disability, facial dysmorphism, general hypotonia, an unusual gluteal crease and cardiovascular structural defects. To assess systems-based analysis within small cohorts, we sequenced the blood transcriptomes of all family members of the six pedigrees, then analyzed each pedigree individually. We tested the validity of the family-specific results by looking for recurrence across the cohort. To evaluate the differentially expressed genes we used both standard gene set enrichment methods and a novel meta-analytic co-expression assessment that exploits thousands of publicly available expression datasets to detect outlier events. Using the standard systems approach we find that differentially expressed gene sets are highly disjoint between pedigrees and that enrichment signals are mostly unique, with recurrent enrichment occurring only in large and non-specific gene sets. In contrast, with our outlier analysis we find recurrence of two genes, *CACNA1I* and *IGFBP3*, that are both downregulated in 5 out of the 6 pedigrees. We call these “functional outliers” as they appear to be uniquely dysregulated in the cohort, specifically acting inconsistently with respect to the primary systems-signal within the data, and therefore missed by standard systems analysis. Our outlier analysis revealed otherwise difficult-to-observe signals in the case of the *TAF1* syndrome cohort. We anticipate that future transcriptomic studies of rare disorders would benefit from the application of this approach.

1428W**Timely detection and bioinformatic mitigation of rapid whole genome**

sequencing genome-wide biases in a clinical lab setting. S. Batalov, S. Chowdhury, Y. Ding, S. Kingsmore, N. Veeraraghavan, RCIGM Investigators. Rady Children’s Institute for Genomic Medicine (RCIGM), San Diego CA.

The speed, accuracy and lowering costs of whole-genome sequencing (WGS) can change the acute care of infants in the intensive care units (ICU). In a clinical setting, it is imperative to establish a robust and scalable system for quality control (QC) of every patient’s genome through all steps of the high volume rapid sequencing lab and *in silico* workflows, and emphasis on rapid QC metrics that can quickly identify potential processing issues in real-time is essential. We present a fast and sensitive workflow called **CoNVince**, to identify problematic genomic libraries that may contain GC-biases that could impact downstream small and large variant calling. **Methods:** WGS was performed via Illumina SBS chemistry and alignment and variant calling performed via EdicoGenomics DRAGEN. Copy number variation (CNV) calling was performed via a combination of the CNVnator and Manta callers with population frequency filtering. A BAM file is an input to the **CoNVince** software that identifies genome coverage biases in a few hundred representative genomic regions for each nucleotide compositional class. Typical robust PCR-free libraries have a small negative, positive or essentially no GC composition bias (within +/-25%). However, libraries with larger bias can produce an abundance of false positive CNV calls. This contrast was exploited as a signal for library quality. **Results:** Retrospective analysis revealed that 1.5-2% of the 1400+ sequenced samples were identified as having high GC-content biases. These samples all resulted in a higher number of CNV calls that correlated with the degree of GC bias. Only DNA library re-preparation and sequencing resolved the issue, while re-sequencing the existing libraries did not. We have developed a CNV workflow (**CoNVince**) which resamples the aligned reads to counter the identified compositional biases systemically in conjunction with using such tools as CNVnator/CNVkit and others. This framework serves to inform sequencing quality as well as reduce false positives in CNV calls from affected samples. **Conclusions:** The ability to implement rapid and reliable QC metrics in the rapid WGS process is essential as hours can make the difference in the impact a genetic diagnosis may have in the management of the patient. With the significance that CNVs play in the genetic disease burden in the intensive care unit, this quality control step can help rapidly identify samples that require additional laboratory processing.

1429T**Re-computing NGS read depth statistics to standardize coverage metrics when using disparate variant call outputs for clinical genomics applications.**

A. Bhattacharya¹, A. Gerasimova¹, Q. Nguyen¹, C. Elzinga², E. Moler¹. 1) Quest Diagnostics Inc, 33608 Ortega Hwy, San Juan Capistrano, CA 92675; 2) Quest Diagnostics Inc, 200 Forest St, Marlborough, MA 01752.

Next-generation sequencing (NGS) produces large quantities of sequencing data of variable quality. One of the requirements of clinically validated NGS pipelines is accuracy of computed coverage depth and variant frequency with specified minimum base quality. Because different tools have different strengths, multiple variant callers can be used to improve accuracy of variant calling. However, the associated read counts for reference and alternative alleles will often differ depending on the approach used. These differences can be dramatic in particular sequence contexts, such as repetitive regions. Here we present a user-friendly Python tool that can use BAM and VCF input files and a user-provided PHRED scaled quality score cutoff to calculate the reference and alternative allele counts for SNPs, INDELs, and structural variants. The tool provides separate per-strand counts to facilitate measuring strand bias, and works in the context of multiallelic sites if present. The tool can also be used at a regional level, performing coverage analysis using BAM sequence and BED region definitions files. In this case, it provides the minimum, maximum, and mean depths for each region, and also provides details of each nucleotide count, and insertion and deletion counts, for every base in each BED region. Recomputing NGS read depth using this tool allowed separation of clinically relevant variants from common variants with a high level of accuracy, including successfully separating myeloid neoplasm-associated mutations from common variants in 47 genes.

1430F

The newborn screening translational research network: Advancing the understanding of genetic disease in newborns and children using a shared infrastructure of genomic and phenotypic information. *A.M. Brower, J.A. Adelberg, D.E. Ekoma, M.S. Watson.* American College of Medical Genetics and Genomics, Bethesda, MD.

Since its inception over half a century ago, state-based newborn screening (NBS) programs annually screen 4 million newborns for genetic conditions that require immediate treatment to prevent morbidity and mortality. Research that capitalizes on the NBS system offers an opportunity to improve understanding of genetic disorders in newborns and children, develop novel technologies, and study the effectiveness of new treatments. Researchers have begun to explore the use of genomics in NBS, and state-based NBS programs have begun screening for disorders that have a large amount of genetic heterogeneity, including onset of symptoms in childhood or beyond. Natural history studies that capture longitudinal information on cases resulting from this expanded NBS will establish the clinical validity and clinical utility of identified disorders, and pilots of novel NBS technologies will provide analytical validation. Because the majority of the conditions that are part of, or candidates for, newborn screening in the United States are rare or very rare, a shared informatics system is critical. The ideal system will support the collection, analysis and sharing of genomic and phenotypic data from the basic research designed to advance understanding of the disease process, to translational research to develop technologies to screen and therapies to treat, to public health implementation of a comprehensive newborn screening program to identify newborns at risk. To support these efforts, the National Institute of Health *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD) Hunter Kelly Newborn Screening Research Program created the Newborn Screening Translational Research Network (NBSTRN). The American College of Medical Genetics and Genomics (ACMG) operates the NBSTRN Coordinating Center (NBSTRN-CC) that is working to develop and implement a suite of tools to efforts to translate basic research discoveries to the state-based newborn screening programs. We report on the use of standardized vocabularies, interactive computer systems, and robust security measures used by over 100 researchers across 8 studies are utilizing genomic and phenotypic data. Data from 1900 participants with over 1 million data points have been collected. The use of the LPDR by researchers will create unique cohorts of individuals with genetic disease and accelerate understanding of these conditions. .

1431W

A bioinformatics framework for identification of mishandled samples in clinical NGS based tests. K. Cao¹, P. Jayaraman¹, M. Gonzalez¹, F. Lin¹, Z. Yu¹, S. Baker², C. W¹, M. Li^{1,2}, M. Sarmady^{1,2}. 1) Children's Hospital of Philadelphia, Philadelphia, PA; 2) University of Pennsylvania, Philadelphia, PA.

Next-generation sequencing (NGS) workflow in the clinical laboratory is very complex and involves multiple manual steps. It is therefore, susceptible to contamination and human errors including sample mislabeling, swapping or DNA/RNA mix-ups. Misdiagnosis as a result of such incidents can lead to serious medical consequences. We have developed a bioinformatics framework to detect the above incidents. Each one method in the framework can be applied independently to either target panel or whole-exome based tests. Given the high sensitivity of NGS assays, DNA of the sample mixed with others could result in false mosaicism and/or somatic changes and ultimately false positive calls and potentially missing true positive calls. We applied the tool *VerifyBamID* on whole-exome bam files with predetermined DNA mixture levels ranging from 0.5-5%. We assessed the correlation of contamination level between computational prediction and *in-silico* mixture. We then identified the acceptable contamination level that has least impact on variant genotyping as the QC threshold in our NGS bioinformatics pipeline. Prediction on gender not only helps define hemizygosity of variants found in sex chromosome but also can quickly identify whether the sample is mislabeled or swapped when it shows opposite gender to the patient's chart. We developed an algorithm to predict the gender using selected variants presented in non-pseudoautosomal regions of *ChrX* using whole-exome data. We determined the Het/(Hom+Het) ratio in females and males using a cohort of samples with known gender. We then predicted the sample's gender by comparing the het ratio in the sample to predetermined thresholds for male and female. Sample identity confirmation is urged for all NGS tests in clinical laboratories. We determined the sample identity by cross-matching the calculated genotypes of sixteen preselected fingerprint SNPs to the results generated by TaqMan assay using second DNA extraction on the same sample. Specifically, for cancer genetic tests using both targeted panel and fusion panel, we cross-matched genotypes of a set of comparable fingerprint snps in DNA and RNA NGS data to confirm the sample identity. In summary, we proposed a framework that enables early detection of mishandled samples in NGS-based diagnostic tests in clinical laboratories before they enter downstream interpretation, and therefore minimizes the possibility of misdiagnoses due to sample miss-processing.

1432T

HLA-Helper: HLA sequence conversion and association analysis toolkit.

W. Choi¹, Y. Luo², S. Raychaudhuri², B. Han¹. 1) Department of Medicine, Seoul National University College of Medicine, Korea, Seoul, South Korea; 2) Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA.

Human Leukocyte Antigen (HLA) genes encode the Major Histocompatibility Complex (MHC) protein, which controls human immune responses and affects susceptibility of various diseases. In an association analysis to fine-map which allele or amino acid position is driving the disease, researchers use HLA typing technologies to genotype HLA or computational approaches to impute HLA from intergenic SNPs. Each high-resolution HLA allele represents a specific DNA or amino acid sequence. Thus, given typed or imputed alleles, the alleles must be converted to sequence information to be used in association analysis of the exonic SNPs and amino acids. Here we introduce a new software package HLA-Helper that helps to process HLA sequence information from typed or imputed HLA information. The package accepts HLA input information in a wide spectrum of formats ranging from imputation technologies (SNP2HLA, HIBAG, HLA*IMP) and typing technologies (Serological low-resolution typing, 2-field high-resolution typing, and 3- or 4- field sequence-based typing), and undertakes the converting job to generate output in binary markers which can be used in various statistical tests. Researchers can choose either the latest version or a specific previous version of IMGT/HLA database for the sequence information. The package also includes the basic association analysis tools for the univariate, multivariate, and conditional association tests and the plot generation tools for the Manhattan plot, amino acid plot, and association heatmap plot.

1433F

Hi-Lite, generalized linear model based similarity measurement for Hi-C samples. C. Crowley¹, Y. Yang², M. Hu³, Y. Li^{1,2}. 1) Biostatistics, University of North Carolina Chapel Hill, Chapel Hill, NC; 2) Genetics, University of North Carolina Chapel Hill, Chapel Hill, NC; 3) Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic Foundation, Cleveland, OH.

Genome-wide three-dimensional (3D) organization of DNA chromatin can directly influence their genomic function. Chromatin conformation capture (3C) and derived technologies, such as Hi-C, provide a genome-wide view of chromosome spatial organization (i.e., interactome). Hi-C data, in the raw form, is usually presented in the form of a contact frequency matrix where the read counts (constituting the matrix elements) contain various sources of uncertainties, biases, and noises. It is therefore challenging to quantify the similarity of the interactome between two different Hi-C experiments, which can be two replicates of the same sample, samples from two different time points, or two different tissues or cell lines. To address this challenge, we present a generalized linear model (GLM) based approach, named as Hi-Lite, to measure the interactomic similarity between different Hi-C samples. We have re-analyzed our most recent published compendium of interactome datasets. Our results suggest that Hi-Lite provides more biologically relevant clustering results compared to other similarity (or dissimilarity) metrics, including Euclidian (i.e., L2) distance, L1 distance, Pearson correlation, and Spearman correlation.

1434W

Accessing and extracting informations from binary sequence alignment/map (bam file) in a non-unix environment with python 3.6+. J.G. Damasceno¹, F.F. Gouveia Filho², G.M. Novo Filho¹, A.T. Dias¹, A.M. Nascimento¹, E.A. Zanardo¹, F.A.R. Madia¹, O.S. Akl¹, M.M. Montenegro¹, Y.G. Oliveira¹, L.D. Kulikowski¹. 1) Laboratório de Citogenômica LIM 03, Departamento de Patologia, Faculdade de Medicina da Universidade de São Paulo, São Paulo, SP, Brazil; 2) Núcleo de Tecnologias Estratégicas em Saúde, Universidade Estadual da Paraíba, Campina Grande, PB, Brazil.

The Sequence Alignment/Map (SAM) file is the most suitable way to store sequence alignments information provenient from massive parallel sequencing methodologies. In its binary form (a BAM file), it is a compressed gzip file used to store the same information in less computational space, while still providing quick access to the data. SAMTools is used to access, read and extract information from both file formats in a quick and reliable way, while other tools like Pysam (for Python 3+) and Rsamtools (for R) utilize other languages to achieve the same results. Both SAMTools and Pysam are UNIX only, while Rsamtools is available in many different platforms. John Longinoto developed the non-UNIX Pybam algorithm (Python 2) that decodifies the binary file back to a text (read-friendly), although Pybam became outdated by current standards and Python upgrades. Inspired by this approach, we developed an adapted algorithm that works with Python 3.6+ in a Windows distribution, allowing the user to transform and handle the sequence data stored inside the file. The algorithm takes less than a second to write the file into memory, has O(N) complexity and an average of 1*10⁶ loops/minute while reading the data with no memory overuse. Working in an UNIX environment is common to bioinformaticians, but is not to the average researcher that will be held back by a lack of working tools. The development of bioinformatics tools that work on Windows environments are crucial in that aspect, as less than 12% (April, 2018) of the world utilizes UNIX machines.

1435T

Storage and use of dbGaP data in the cloud. *M.L. Feolo¹, A. Stine¹, C. O'Sullivan¹, S. Kravitz², J.B. Holmes¹, B.L. Kattman¹, S.T. Sherry¹.* 1) National Center for Biotechnology Information, Bethesda, MD; 2) MITRE Corp., McLean, VA, USA.

The National Center for Biotechnology Information (NCBI) database of Genotypes and Phenotypes (dbGaP) is an NIH-sponsored archive charged to store and provide access to information produced by genome-scale studies that investigate the interaction of phenotype and molecular data. The next-generation sequence data deposited to dbGaP are processed and distributed by NCBI's Sequence Read Archive (SRA) or by various NIH "trusted partners". The dbGaP has progressed from a central archive to one with federated storage solutions. This progression was done initially by incorporating NIH contracted trusted partners that provision data on servers external to NCBI, and now to cloud-hosted data driven by the increasing scale of sequencing projects. Some studies in dbGaP now contain raw sequence data (e.g. BAM or CRAM) of the size that are time-consuming to download, and may best be analyzed using scalable, cloud computing. Fortunately, both NIH policy and commercial cloud capabilities have evolved to provide solutions to this problem. The Sequence Data Delivery Pilot (SDDP) is an NIH funded project to develop cloud-hosted data solutions in support of the NIH Genomic Data Sharing policy. In coordination with the SDDP, NCBI's SRA has extended accessions to allow reference to cloud locations for BAM, CRAM, and VCF files. The SRA accession space, and search methods on the NCBI sites are otherwise unchanged. In addition, a partnership with MITRE has resulted in a SDDP released software package (Fusera) implementing FUSE in the user's cloud computing account that allows read access for dbGaP/SRA data from cloud accounts without the need to move files from the original location. Approved researchers that require a local copy of the files can use "srapc" utility to copy the files to the their own cloud storage/compute account. Changes to dbGaP/SRA in collaboration with SDDP provides a new paradigm in data sharing by providing direct access to large cloud based datasets via dbGaP approvals, eliminating the need for download. The dbGaP still supports studies that are not using cloud storage, and a single dbGaP study accession may contain phenotypes and genotypes (i.e., .VCF or PLINK files) stored at NCBI, and raw sequence data available in the cloud. This presentation will describe how dbGaP/SRA has been successfully extended to provision storage and direct access of data on Google (GCP) and Amazon (AWS) in support of the SDDP and the NIH GDS policy.

1436F

High accuracy somatic variant detection with TNscope. *D.N. Freed, R. Aldana, R. Pan, H. Chen, H. Feng, L. Chen.* Sentieon Inc., Mountain View, CA.

Accurate identification of somatic variation is critically important in Oncology where somatic variation guides treatment decisions. However, accurate identification of somatic variation remains difficult due to its low allele frequency and sequence noise. We developed TNscope, a haplotype-based somatic variant caller, to provide highly accurate identification of somatic variation. Sentieon TNscope provides significant improvements over existing tools. New algorithms for local *de novo* assembly enable full evaluation of all possible haplotype candidates, resulting in increased sensitivity for low allele frequency variant candidates. All variant candidates are evaluated by rigorous PairHMM-based analysis without any down-sampling. These variant candidates are further annotated with novel features for improved variant filtering. Structural variants are identified through a combination of split-read and paired-read methods. Filtering of SNVs and Indels is performed with a pre-trained machine-learning model for increased accuracy. The engineering version of this method was used in the most recent ICGC-TCGA DREAM Mutation Calling challenge, and ranked first on the leaderboard for all three categories: SNV, Indel, and SV. To evaluate the performance of TNscope, we compared Sentieon TNhaplotyper -- a tool that matches the results of MuTect2, with Sentieon TNscope with a post-processing model trained on an *in-silico* mixture of HG002 and HG001. We then evaluated the performance of both tools on a tumor-normal pair of GIAB samples with an *in-silico* mixture of HG005 and HG004 as tumor and HG004 as normal, with tumor fractions of 0.1, 0.15, and 0.20, and both tumor and normal samples at 100x depth. TNscope significantly outperforms TNhaplotyper. On the three tumor fractions, TNhaplotyper achieves F1-scores of (59.8, 73.0, 77.0) for SNPs and (39.7, 48.6, 52.3) for Indels, while TNscope achieved F1-scores of (91.7, 97.4, 98.7) for SNPs and (86.9, 93.2, 95.2) for Indels.

1437W

Phenotype-based variant triaging for use in clinical whole genome sequencing (cWGS). V.G. Gainullin¹, B.R. Lajoie¹, A.L. Halpern¹, A. Scocchia¹, A. Chawla¹, M. Rajan¹, S.S. Ajay¹, A. Warren¹, J. Avecilla¹, F. Mullen¹, V. Rajan¹, R.T. Hagelstrom¹, M.L. Cremona¹, D.L. Perry¹, J.J. Belmont¹, D.R. Bentley², R.J. Taft¹. 1) Illumina Inc., 5200 Illumina Way, San Diego, CA, USA; 2) Illumina Cambridge Ltd., Chesterford Research Park, Little Chesterford, UK.

Clinical whole genome sequencing (cWGS) detects a greater range of genetic variation than either targeted or whole exome sequencing, and as the time required for sequencing and secondary analysis continues to shorten, interpretation is becoming the limiting factor for scalable deployment. We have found that a typical rare, undiagnosed and genetic disease (RUGD) cWGS analysis test detects approximately 4,000,000 variants per sample, with >10,000 variants passing preliminary filtering based on allele frequency and quality filters. Subsequent filtering steps result in ~700 variants that meet stringent genotype, inheritance, consequence and pathogenicity prediction filters, of which an average of 75.6 variants map to a gene linked to at least one of the patient's phenotypic features. Although multiple gene disease relationship (GDR) resources exist, no single resource captures recently described GDRs or the complete phenotypic spectrum associated with a locus. To address this problem we developed an approach to compute similarity indices reflecting the phenotypic overlap between the patient and established GDRs derived from OMIM GeneMap, HPO and PubMed Central abstracts. Patient phenotypes are populated using a catalog of synonyms, and novel phenotypic terms are discovered through word embedding models. Additionally, HGNC loci symbols are mapped between any current or prior names, or other aliases. Gene-phenotype pairs are weighted by the strength of the support in the primary literature. Genes without GDRs inherit the phenotypes linked to their interaction partners and/or orthologous mouse genes to aid in the identification of novel research candidates. Preliminary evaluation in 14 resolved cases (5 proband-only and 9 trios) showed that this approach results in greater sensitivity than using any individual GDR resource. Across 14 cases, the causal variant had an average rank of 2.7 of 75.6 strong candidate variants per case (top 3.5%). Taken together, our results demonstrate that computational phenotype-based interpretation is likely to be an effective method to significantly reduce the curation burden of cWGS RUGD cases.

1438T

Comparison of phenotype-driven variant prioritization tools for diagnostic exome sequencing. Y. Guo, M. Harr, D. Watson, X. Wang, B. Almoguera, K. Stumm, C. Chung, L. Tian, F. Mafra, R. Pellegrino, H. Hakonarson, A. Santani. The Center for Applied Genomics at the Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA.

Exome sequencing has been widely used in clinical testing for identifying genetic variants that cause Mendelian disorders. The amount of genomic data collected in some clinical labs has increased to a tipping point where semi-automatic or automatic bioinformatics pipelines are required to assist manual variant prioritization by human analysts. Multiple tools are designed to implement such tasks and our current work aims at evaluating and comparing those tools using real cases. We selected 74 Mendelian disease cases of different inheritance models as a test dataset, each with positive genetic testing result, detailed clinical phenotype in the format of Human Phenotype Ontology (HPO) terms, and whole exome sequencing data generated from Illumina HiSeq machines. Four different variant prioritization/clinical correlation tools (Phenolyzer, Phenomizer, Exomiser, and Moon) were assessed and the results were compared to the genetic diagnosis. The tools differentiate in 4 aspects, a) final ranking of the diagnosis in the resulting variant/gene list, b) availability to be incorporated into a bioinformatics pipeline, c) whether it is publicly accessible or a commercial product, and d) customizability and scalability. Moon achieve the best ranking of the known diagnosis. All four tools allow HPO terms as input, and Phenolyzer, Phenomizer and Moon can take both HPO terms and free text of phenotypes as input. Exomiser and Phenolyzer provide gene/variant lists as outputs and can be run in batch mode, enabling easy integration into script-based pipelines. Phenomizer requires manual input while in Moon users can either input manually or import data with API. Exomiser and Moon have the ability to assess disease associations not just for genes, but also at the variant level, taking into account pathogenicity estimates of the variants. Moon is a commercial software while the others are open to the public. And Phenolyzer showed flexibility in customization and scalability with the option to include additional databases. We assessed 4 variant prioritization tools for the identification of causal variants of Mendelian disorder cases with whole exome sequencing data. Comparison with the known diagnosis showed strengths and areas for improvement for each tool. A comprehensive, open-to-the-public method is needed to achieve optimal ranking of diagnosis in the test dataset, capability to be integrated into bioinformatics workflow, and flexibility in customization and scalability.

1439F**Are we close to constructing a fully diploid view of the human genome?**

B.T. Hannigan, C. Fungtammasan, N. Hill. DNAnexus, Mountain View, CA.

The publication of the draft human genome sequence in 2001 ushered in the era of large-scale comparative human genomics. Over the following decade and a half, researchers have continued to refine the initial draft to create the series of human reference genomes used today to diagnose rare genetic disorders, examine a patient's cancer risk profile, or even to monitor the health of a developing fetus. While this resource has proved to me invaluable, there are still a number of improvements that researchers have proposed. One such improvement is to move from a haploid view of the genome to a diploid view. The latest version of the human reference genome, GRCh38, moves in that direction with its inclusion of alternative alleles at certain genomic loci, but some researchers, such as those involved in the Vertebrate Genome Project, have proposed a more radical approach to the creation of reference genomes: a fully phased diploid genome. Here, we build on previous work by other groups to build a phased genome assembly for the Ashkenazi son of the Personal Genome Project by applying recent algorithm and technology improvements such as trio-binning and 10X linked-reads. We examine the quality of the assembled genome as well as the structural variants identified, with a particular focus on challenging but biologically important genomic regions such as those containing the HLA genes.

1440W**Tissue transcriptomics: Potential for postmortem interval predictions.**

N. Herrick¹, R. Srivastava², G. Sandusky³, S.C. Janga², S. Walsh¹. 1) Department of Biology, School of Science, Indiana University - Purdue University Indianapolis, Indianapolis, IN; 2) Department of BioHealth Informatics, School of Informatics and Computing, Indiana University - Purdue University Indianapolis, Indianapolis, IN; 3) Pathology and Laboratory of Medicine, Indiana University School of Medicine, Indianapolis, IN.

Accurate prediction of the postmortem interval (PMI) is one of the most important and complex tasks forensic coroners perform. Investigators have attempted to determine PMI, the elapsed time between death and discovery of a body, with numerous biological, chemical, biochemical, and/or physical indicators for civil and criminal investigations (Hunter, M.C. et al., 2017). Maintaining and developing ample methods to determine PMI is advantageous as death is entirely circumstantial and different approaches are necessary due to evidence type availability. Transcriptomics offers a better alternative for the condition of death as differential expression of transcripts within various tissues can ultimately provide more precise PMI indicators, as shown by Ferreira, P. G. et al., 2018, and can ultimately be used to produce trained prediction models regarding the tissue type, accuracy, and/or context. The Genotype-Tissue Expression (GTEx) Project provides RNA-seq data (dbGaP accession number phs000424.v7.p2) across 9,777 sample tissues from 550 subjects utilized for the development of PMI prediction models. Here we have mined this data to further refine potential key predictors associated with PMI using blood and salivary gland tissues. This short list of candidate markers has been elucidated from the dataset and prediction models generated using advanced machine learning techniques. Validation of these systems (biological assays and prediction models) was applied on an additional cohort to measure prediction performances. This cohort includes newly obtained postmortem cell types that possess time of death metadata. Tissue transcriptomics derived and analyzed from postmortem samples that utilize this novel tool will immensely aid forensic investigators in PMI prediction and contribute to a broader understanding of postmortem biological processes.

1441T

A model based association test considering multiple causal variants. *J.J. Joo¹, F.H Hormozdiari², E.E Eskin^{3,4}*. 1) Department of Computer Science and Engineering, Dongguk University-Seoul, 04620 Seoul, South Korea; 2) T.H. Chan School of Public Health, Harvard University, 665 Huntington Avenue, Boston, MA 02115; 3) Department of Computer Science, University of California, Los Angeles, Los Angeles, CA 90095, USA;; 4) Department of Human Genetics, University of California, Los Angeles, Los Angeles, CA 90095, USA.

The standard genome-wide association studies (GWAS) search for an association between a single variant and a phenotype of interest. Recently, several studies reported that at many risk loci, there are more than single causal variant. For the loci with multiple causal variants with small effect sizes, the standard GWAS approach may be inappropriate as it is underpowered to detect small effects. Alternatively, an approach considering effects of multiple variants may increase statistical power by incorporating effects of multiple causal variants. In this work, we propose a new statistical method, MARS(-Model-based Association test Reflecting causal Status), that tries to find an association between variants in risk loci and a phenotype considering the causal status of the variants. Applied our method to several simulated datasets, our method successfully controlled type I error as well as improved statistical power compared to standard univariate GWAS. Applied to GTEx data, we identified more eGenes than the one reported by GTEx.

1442F

Med2RDF: A semantic knowledge base for genomic medicine. *T. Katayama¹, M. Kamada², S. Kawashima¹, R. Kojima², M. Nakatsu², Y. Okuno²*. 1) Database Center for Life Science, Joint Support-Center for Data Science Research, Chiba, Japan; 2) Kyoto University, Kyoto, Japan.

We have compiled a collection of Japanese genomic variations and annotated the variants as two newly developed databases toward a realization of the genomic medicine. The TogoVar database, developed by National Bioscience Database Center (NBDC) and Database Center for Life Science (DBCLS), provides a comprehensive archive of the Japanese variants. While, the Medical genomics Japan Variant Database (MGeND), developed by Kyoto University, accumulates variants which have been curated by the experts of each disease domain. For conducting curation of variants of uncertain significance (VUSs), it is required to integrate interpretation of expertise related to mutation, treatment results, and clinical information of the patients. Because this procedure takes a lot of labor, we plan to make it computer aided especially with the support of artificial intelligence (AI) and machine learning (ML) technologies. In order to realize this, we first need to integrate existing knowledge which biomedical experts apply for curation, based on medical guidelines. In which, relations among mutations, genes, functions, diseases and drug compounds, and frequency of mutations, approval status of trials are used as the judging criteria. These are stored in the distributed biomedical databases, however, it is hard to integrate and use them for effective knowledge extraction because each database has its own data format and structure. To resolve this issue, we introduced the Semantic Web in which heterogeneous data can be unified by the Resource Description Framework (RDF). RDF data uses globally unique Uniform Resource Identifiers (URIs) to identify concepts and objects in the datasets, ontologies to describe semantic types, and links to relate information across multiple databases seamlessly. We transformed major biomedical databases into RDF by developing ontologies to describe variations. Resulting RDF datasets, which includes data on clinical relevance, mutation frequency, protein interactions, and gene-drug relations, are distributed at the Med2RDF project website (<http://med2rdf.org>) developed by DBCLS and Kyoto University in collaboration. In the poster, we will report the results of (1) integration of these datasets in the TogoVar database, and (2) applications on predicting pathogenicity of variants and relations of variation-drug based on our in-house developed AI/ML methods. Those are expected to help experts and curated results will be incorporated into the MGeND database.

1443W

Empirical evaluation of variant calling accuracy using ultra-deep whole-genome sequencing data. *T. Kishikawa^{1,2}, Y. Kamatani^{3,4}, M. Kubo⁵, Y. Okada^{1,4,6}*. 1) Department of Statistical Genetics, Osaka University Graduate School of Medicine, Suita, Japan; 2) Department of Otorhinolaryngology - Head and Neck Surgery, Osaka University Graduate School of Medicine, Osaka, Japan; 3) Kyoto-McGill International Collaborative School in Genomic Medicine, Kyoto University Graduate School of Medicine, Sakyo-ku, Kyoto, Japan; 4) Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 5) RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 6) Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Suita, Japan.

In the study design of whole-genome sequence (WGS), a sequencing depth is a crucial parameter to define variant calling accuracy and study cost, while standard recommendations has not been established. We empirically evaluated variant calling accuracy of the WGS pipeline using ultra-deep WGS data (approximately 410×). By randomly sampling sequence reads, we created a series of simulation WGS datasets with a gradual variety of depths ($n = 54$; from 0.05× to 410×). We evaluated genotype concordances of the WGS data with those in the SNP microarray data or the WGS data using all the sequence reads. The WGS data with higher depths showed higher concordance rates, and > 13.7× depth achieved as high as > 99% of concordance. Comparisons with the WGS data using all the sequence reads showed that SNVs achieved > 95% of concordance at 17.6× depth, while indels showed only 60% of concordance. These results recommended the medium degree of the WGS depth setting (around 15×) for an achievement of both accurate SNV calling and cost-effectiveness, while relatively higher depths are required for accurate indel calling..

1444T

Mendelian inheritance errors in whole genome sequenced trios are enriched in repeats and cluster within copy number losses. *P. Kothiyal, W.S. Wong, D.L. Bodian, J.E. Niederhuber*. Inova Translational Medicine Institute, Inova Health System, Fairfax, VA.

Background: Next-generation sequencing (NGS) enables advances in the clinical application of genomics by providing high-throughput detection of variants in the genome. However, NGS technologies, especially whole genome sequencing (WGS), are often associated with a high false positive rate. Trio-based WGS data can contribute significantly towards the development of improved quality control methods that can also be applied to non-family WGS. Mendelian Inheritance Errors (MIEs) in parent-offspring trios are commonly attributed to erroneous sequencing calls, as the rate of true de novo mutations is extremely low compared to MIE incidence. Here, we analyzed WGS data from 1,314 mother, father, and offspring trios across ethnically diverse human populations with the goal of studying the characteristics of MIEs. **Results:** We filtered MIEs on read depth, genotype quality, and allelic balance and observed that quality based filtering has a greater impact on frequent MIEs compared to MIEs unique to a trio. Our results indicate that MIEs are enriched in regions with repeat elements and that MIE density correlates with the density of Short Interspersed Nuclear Elements (SINEs). We also observed clustered MIEs in regions overlapping large deletions. We created population-specific MIE profiles and discovered genomic regions that represent different MIE distributions across populations of differing ethnicity. Finally, we have summarized these MIE distributions to produce population-specific MIE tracks that can be loaded in UCSC Genome Browser. **Conclusions:** While MIEs are frequently used for benchmarking error rate, and have been explored for detection of deletions, we aimed to understand the properties of MIEs before discarding them as errors or including them in alternative applications. With the implication of de novo mutations in various disorders, they are being studied more routinely in clinical practice. We aim to provide a reference that can be used for annotating MIEs with frequencies in a larger population to aid in the filtering of candidate de novo mutations. Finally, the MIE tracks can be used for flagging calls in proximity to clustered MIEs before allele frequency and admixture calculations, annotating candidate de novo mutations, discovering population-specific putative deletions, and for distinguishing between regions that have errors due to sequence quality vs. chromosomal anomalies.

1445F

INTERSPIA: A web application for exploring the dynamics of protein-protein interactions among multiple species. *D. Kwon, D. Lee, J. Kim, J. Lee, M. Sim, J. Kim.* Konkuk University, Seoul, Korea, Republic of.

Proteins perform biological functions through cascading interactions with each other by forming protein complexes. As a result, interactions among proteins, called protein-protein interactions (PPIs) are not completely free from selection constraint during evolution. Therefore, the identification and analysis of PPI changes during evolution can give us new insight into the evolution of functions. Although many algorithms, databases and websites have been developed to help the study of PPIs, most of them are limited to visualize the structure and features of PPIs in a chosen single species with limited functions in the visualization perspective. This leads to difficulties in the identification of different patterns of PPIs in different species and their functional consequences. To resolve these issues, we developed a web application, called INTER-Species Protein Interaction Analysis (INTERSPIA). Given a set of proteins of user's interest, INTERSPIA first discovers additional proteins that are functionally associated with the input proteins and searches for different patterns of PPIs in multiple species through a server-side pipeline, and second visualizes the dynamics of PPIs in multiple species using an easy-to-use web interface. INTERSPIA is freely available at <http://bioinfo.konkuk.ac.kr/INTERSPIA/>.

1446W

SAMPLIT: Rapid structural variant visualization for short, long, linked, and phased reads. *R.M. Layer^{1,2}, J.R. Belyeu^{3,4}, A.R. Quinlan^{3,4,5}.* 1) Department of Computer Science, University of Colorado, Boulder, CO; 2) BioFrontiers Institute, University of Colorado, Boulder, CO; 3) Department of Human Genetics, University of Utah, Salt Lake City, UT; 4) USTAR Center for Genetic Discovery, University of Utah, Salt Lake City, UT; 5) Department of Biomedical Informatics, University of Utah, Salt Lake City, UT.

Structural variant (SV) detection remains a challenging problem. Biological and technical artifacts complicate our ability to differentiate signal from noise causing the accuracy of SV detection to remain worse than that for smaller variants. As a result, visualization has become a standard step in the SV validation process. Inspecting the raw alignment data, genome annotations, and the region surrounding a candidate SV can help remove false positives and confirm breakpoints and genotypes. We created SAMPLIT to support rapid, focused visualization of SVs, that highlights the alignment signals that are hallmarks of SVs including sequence coverage, discordant paired-end reads, split reads, and alignment gaps. For each SV, SAMPLIT creates a subplot for each sample that includes a sample-specific coverage profile and alignments organized along the y-axis by "event size" (insert size for paired-end reads, gap size for split-reads). This organization creates distinct clusters of normal and discordant alignments. Discordant alignments are color-coded by the type of SV they support. SAMPLIT also supports the simultaneous visualization of: - short Illumina paired-end reads, long reads from PACBIO and Oxford Nanopore, and linked reads from 10X - phased BAMs with any number of haplotype assignments - genome annotations such as transcripts, repetitive regions, and mapping quality tracks With SAMPLIT, users can quickly and easily inspect SVs across many samples and technologies to determine if any sample contains discordant alignments that support the candidate SV, which can be especially helpful when evaluating de novo variants in families and somatic SVs tumor/normal samples. Source code and documentation at <https://github.com/ryanlayer/samplit>.

1447T

Hybrid scaffolding of optical maps provides high contiguity and corrects errors in sequence assemblies. *J. Lee, A. Pang, A. Hastie, J. Wang, T. Anantharaman, H. Cao.* Bionano Genomics, San Diego, CA.

High quality genome assembly of novel genomes has exploded in recent years as a result of novel and maturing technologies. To disambiguate homologous regions of these novel genomes, long reads and linked reads are used to assemble contigs and generate scaffolds with up to 10 Mbp N50 values. Assembling these sequences into chromosome arm or chromosome length can only be accomplished using Bionano Genomics long read mapping or one of the Hi-C based methods. Bionano chromosome assemblies can often be achieved by utilization of the new DLS chemistry and/or by combining with NLRs chemistry. Bionano Genomics, with its long molecules N50 in the range of hundreds of kilobases with the DLS chemistry, is unique in that it can generate megabases long contiguous assemblies with the well understood overlap layout consensus algorithms. These assemblies are then used to scaffold sequences into chromosome or chromosome arm length assemblies by the Hybrid Scaffold pipeline. Alternatively, Hi-C based methods leverage crosslinking of DNA that is in close proximity in vivo through chromatin folding, which is then sequenced using short read sequencing. Since the long range interaction in Hi-C is based on biological connections and encoded by short reads, significant inference is required to reconstruct the interaction information. In comparing Hi-C scaffolds with Bionano DLE Maps, many discrepancies were found. Inverted and rearranged segments of various sizes were identified when aligning the Hi-C scaffolds to the Bionano DLE Maps. Aligning the pre Hi-C scaffolded sequences with BNG DLE maps, we see that most of these divergent breakpoints were at ends of the pre Hi-C scaffolded contigs, suggesting that orientation and arrangement discrepancies were introduced during Hi-C scaffolding. Since the Bionano method leveraged native intact molecules, spanning breakpoints of these segments by Bionano molecules suggests mis-orientations in the Hi-C scaffolds. In another case, a third long read sequencing technology agrees with BNG structure at the regions of discrepancies, further supporting the accuracy of the Bionano assembly.

1448F

QC software for analysis of sequence data in family-based studies. *Q. Li, J. Bailey-Wilson.* Computational and Statistical Genomics Branch NHGRI/NIH, Baltimore, MD.

In recent years, with the increased popularity of DNA sequence data, the research community has paid more attention to family based studies of complex traits. Pedigree data pose new challenges in the quality control steps for sequence variants. Not only do we need to screen data based on various genotyping calling metrics, we also need to use the pedigree structure to detect any markers with inconsistent inheritance patterns. In this work, we present our QC pipeline software, written in R, for sequence data analysis. It includes functions to extract various genotyping calling measurements, to detect Mendelian inconsistency, and to detect any excessive allele sharing due to remote inbreeding. Our software uses the SQLite database for easy access and short processing time for VCF files. Variant level data is stored in a different table from variant-and-sample level data. Sample level data (related to QC metrics) are broken into multiple tables, each table corresponding to only one specific data quality metric. We used our software to conduct quality control on whole exome sequence data on odd numbered chromosomes in extended pedigrees from the Genetic Analysis Workshop 18 (GAW18). We also simulated Mendelian errors and genotyping errors in the sequence data to evaluate our QC functions. Our simulations showed the functions can detect those errors with high accuracy and within a reasonable time frame. We compared our data extraction functions with what are available in other similar software, such as Golden Helix SVS. Our functions perform equally well in term of accuracy and are much more flexible to use. These software are freely available and will assist researchers in large family studies using whole exome and whole genome sequence data.

1449W

MedGen: A powerful tool for clinical genetics. A.J. Malheiro, D.R. Maglott, M. Benson, C. Chen, S. Chitipiralla, D. Hoffman, M.J. Landrum, B. Holmes, V. Gorelenkov, W. Shi, B.L. Kattman. National Center for Biotechnology Information (NCBI), NLM, NIH, Bethesda, MD.

MedGen is a free and easily accessible database of information about diseases and phenotypes with a genetic component, housed at NIH's National Center for Biotechnology Information (NCBI). MedGen facilitates the clinical process by providing unrestricted access to related data from multiple sources such as ClinVar and the NIH Genetic Testing Registry (GTR). MedGen has records for Mendelian disorders, pharmacological response phenotypes, complex disorders, and clinical features or findings. The data in MedGen are updated daily; updates are processed as data becomes available from the source. The data are available for bulk download on the ftp site (<ftp://ftp.ncbi.nlm.nih.gov/pub/medgen/>) and can be easily integrated in electronic medical records. Individual records may be viewed on the web (<https://www.ncbi.nlm.nih.gov/medgen/>). A clinician can use MedGen to find information about a genetic phenotype: 1. by name (autocomplete dictionary available); 2. by causative gene (use gene symbol followed by [gene] to get a list of diseases caused by or with some association to that gene); 3. by clinical feature (enter the name of the feature followed by [clinical features] to retrieve a list of diseases with that feature); and 4. by identifier from another database (e.g. OMIM, HPO, OrphaNet). An advanced search feature is available to support complex queries. On a search result page, there are filters to tailor the view of the data to the current needs of the user. Each disease record has a list of all features of a disorder based on data from HPO and OMIM. One can find all disorders with particular feature(s) using the clinical feature field in the Advanced Search (e.g. syndactyly). This presentation will demonstrate how to leverage information in MedGen throughout the clinical process. To research a case: access disease and phenotype information from multiple resources including *GeneReviews*, Genetics Home Reference, Medical Genetics Summaries, OMIM and pre-calculated PubMed queries. To care for a patient: access actionable information such as practice guidelines, recommendations, position statements and evidence-based reviews from authoritative sources as well as a list of relevant orderable clinical and research tests from the NIH GTR and a list of variants for the relevant genes from ClinVar. To find more details from dedicated sites: follow content-specific links to ClinicalTrials.gov, GARD, Orphanet, MedlinePlus, Malacards, PharmGKB, and patient-oriented resources.

1450T

Improving next-generation sequencing determination of antigens in the MNS blood group system through the use of a graph-based reference assembly. S. McGee¹, J.D. Smith¹, K.W. Lannert², J.M. Johnsen^{2,3}, M.M. Wheeler¹, D.A. Nickerson¹. 1) Department of Genome Sciences, University of Washington, Seattle, WA; 2) Bloodworks NW Research Institute, Seattle, WA; 3) Department of Medicine, University of Washington, Seattle, WA.

Graph-based reference assemblies aim to improve short read mapping and the subsequent detection of genetic variants. Specifically, graph-based assemblies have been shown to improve variant detection by accounting for known variation in highly polymorphic and homologous regions of the genome. In this study, we hypothesized a graph-based approach would improve variant detection in the segmentally duplicated glycoporphin genes (*GYP A*, *GYP B* and *GYP E*). These genes are located in tandem on chromosome 4 and share approximately 97% sequence identity. Mapping quality and genotyping difficulties have previously been reported for a cluster of single nucleotide variants (SNVs) in exon 2 of *GYP A*. These variants encode clinically-relevant antigens in the MNS blood group system (the M and N antigens); however, genotype analyses of *GYP A* exon 2 variants are currently confounded by high homology to exon 2 of *GYP E*. To address this, we constructed a graph-based reference sequence utilizing SNVs and indels detected in NGS data from approximately 1100 samples for which M and N antigens were previously serologically typed. Using the software package VG, we re-mapped a subset of the 1100 samples (N=100) to this *GYP* graph-based reference assembly. Our results show significant improvement in depth and mapping quality of the reads over *GYP A* exon 2. In addition, there was striking improvement in correlation to M and N phenotype data derived from serology. In previous work, variant calls were determined using GATK HaplotypeCaller and showed concordance with serology of 40% for M+N- (homozygous alternate calls) and 14% for M+N+ (heterozygous calls). With the graph-based approach, the concordance of antigen determination increased to 93% for M+N- and 89% for M+N+, improvements of 2.3 fold and 6.4 fold, respectively. These results illustrate how a graph-based approach can improve mapping quality and variant detection for genetically complex and clinically-relevant genomic loci. We are currently extending these analyses to additional blood group loci and predict further application of this approach will detect novel variation and improve blood group antigen determination from NGS data.

1451F

Scientific benefits resulting from standardizing electronic health records to facilitate multi-site collaboration. F.D. Mentch¹, H. Qiu¹, L.M. Vazquez¹, D.J. Abrams¹, J.J. Connolly¹, P.M.A. Sleiman^{1,2}, H. Hakonarson^{1,2}. 1) Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

National and international projects to research human health, many including genomic components, are proliferating. Multi-institution collaborations are enhanced if contributing sites can more efficiently and accurately share phenotype data. One way to facilitate data sharing is for each site to migrate electronic health records (EHRs) to common vocabularies and data storage formats. As it takes substantial effort to standardize EHR data, the implementation costs can be perceived as prohibitive. We report on the scientific benefits we found after migrating to standardized vocabularies and a common data model which offset some of the concerns about implementation cost. For example, as a large biobank site (n>500,000) with extensive EHR data (n>110,000), we found a benefit in the global calculation of drug and health condition trajectories, which we would otherwise do piecemeal, and in the consolidation of concepts across the domains of medications, procedures and diagnoses. We found an uneven loss of information in the transformation particularly in diagnosis translations that is greatest where the ICD coding system is weakest. The migration to a common vocabulary for procedures and lab values was found to be particularly advantageous for our site. Perhaps the greatest value in the migration process was tapping into the codes underlying the clinical EHR data capture system of the hospital, giving a closer connection to the clinical descriptions chosen at the time of patient care and mapping these more accurately to a global standard vocabulary.

1452W

Astrolabe: Automated pharmacogenetic allele calling of CYP2D6, CYP2C9 and CYP2C19 from NGS data. N. Miller^{1,2}, G. Twist¹, N. Gonzaludo³, R. Lata⁴, E. Boone⁴, R. Gaedigk^{2,4}, E. Farrow^{1,2}, A. Gaedigk^{2,4}. 1) Center for Pediatric Genomic Medicine, Children's Mercy, Kansas City, Kansas City, MO; 2) School of Medicine, University of Missouri-Kansas City, Kansas City, MO; 3) Illumina, Inc., Foster City, CA; 4) Division of Clinical Pharmacology, Toxicology & Therapeutic Innovation, Children's Mercy, Kansas City, Kansas City, MO.

CYP2D6, CYP2C9 and CYP2C19 contribute to the metabolism of over 30% of drugs in clinical use including antidepressants, antipsychotics, opioids, antiemetics, anti-arrhythmics, β -blockers, cancer chemotherapeutics, the anti-coagulant warfarin and the antiplatelet agent clopidogrel. The activity of these enzymes varies widely among individuals and there is strong evidence that genetic variation significantly contributes to drug efficacy and risk of adverse events. Expert guidelines have been published to improve the use of substrates metabolized by these three genes. The Pharmacogene Variation Consortium (PharmVar) has defined over 100 allelic variants for CYP2D6, 60 alleles for CYP2C9 and 35 for CYP2C19. CYP2D6 allele calling is complicated by copy number variation, gene rearrangements and conversions and the presence of non-functional homologs and pseudogenes. Given these complexities, routine determination of pharmacogenetic alleles by genome sequencing is non-trivial. Previously, we developed a probabilistic scoring tool for pharmacogene haplotype calling from WGS (Constellation, now renamed Astrolabe) and evaluated its performance for CYP2D6. Here, we describe a series of improvements for CYP2D6 and the addition of support for CYP2C9 and CYP2C19. Enhancements include the incorporation of machine learning classifiers for CYP2D6 structural variant detection, configurable weighting of key haplotype variants, improvements in overall software quality and performance, expansion to GRCh38 and the incorporation of versioned haplotype definitions from PharmVar. Compared to a set of 89 samples genotyped with orthogonal methods, Astrolabe calling of CYP2C9 and CYP2C19 had 100% and 96% accuracy, respectively. Results from two different simulations of all possible diplotypes showed calling accuracy of 100% for CYP2C9, 99.8% for CYP2C19 and 95.7% for CYP2D6. Astrolabe performance was also evaluated against 96 Coriell Cell Repository cell-lines previously genotyped by the Centers for Disease Control and Prevention-based Genetic Testing Reference Material Coordination Program (GeT-RM) (Pratt VM et al., J Mol Diag 2016, 18:109-123). Whole genome sequences were released by Illumina and data for selected consented samples are available on ENA (Study PRJEB19931, <https://www.ebi.ac.uk/ena/data/view/PRJEB19931>). Astrolabe software is freely available for research use from the Children's Mercy Genome Software Portal (<https://www.childrensmc.org/genomesoftwareportal/>).

1453T

Fast model-based detection of identity-by-descent. *J. Nait Saada, P. Palamara.* Department of Statistics, University of Oxford, Oxford, UK.

Identical-By-Descent (IBD) segments are a fundamental measure of genetic relatedness, and IBD detection is of great interest in many applications, including demographic inference and haplotype-based association. Algorithms that scale to large datasets, such as GERMLINE [Gusev et al., *Genome Res.*, 2009] and RaPID [Naseri et al., *Biorxiv*, 2017] sacrifice modeling accuracy in favor of computational speed. On the other hand, efficient model-based algorithms for IBD detection such as RefinedIBD [Browning and Browning, *Genetics*, 2013] enable detecting shorter IBD segments but do not scale as well. We have developed a new algorithm, called Preprocessed Ascertained Sequentially Markovian Coalescent (PASM), that enables accurate IBD detection while retaining computational efficiency. PASM utilizes string matching to detect candidate IBD segments using the PBWT algorithm [Durbin et al. 2014, *Bioinformatics*] and applies the ASMC algorithm [Palamara et al., *Nature Genetics*, in press] to filter out candidate segments using coalescent modeling. We compared the performance of several methods using coalescent simulation, including PASM, GERMLINE, RefinedIBD, RaPID and ASMC. We simulated 10 independent datasets for a 30Mb chromosome and 300 haplotypes, based on a model of European demographic history and recombination rates from a human chromosome 2. We mimicked SNP ascertainment by subsampling variants to match UK Biobank allele frequencies. We used a grid-search over possible parameter values to optimize the accuracy of all methods, measured using area under the precision-recall curve. We defined a region to be IBD if the most recent common ancestor for a pair of individuals is less than 100 generations. PASM's performance (within the recall range [0.01, 0.50] shared by all methods) was 40.51% (SE=3.61) higher than RaPID, 8.85% (1.51) higher than RefinedIBD, 8.94% (1.01) higher than GERMLINE, and 4.95% (0.62) less than ASMC. Comparing accuracy at a deeper time scale (IBD=250 generations), PASM still outperformed other methods (shared recall range = [0.00, 0.26]); PASM's accuracy was 56.37% (4.96) higher than RaPID, 16.07% (0.96) higher than GERMLINE, 2.76% (0.36) higher than RefinedIBD, and only 1.73% (0.18) lower than ASMC. In addition, PASM's computational time scaled well in a larger dataset of 5000 haplotypes, requiring around 21 min, compared to 32s for GERMLINE, 20min for RaPID, 1h and 25min for RefinedIBD and 13h for ASMC.

1454F

The road to atomistic pathology. *M.-S Park, A. Chamberlin, I.D Petrik, D. Shinde, R. Lassiter, F. Hernandez, T. Pesaran, C.-L Gau, H.-M Lu, R.B Fenwick.* Ambry Genetics, Aliso Viejo, CA.

Variant classification of missense alterations remains a significant challenge. The predictive capacities of methods that focus on the sequence (e.g. in-silico predictors) are limited and are typically combined with additional lines of evidence (e.g. population frequency, family studies, and functional studies) to obtain robust classification. However the utilization of alternative evidence categories is limited by availability and cost. Biophysical principles provide a promising complementary source for additional lines of evidence, because protein structures provide a rich set of features that elucidate protein function and dysfunction. We present the analysis of a series of missense variants using biophysical methods, such as model building (Kim, Chivian, & Baker, 2004), and stability calculations (Schymkowitz et al., 2005), combined with structural-biological insight. This integrated biophysical approach offers additional predictive power as well as mechanistic insight into the mode of disruption. We find that analyzing variants in the context of the protein structure yields an atomistic view that explains why variants at similar locations in the sequence have different clinical outcomes. We have successfully applied this approach in the past to explain challenging variants in a variety of genes (Martin et al., 2017; Peng et al., 2018; Powis et al., 2018). We illustrate for several cases how active sites can be affected while leaving stability unchanged and how variants can have significant impacts on protein folding and stability. **Keywords:** Protein structure, Bioinformatics, Genetic testing, Molecular pathophysiology, Mutation detection.

1455W

Interrogation of biases from different platforms and different types of library preparations in transcriptome sequencing study. G. Park, M. Roh, R. Kim, D. Seo. MacroGen Clinical Lab, Rockville, MD.

Nowadays, a gene expression profiling is being conducted with a variety of sequencing platforms and library preparation methods. The sequencing databases like GEO, SRA, and EMBL-EBI have been providing a publicly available transcriptome sequencing data, which enables users to accelerate data sharing and reusability. However, users should still carefully consider using both new sequencing data and public data together due to data discrepancies and biases, which might possibly come from different types of sequencing platform, library preparation methods, sequencing method either paired-end and single end, and length of reads. In this study, we looked into the gene expression profiling from sequencing data produced from the platforms, Illumina HiSeq2500 and NovaSeq, and different library preparation methods such as stranded mRNA, non-stranded mRNA, and globin-zero total RNA kits. Based on our observations, platform biases did not make a huge difference in gene expression profiling; however, different library preparation methods did. The presence of globin significantly affected the gene expression profiling because of the expression of globin related genes. Even though method with strand information kit showed different gene expression profiling, the result had similar overall expression pattern because of the 20% overlap of genes between the top and bottom reference sequences. In conclusion, biases from different platforms of transcriptome sequencing can be insignificant as this can be corrected after removing the batch effects, but different library preparation methods will affect significantly the gene expression profiling. Therefore, the library preparation methods need to be carefully evaluated and selected before conducting meta-analysis of transcriptome sequencing study with publicly available sequencing data.

1456T

Integrated Bayesian analysis outperforms existing methods for variant classification. D. Qian, Y. Tian, S. Li, J. Clifford, B.A.J. Sarver, A. Chamberlin, R.B. Fenwick, T. Pesaran, C. Gau, H.M. Lu, M.H. Black. Ambry Genetics, Aliso Viejo, CA.

Introduction: Common practice for variant classification follows the scheme-based ACMG guidelines (Richards et al. 2015) or utilizes a naïve Bayesian classifier (NBC; Tavtigian et al. 2018) with 7-category evidence data. However, methods to make a definitive variant classification remain a challenge for clinical testing largely due to the lack of effective ways to incorporate patterns of pathogenic information in classified variants in concert with available evidence data. **Methods:** We report a novel integrated Bayesian learning and aggregation analysis using multifactorial variant prediction (MVP) to capture the uncertainty of *in silico* effects and aggregate both qualitative and quantitative evidence in a data-driven probabilistic process. To evaluate classification performance, we assembled a dataset of 1,670 missense variants in 67 genes unambiguously classified by ClinVar consensus as benign (B, n=126), likely benign (LB, n=673), likely pathogenic (LP, n=343) or pathogenic (P, n=528). We also collected 1,056 missense variants of uncertain significance (VUS) supported by 2 or more ClinVar submitters and no conflicting submissions. Each classified variant or VUS in the evaluation datasets had at least 1 type of evidence collected from public resources and a cohort of 176,280 patients who underwent genetic testing at a single diagnostic lab. **Results:** The prediction performance of scheme-based ACMG and NBC methods were roughly equivalent in both classified variants and VUS. While the accuracy of MVP and NBC were comparable for predicting variants with known classifications (99.9% and 99.8%, respectively), the yield of MVP was higher (90.8% vs. 74.2%, respectively). When predicting VUS in the combined set of classified variants and VUS, the accuracy was higher for MVP than NBC (83.4% vs. 62.7%) although both yields were moderate (33.5% vs. 33.1%). Notably, MVP had better performance in predicting pathogenicity; for the 871 LP/P variants, the true positive rates were 90.9% and 53.7% for MVP and NBC, respectively, highlighting a substantial number (n=324, 37.2%) of LP/P variants unclassifiable by NBC that were correctly classified as pathogenic by MVP. **Conclusions:** Taken together, our powerful Bayesian MVP analysis had greater accuracy and considerably higher yield of pathogenic variant classification than other existing approaches, and as such may serve as an automated, data-driven tool to objectively classify missense variants and reduce VUS rates.

1457F

Statistical considerations for the analysis of massively parallel reporter assays data. D. Qiao¹, M.H. Cho^{1,2}, E.K. Silverman^{1,2}, X. Zhou¹, N.H. Laird³, P. Castaldi¹. 1) Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; 2) Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; 3) Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA.

Background: Non-coding DNA is now known to contain a multitude of functional gene regulatory elements that alter gene expression. To identify and functionally characterize these regulatory elements, massively parallel reporter assays (MPRA) utilizing chip-based technologies and next-generation sequencing were developed. Using MPRA, hundreds of thousands of oligonucleotide sequences (oligos) can be examined quantitatively in a single experiment, but the statistical methods to analyze MPRA data, specifically in relation to the regulatory effects of different SNP alleles, have not been fully studied. The methods used to identify allele-specific gene regulatory effects in MPRA include the Mann-Whitney test and t-test on RNA/DNA ratios, Fisher's exact test and QUASAR-MPRA using total oligo counts, and the mpralm method. **Methods:** In this work, we explore fundamental issues with ratio-based representations of gene regulatory effects, and we demonstrate how the baseline allelic imbalance present in most if not all MPRA libraries produces biased results using existing methods. We propose a novel, adaptive testing approach, and we examine the performance of this approach versus all commonly used MPRA methods using simulations. **Results:** We show that, unlike existing methods, the adaptive approach is robust to allelic imbalance and therefore more reliable. We also examine the power of all methods across various realistic scenarios, and we demonstrate that while the mpralm method is biased in certain situations, it also is more powerful across many scenarios. **Conclusions:** We conclude that our recommended approach to MPRA analysis of allelic effects is to use both the adaptive approach and mpralm to maximize power while identifying likely false positive results. We have implemented these tests along with routines for simulating MPRA data in the Analysis Tools for MPRA (atMPRA), which is an R package for the design and analysis of MPRA experiments, available at <http://github.com/redaq/atMPRA>.

1458W

Whole genome sequencing data analysis based on GRCh38 and difficulty on haplotyping for pharmacogenes. N. Qu, M. Jian, J. Guo, Q. Shi, A. Ou, X. Liu. BGI-Shenzhen, Shenzhen, Guangdong, China.

GenBank released GRCh38 (Genome Reference Consortium Human Reference 38) in 2013, whose context ID in UCSC Genome Browser is hg38, which is the most recent genome version and is a more accurate representation of the human genome. The GRCh37 is a single representation of multiple genomes, while the GRCh38 provides alternate contigs ("alt_sequences") for some genomic regions for whose variability prevents adequate representation by one single reference. We analysed NA12878's WGS data using GRCh37 and GRCh38 at the same time, and compared the SNPs and Indels to GIAB golden standard and sanger verification result. We found that the result using GRCh38 presented a high false negative rate. After removing GRCh38's alternate contigs from reference genome, it results in lower false-negative and lower false-positive than GRCh37. All our databases were transferred into GRCh38 version by liftover software. Such as ExAC, dbNSFP, etc. In addition, we developed an annotation software named bcfanno by C language, which can annotate a WGS sample in 2 hours on a stand-alone system. CPIC (The Clinical Pharmacogenetics Implementation Consortium) has stated that if the subject has two haplotypes which are CYP2C9*2 (rs1799853) and CYP2C9*3 (rs1057910), the subject then has a phenotype as poor metabolizer towards warfarin. Or intermediate metabolizer when the two SNPs are on the same haplotype while the other haplotype is CYP2C9*1, which means no related variations. However, due to the limitation of NGS, it is difficult to distinguish whether they are on the same DNA strand. Normally, it is assumed that all the tested variations are on either the same or separated strand. But neither of these assumptions is accurate enough. Else more, sequencing depth of all locus related to conclusions are need to check. For example, NGS result shows that there was only a heterozygous mutation on the CYP2C9 gene: rs1799853. Then we thought the subject has a phenotype of intermediate metabolizer. However, we next found that the sequencing depth of another loci rs1057910 was actually 0X. And the sanger sequencing result of rs1057910 was homozygous variation but not real negative. So the subject should be poor metabolizer. For those uncovered locus, another sequencing method is required. Or at least we should inform the subject that such locus is not successfully called.

1459T**STAASIS: A web based server for identifying Illumina genotype arrays and initial strand orientation.** *N.W. Rayner^{1,2,3}, N.R. Robertson^{1,2}, M.I. McCarthy^{1,2,4}*

1) Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford Oxford, OX3 7BN, UK; 2) Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of Medicine, University of Oxford, Churchill Hospital, Oxford, OX3 7LJ, UK; 3) Department of Human Genetics, Wellcome Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK; 4) Oxford NIHR Biomedical Research Centre, Oxford University Hospitals Trust, Oxford, OX3 7LJ, UK.

Array genotyping remains a popular, cost effective option for large cohorts and along with imputation and meta-analysis has proved extremely successful in identifying novel disease associated loci. A key requirement for both imputation and meta-analysis is that all data are aligned to the same genomic build and strand, an issue we have previously addressed by developing strand files for Illumina and Affymetrix genotyping arrays across all recent genome builds and making them freely available to download. However these strand files assumed that all Illumina data are by default created using the TOP genomic alleles; this is not always the case. Illumina have a number of other possible allele alignments and from the many requests for assistance received to date these alternate alignments appear to be used in a sizeable minority of cases. Another complication is that the array used for genotyping may not be known, or partially known e.g. an OmniExpress array, of which there are 13 versions. This along with the allele alignment issue can make it difficult to select the right strand file to use. To address the requests for help with these issues we had developed programs to automatically identify the array and allele alignment. These though had to be manually configured and run, and with the increase in the number of arrays became increasingly slow to run. We have now redeveloped these programs into a single web server, STAASIS, written in Perl, Java and JavaScript that utilises a client server model with in-memory hashes to reduce the run time from several hours previously to a minute or less. The program accepts a list of variants or a plink format bim file, and whilst both allow for array identification only the latter enables the allele alignment check to be performed. Once submitted the list of variants is checked against the content of all current and previous Illumina arrays, currently numbering more than 160, to identify the most likely array. The alleles from the bim file, if supplied, are compared to all possible allele combinations to determine their orientation. Using 110 bim files we successfully identified the array and orientation in all but one, where the number of variants left after quality control was too low for a definite match. In future we plan to add new features such as Affymetrix chips, position based checks and by extension genome build identification, as well as including the forward genomic strand amongst the possible strand options.

1460F**Machine learning approaches for comparative genome structure analysis.** *C. Rojas¹, M. N. Tran¹, L. Huynh¹, F. Hormozdizadeh^{1,2,3}*

1) University of California, Davis, Davis, CA; 2) MIND Institute, University of California, Davis, USA; 3) Biochemistry and Molecular Medicine, University of California, Davis, USA.

The development of high-throughput chromosome conformation capture techniques (Hi-C) has provided a wealth of data on the three-dimensional architecture of genomes. We can use this data to analyze the topological structure of genome and understand genomic interactions. However, the accurate approach to find conserved or specific genomic interactions in two or more Hi-C contact matrices is an open question. We introduce a convolutional autoencoder an unsupervised machine learning technique to produce a similarity function to compare areas between pairs of Hi-C matrices. Our model is trained on sub-blocks of the Hi-C matrix that are treated as high-dimensional vectors and that are transformed into lower dimensional vectors. We show that our autoencoder outperforms statistical methods such as principal components analysis (PCA) and root-mean-square error (RMSE) for finding genomic interactions which are specific to one of the matrices. This method is useful and accurate in finding genomic interactions specific to one genome which potentially result in changes in gene expression by comparing Hi-C matrices from two or more tissues or species.

1461W

OligoPVP: Phenotype-driven analysis of individual genomic information to prioritize oligogenic disease variants. *P.N. Schofield¹, I. Boudelloua², M. Kulmanov³, G. Gkoutos³, R. Hoehndorf³.* 1) Physiology Development and Neuroscience, University of Cambridge, Cambridge, United Kingdom; 2) Computational Bioscience Research Center, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia; 3) College of Medical and Dental Sciences, Institute of Cancer and Genomic Sciences, Centre for Computational Biology, University of Birmingham, B15 2TT, Birmingham, United Kingdom Institute of Translational Medicine, University Hospitals Birmingham, NHS Foundation Trust, B15 2TT, Birmingham, United Kingdom.

Identification of variants associated with genetic diseases is a major challenge; in particular in the analysis of clinical sequence data from individual patients. An increasing number of disorders have been identified for which two or more variants in one or more genes are required to cause the disease or significantly modify its severity or phenotype. It is difficult to discover such interactions using existing approaches. We have previously developed a pipeline, DeepPVP, that combines automated inference with deep neural networks to identify the likely causative variants in whole exome or whole genome sequence data. This exploits phenotypic similarity between patient phenotypes and the large body of genotype-phenotype data available for both humans and model organisms. Here, we present OligoPVP, which builds on Deep PVP to highlight potential causative variants in multiple genes for which there is evidence of direct protein-protein, network or regulatory interaction, predicated on the observation that variants in genes within the same pathway or network show similar or related disease phenotypes. Using synthetic genomes constructed from information in the Digenic Disease Database (DIDA: <http://dida.ibsquare.be/>) and disease annotations to the human phenotype ontology, we are able to recover established digenic and trigenic variants in the same genomes. Out of 164 whole genome sequences where two variants were inserted, we find both causative variants (i.e., the two variants we inserted) as the highest ranked variants in 88 cases (53.66%) and within the top ten ranks in 107 cases (65.24%). For the 25 cases of triallelic diseases, we find all three causative variants within the first three ranks in 10 cases (40.00%) and we find all three causative variants within the top ten variants in 14 cases. Our results show that OligoPVP can efficiently detect digenic and higher order interactions between variants in different genes using a phenotype-driven approach, and identify combinations of potentially etiologically important variants in whole genomes.

1462T

Automated disease-specific variant classification for rare autosomal recessive disorders. *M.C. Schu, S.W. Erickson, N. Whitehead.* Center for Genomics in Public Health and Medicine, RTI International, Research Triangle Park, NC.

Compiling accurate prevalence estimates and comprehensive inventories of disease-causing variants is an inherent challenge in characterizing rare Mendelian disorders. Because of the low incidence (e.g. 1 case in every 10,000 births or smaller), even large genetic screening studies that actively recruit for cases are likely to miss the full range of genomic aberrations that lead to the disorder. If, however, the causal gene is known for a disease of interest, population genetic databases can be leveraged to help fill in the gaps from large screening efforts and infer the population frequency of predicted disease-causing variants. Nevertheless, even with sophisticated functional classification tools, it is difficult to determine which predicted deleterious variants will alter the protein fitness to a degree that is sufficient for inducing the disease state in carriers. Manual curation of predicted deleterious variant might be used to improve the specificity of functional predictions; however, drawbacks to this approach include scalability limitations and concerns regarding reproducibility. As an alternative method for expanding the inventory of disease-causing markers, we present an automated method for predicting novel variants that are causal for autosomal recessive disorders. Our approach begins by compiling a training set of variants whose pathogenic status with respect to the autosomal recessive disease of interest has been well established in either ClinVar or other comparably rigorous literature sources. This list of confirmed causal and benign variants is then run through a battery of variant functional prediction tools, including PROVEAN, SIFT, and CONDEL. From the scores of these algorithms, a composite model is constructed that predicts the disease state of a hypothetical homozygous carrier for a given variant. The key innovation of our approach is directly predicting the disease state induced by the variant rather than the intermediate impact of the mutation on protein function. The resulting disease-specific classifier can be applied to the user's population database of interest to predict novel variants that are likely to be causal for the disorder in question. We have applied this approach to multiple rare genetic disorders and demonstrate the utility of the method for improving disease prevalence estimates and identifying new targets for genetic screening panels.

1463F

BayesDel and REVEL outperform other *in silico* meta-predictors for variant classification. Y. Tian, A. Chamberlin, R.B. Fenwick, T. Pesaran, C. Gau, S. Li, H.M. Lu, M.H. Black, D. Qian. Ambry Genetics, Aliso Viejo, CA.

Introduction: Many *in silico* predictors have been developed for assessing the pathogenicity of missense variants. However, evidence-based thresholds for assessing *in silico* scores are largely unavailable, and diagnostic laboratories utilize different scores and criteria for variant interpretation. **Methods:** We compiled a dataset of 4,094 missense variants in 66 genes with unambiguous ClinVar classifications, and evaluated 5 commonly used *in silico* meta-predictors: CADD, REVEL, Eigen, MetaSVM and BayesDel, and the combination of 2 individual predictors: SIFT/Polyphen2. We used logistic regression to compute gene-specific and global thresholds for assigning benign and pathogenic evidence based on the estimated probabilities of pathogenicity (0.2 and 0.8 for gene-specific, and 0.1 and 0.9 for global thresholds). We then assessed prediction performance using leave-one out cross-validation.

Results: Gene-specific thresholds for benign and pathogenic evidence varied widely among genes. For example, gene-specific REVEL and BayesDel thresholds for benign evidence ranged 0.027 to 0.710 and -0.517 to 0.252, respectively; gene-specific REVEL and BayesDel thresholds for pathogenic evidence ranged 0.501 to 0.969 and 0.002 to 0.550. REVEL and BayesDel had the highest positive and negative predictive values and overall prediction performance (PPV=97%, NPV=97%, and OPP=0.90 for each, respectively) and nearly equivalent yield (76.1% and 75.4%, respectively). Notably, BayesDel had slightly less variation in performance statistics than REVEL across all genes analyzed (OPP SD=0.04 vs. 0.06, respectively). MetaSVM had only slightly lower PPV, NPV and yield (96%, 96%, 74.1%), but greater variation in performance across all genes (OPP SD=0.09). Eigen and CADD had similarly high PPV and NPV (range: 91-97%), but substantially lower yield (53.1% and 52.7%, respectively). The combination of SIFT and PolyPhen2 also had similar NPV (95%), but substantially lower PPV (79%), OPP (0.80), and yield (62%).

Conclusions: REVEL and BayesDel outperform other commonly used *in silico* predictors, and their corresponding gene-specific or global thresholds for predicting pathogenicity can be used to provide accurate evidence for variant assessment in accordance with ACMG/AMP guidelines.

1464W

Quality control for 141,456 human exomes and genomes in the Genome Aggregation Database: An open-source, high-throughput computational workflow. G. Tiao¹, L. Francioli^{1,2}, K.J. Karczewski^{1,2}, K. Laricchia¹, J. Alföldi¹, L. Gauthier¹, T. Poterba¹, D. King¹, A. Wang¹, C. Seed¹, M. Daly^{1,2}, B. Neale^{1,2}, D. MacArthur^{1,2}, The Genome Aggregation Database (gnomAD) Consortium.

1) Broad Institute of MIT and Harvard, Cambridge, MA; 2) Massachusetts General Hospital, Boston, MA.

The Genome Aggregation Database (gnomAD) is the largest public collection of sequenced humans, currently spanning 141,456 unrelated individuals of diverse ancestry. The database is accessible online (gnomad.broadinstitute.org) and is widely used as an aid in clinical variant interpretation and as a general resource for human genetics research. With 125,748 exome and 15,708 whole-genome samples, the size of the gnomAD 2.1 dataset has made it impractical to analyze using standard software programs. In addition, current quality control (QC) analysis of sequencing data typically requires multiple pieces of software that are poorly integrated with each other. Leveraging Hail (<https://hail.is/>), an open-source framework for distributed genetics data analysis, we have built a fast, efficient, and comprehensive computational pipeline that automates sequencing data quality control and scales to hundred of thousands of samples. The pipeline can be parallelized locally or on a cloud platform; generates results that are highly reproducible; and requires little manual intervention to run. Our workflow is written entirely in python using the Hail library and requires no other software. Our workflow is divided into a sample QC component and a variant QC component. The sample QC component automatically infers sex, duplicate samples, familial relationships, and ancestry and then uses this information to compute relevant quality control metrics and generate plots to exclude low-quality samples. The variant QC portion of the pipeline computes allele-level metrics directly from genotype quality metrics and performs allele-level filtering using a random forest model. Variant quality metrics plots are generated. We show that this approach outperforms standard methods such as the Genome Analysis Toolkit (GATK) Variant Quality Score Recalibration (VQSR). Given its advantageous features and the modularity of its components, we believe the workflow will prove useful as a template to other groups working to analyze large genomic datasets and have released the code at https://github.com/macarthur-lab/gnomad_qc.

1465T

Descriptive analysis of identification and characterization in-silico of genomic susceptibility focus associated with rare diseases, a bioinformatic approach. F. Tobar-Tosse, C. López-Burbano, P. Hurtado-Villa. Department of Basic Science for Health, Pontificia Universidad Javeriana Cali. Cali - Colombia.

Introduction: Rare Diseases (RDs) are a group of several diseases with low prevalence, and depending on the country their definition varies ranging from 1 in 500.000 to 1 in 2.000¹, but all together affects a significant number of people. The World Health Organization (WHO) established that 8% of the world's population are affected by RDs² making a cumulative impact importance for public health. The 80% of RDs have genetic etiology, most of them related to single gene disorders³ and the rest with unknown mechanisms; this diversity makes it important to perform a detailed genetic and molecular approach in each disease that may come from areas such as bioinformatics, which facilitate the understanding of the physiopathology of these entities. **Objective:** to describe patterns among focus of genomic susceptibility associated with RDs, in order to characterize sensible regions to structural variations on the human genome through a bioinformatic analysis of the information available in public databases about RDs. **Methods:** a descriptive repository of phenotypic, clinical, molecular and epidemiological characteristics of RDs was created with information from public databases (*Office of Rare Diseases Research* and *OMIM*). Also, exploring specific databases (*Genebank*, *Genome Browser*, *Kegg*), a molecular repository was created, where genes, proteins, metabolic processes and gene expression (if there was a registry) for each disease were identified. Finally, a descriptive statistical analysis was made. **Results:** We found 145 RDs of importance due to its localization on region sensible to structural variation. The patterns identified were 14 and the genes within these patterns were 25. The genes involved in most biologic processes were *PIK3CA* (68,02%), *COL2A1* (9,72%) and *STAT3* (3,85%). Autosomal Dominant was the most recurrent inheritance pattern, while oncogenic diseases had the major number of biologic processes affected. **Discussion:** structural genomic studies have been established as effective methods for study RDs, since they allow to classify and describe more precisely the genes involved in each of these disorders³ as well as to establish possible associations between genomic elements and RDs. **Conclusions:** Here we present the first approach to established specific focus of genomic susceptibility for RDs. Further investigation in phenotypic correlation is needed in order to fully understand their genetics and epigenetic factors that enhance an aberrant phenotype.⁴.

1466F

Bioinformatic software pipelines strategies for analysis of amplicon-based assays for clinical applications. F. Vezzi. Devyser AB, Stockholm, Sweden.

Introduction: Next Generation Sequencing (NGS) is the new black when it comes to finding the genetic cause of inherited diseases as well as oncology. To effectively use NGS in everyday diagnostics a combination of easy laboratory workflow and easy to interpret data analysis are needed. To meet the first requirement, Devyser AB developed four single tube library preparation assays for both oncology (BRCA) and hereditary (Thalassemia THAL, Familial Hypercholesterolemia FH, and Cystic Fibrosis CFTR) conditions. To meet the second parameter, a combination of technical solutions and multiple supported analysis software allow to easily identify both Single Nucleotide Variants (SNV) and Copy Number Variants (CNV) in a stable and reliable way. A study on five different software tools has been conducted to evaluate the ability of the tools to correctly call SNV and CNV. **Method:** Four out of five tools are third party proprietary tools, while the last one is an internally developed tool used for QC and benchmarking. For each assay, we prepared and sequenced samples with previously characterized variants and analyzed each sample with all the tools supporting the kit. 158 samples have been analyzed with five tools for BRCA germline variants, 88 FFPE samples have been analyzed with four tools for somatic variants, 188 samples have been analyzed with five tools for CFTR variants, 24 samples have been analyzed with four tools for FH related variants, and 279 samples have been analyzed with three tools for THAL variants. **Preliminary Results:** All the tools show a clinical sensitivity >99.9% for SNVs, thus showing their ability to correctly identify SNV variants. Analyzing samples for germline variants, the number of False Positive (FP) calls is negligible. When analyzing samples for somatic variants and searching for low frequency variants we noticed the presence of tool-specific variants that can be discarded via a frequency database approach. Reliable calling of CNV in amplicon based panels is a difficult task. In order to improve CNV calling, all Devyser kits are designed to have a low amplicon run variability. In addition to this, special amplicons, dubbed Direct Detection amplicons (DD), are designed to span over important CNVs with known breakpoints. Tools able to combine standard CNV analysis to DD are not only able to correctly identify all possible clinically relevant CNVs, but can also facilitate the reporting of complex variants, for example in Thalassemia.

1467W

A suite of automated sequence analyses reveals a difference in the genetic burden of probands versus unaffected siblings. A. Wu, F. Gu, M. Gordon, L. Vlahos, S. Macnamara, A. Kim, E. Burke, M. Malicdan, C. Lau, D. Adams, C. Tift, C. Toro, T. Markello, W. Gahl. National Human Genome Research Institute, Bethesda, MD.

Continuing problems in genome-wide analyses for disease discovery prevent complete analysis of sequence data in small, informative regions of the genome. Most of these regions are typically excluded by the use of a global quality score threshold, potentially excluding true positive deleterious variants. The intramural NIH Undiagnosed Diseases Program (UDP) has developed a phenotype-blind, automated, agnostic pipeline that allows high-confidence genotyping in regions of low coverage and/or low variant quality score, and permits analysis in coding and noncoding regions for variant Mendelian models, including homozygous recessive, de novo dominant, compound heterozygous recessive, X-linked, hemizygous, and, in very well-sequenced positions, novel exon deletions. The recovery of true positive genotype calls involves combined strategies, including re-genotype calling with ethnically matched population prior probabilities and pedigree-aware conditional probabilities in nuclear families. Algorithms were developed that interrogate local BAM file regions to include true positive and reject false positive variant calls without using global quality scoring; instead the algorithms use pedigree inheritance, random base error density, and ethnically matched population frequency prior probabilities obtained from 1300 exomes of the UDP cohort. The pipeline also detects larger deletions on the exon-level for individuals with no coverage in regions well covered in other family members and in 1300 samples using identical chemistry. To evaluate the pipeline's effectiveness in capturing truly deleterious candidate variants as part of a person's genetic burden, two symmetrical analyses were performed using the pipeline. First, the proband was the subject of analysis ($n=97$) and, second, the proband's unaffected sibling(s) was the subject. This test resulted in short lists of deleterious candidate variants that numbered 6.6 per proband versus 5.8 per sib ($p<0.05$). This indicates that, for most of these cases, the true causal change for each proband is likely present in their respective output from our pipeline. These additional potential solutions, if validated, would nearly double the current 30% success rate for the UDP in quartet pedigree cases. We offer this pipeline to the community for new disease discovery.

1468T

Enhanced knowledge generation of rate-changes in transcriptional regulation using ontology rules, sentence structure and deep neural networks. Y. Zhang, W. Liu. Hubei University of Medicine, shiyan, China.

Considering the complexity of biosystems, it's very challenging to automatically and accurately generate biological knowledge from bio-literatures. The existing biological databases do not record temporal information of gene regulations, which are very important to understand the underlying mechanism of many diseases and biological processes. We previously constructed a corpus of time-delays related to the transcriptional regulation (bio-events) of yeast from the PubMed abstracts, summarized the knowledge rules of the bio-events as rate-changes in transcriptional regulation ontology, and obtained 86% accuracy by using the decision tree classifier with the ontology rule features. Deep neural networks (DNN) achieve great success in many machine learning applications including knowledge generation. The word2vec model learned the word embedding features from documents can achieve 50-70% accuracy on many text classification tasks. However, the sentence structure and domain knowledge are rarely considered in DNNs of document classification. We proposed to combine word2vec features, sentence structure, and our ontology rule features to improve DNNs for knowledge generation of rate-change in transcriptional regulation. Experimental results show that on predicting transcriptional regulation, the word2vec in DNN model achieves 73% accuracy, while our combined features in DNN with same parameters achieves 96% accuracy; on predicting the rate-changes in transcription regulation events, word2vec in DNN achieves only 59% accuracy, and our combined features in DNN achieves 90% accuracy. This shows the power of ontology rules and sentence structure features with DNN in knowledge generation.

1469F

MUNIn (Multiple tissue UNifying long-range chromatin Interaction detector): A statistical framework for identifying long-range chromatin interactions from multiple tissues. Y. Yang¹, M. Hu², Y. Li^{1,3,4}. 1) Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Department of Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic Foundation, Cleveland, OH; 3) Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC; 4) Department of Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, NC.

Chromatin spatial organization (interactome) plays a critical role in genome function. Deep understanding of interactome can shed insights into transcriptional regulation mechanisms and human disease ontology. One essential task of interactomic data analysis is to identify long-range chromatin interactions (i.e., peaks). Existing approaches, such as HiCCUPS, Fit-Hi-C and FastHiC, are all designed for single tissue analysis. None of them is able to account for unbalanced sequencing depths and heterogeneity among multiple tissues in a unified statistical framework. To address these challenges, we have developed a novel statistical framework MUNIn (Multiple tissue UNifying long-range chromatin Interaction detector) for identifying long-range chromatin interactions from multiple tissues. MUNIn adopts a hierarchical hidden Markov random field (H-HMRF) model, in which the status of each pair of interacting chromatin loci (peak or non-peak) depends not only on the status of loci pairs in its neighborhood region, but also on the status of the same loci pair in other tissues. To benchmark the performance of MUNIn, we performed comprehensive simulation studies and real data analysis. Compared to tissue-by-tissue analysis, MUNIn can achieve much lower false positive rates for detecting tissue-specific peaks (2.6 ~ 35.4%), and much enhanced statistical power for detecting tissue-constitutive peaks. Our data have demonstrated that MUNIn is a powerful tool for the integrative analysis of multi-tissue interactomic data.

1470W

ChIA-PoP: An integrated tool for ChIA-PET data analysis. W. Huang¹, L. Niu². 1) National Exposure Research Laboratory, US EPA, RTP, NC; 2) Department of Environmental Health, College of Medicine, University of Cincinnati, Cincinnati, OH.

Chromatin Interaction Analysis by Paired-End Tag Sequencing, or ChIA-PET, is a popular assay method for studying genome-wide chromatin interactions mediated by a protein of interest. The main goal of ChIA-PET data analysis is to detect interactions between DNA regions. Here we propose a new method and the associated data analysis pipeline, ChIA-PoP, to detect chromatin interactions from ChIA-PET data. We compared ChIA-PoP with other popular methods, including hypergeometric model (used in ChIA-PET tool), MICC (used in ChIA-PET2), ChiaSig and mango. The results show that ChIA-PoP outperforms those methods by detecting chromatin interactions more accurately. ChIA-PoP software tool is freely available to the public as an R package.

1471T

The NIH Genetic Testing Registry (GTR): A 2018 genetic test submission trends report. M.A. Hoepfner, A.J. Malheiro, M. Benson, S. Chitipiralla, B. Gu, V. Hem, C. Liu, Z. Maddipati, D. Maglott, G. Zhou, J.B. Holmes, B.L. Kattman. Medical Genetics and Human Variation, National Center for Biotechnology Information, Bethesda, MD.

The Genetic Testing Registry (GTR, <https://www.ncbi.nlm.nih.gov/gtr/>) is a free, searchable, online repository of clinical and research genetic tests voluntarily submitted by US and international laboratories. Tests can be molecular, cytogenetic or biochemical and evaluate somatic phenotypes, Mendelian disorders, and pharmacogenetic responses. Housed at NCBI, GTR standardizes test data, promotes transparency, and fills gaps in information about genetic tests. Atomized into orderable test units and indexed by key fields, GTR enables clinicians to find the most appropriate test for the patient. As of May 2018, GTR includes 54,612 genetic tests submitted by 507 laboratories from 40 countries, covering 16,426 genes and 11,208 conditions. We compared test submissions by field entries for June 2014 and May 2018. Entry rates for minimal fields such as 'Analytical Validity' remain consistent as they are required for test registration, and top values for 'Purpose of the Test' showed only minor changes from the earlier year. Primary methodology presents the most striking change: the top 5 methods in 2014 are Bi-directional Sanger Sequencing (B-Sanger; 52.2%), Multiplex Ligation-dependent Probe Amplification (MLPA; 10.2%), Uni-directional Sanger sequencing (U-Sanger; 7.8%), Comparative Genomic Hybridization (CGH; 7.1%), and Next-Generation (NGS)/Massively parallel sequencing (MPS) (6.2%); in 2018 NGS (53.2%) leads followed by B-Sanger (28.3%), MLPA (5.1%), CGH (3%), and U-Sanger (2.2%). This dichotomy is expected, as the clinical community remains dedicated to providing diagnostic services, though the methods to analyze samples have quickly changed. Test counts showing completion of recommended fields such as Target Population, evaluation methods for Variants of Uncertain Significance (VUS), Proficiency Testing Method, Clinical Utility, and Clinical Validity are all increased in 2018 compared to 2014 data, although there is a decrease in completion for these fields as a percentage of total tests (between 30-37%). GTR has massively expanded during recent years with enhanced efficiency in submission for more complex tests. There is a trend towards submission of tests employing high-throughput NGS. Recommended fields such as VUS evaluation and Clinical Utility are being submitted at a lower rate; however, with expansion in the number of tests available, this information is becoming increasingly important to the ordering clinician.

1472F

Variant Ranking Toolkit (VaRTK): An ensemble approach for rapid variant interpretation in clinical diagnosis. M. Manoharan, L. Iyer, GP. Singh, R. Gupta, R. Gupta. Bioinformatics, Medgenome, Bangalore, India.

Background: The growing availability of NGS platform for diagnosis of Mendelian disorders calls for robust methods to facilitate scalable variant prioritization. Distinguishing disease-causing genetic variants from a large number of candidate variants is a complex task. While there are several tools available to identify putative causal variants, their performance vary widely. Therefore new improved methods that can rank the rare coding variants based on their pathogenicity would help in narrowing down the search space and aid prompt diagnosis. **Methods** We developed an ensemble method (VaRTK) based on a logistic regression model trained using rare (MAF ≤ 0.01) putative benign variants from 1000G (8,993 variants) and disease causing missense variations (DM) from HGMD added after Jan 1, 2015 (11,016 variants). We used 18 different rank scores from 14 prediction tools as features. The model was tested using 55,208 DM mutations added in HGMD before 2015 and 108,968 benign variants from 1000G. An independent evaluation of VaRTK was also performed using data from ~1200 clinical samples which was manually interpreted for clinically actionable variants using standard filtering approaches.

Results: The Area Under the Curve (AUC) for VaRTK was 0.88 for the training data and 0.90 for the testing data. For ~1200 clinical samples, VaRTK was able to rank the reported variants among the top5 ranked variants in ~40% of the samples. For the remaining clinical samples, VaRTK prioritized the reported pathogenic variants in the top 100 and likely pathogenic variants in the top 500, thereby reducing the otherwise large search space of ~1000-1500 variants. On the contrary, other popular tools such as SIFT and Polyphen2 predict many variants to be deleterious, therefore decreasing their efficiency to be used as a ranking method. **Conclusions:** The observed results opens up the possibility of using ensemble methods such as VaRTK, MetaLR, MetaSVM in ranking variants for efficient clinical interpretation. This will largely expedite variant identification and enable scalability in a clinical setting reducing the interpretation time by 5-10 fold. Our analysis also highlights the issues with SIFT and Polyphen2 scores due to lack of specificity and recommends the use of ensemble approaches. We are further evaluating the use of other gene-based prioritization scores to enhance the utility our method.

1473W

The genesis of good visualizations - AVADIS Platform. *H.K. Adil, S. Malhan, S. Ramachandrala, A.K. Pabbathi.* Strand Life Sciences Pvt. Ltd., Bangalore, Karnataka, India.

Access, Visualize, Analyze, Discover (AVADIS) is the philosophy that lies at the heart of StrandNGS, our one-stop software for all sequencing data analysis. With the motto of deskilled decision analytics for biologists, we have created an architecture that allows researchers to dive into analysis, visualization, and modeling while the software takes care of the statistical, mathematical, and computational principles. Each visualization option in AVADIS is created by following engineering and design principles to ensure accurate depiction of large-scale data and facilitation of intricate interpretations. Additionally, all plots are configurable, interactive, and ease the exploration across plot types. In addition to the already existing metadata framework, elastic Genome Browser, data-specific gene views, and covariate analyses options, we have added a new functionality called the smooth scatter plot. Here, illustrating this plot we discuss the thought process that goes into the design of a good publication-quality visualization. When scatter plot has several points stacking up at a given (x,y), it is hard to visualize the distribution of points in the plot. Very often, multi-colored heat-scatters are used to indicate the density in unit squares. However, they are discrete and hard to read; leading to quick and false interpretation of the data patterns. To overcome this problem, a smoothed version of scatter plot can be used by applying image-blur techniques that are well known in image processing world. Once the scatter plot data is transformed into image by binning into some predetermined grid size, Gaussian blurring technique is used to get a smoothened image. The density of points at each location is then depicted by a monochromatic hue color scheme designed based on Ostwald color cones to give the most accurate visualization of the scatter density. Given that the AVADIS framework supports lassoing across views, the genes that are of interest could be highlighted in the smoothed scatter by selecting them in this plot itself or any other plot type that is using the same dataset. Therefore, this visualization plot derived from smoothening of closely stacked data points is of immense help in exploratory transcriptomics studies that aim biomarker identification. Using a large-scale single-cell RNA-Seq study with multiple cell types and developmental stages, we illustrate the purpose, perceptions, and aesthetics involved in the design of these visualizations.

1474T

Deep linear mixed models for structured, high-dimensional trait GWAS.

F.P. Casale¹, A.V. Dalca^{2,3}, N. Fusi¹, J. Listgarten⁴. 1) Microsoft Research New England, Cambridge, MA, USA; 2) Computer Science and Artificial Intelligence Lab, MIT, Cambridge, MA, USA; 3) A.A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA; 4) Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA, USA.

Genome-wide association studies (GWAS) have revealed thousands of associations between human diseases and genetic variants, but remain largely limited to simple phenotypes, which limits our understanding of the underlying functional mechanisms. For many diseases, genetic analyses of human anatomy as seen through medical images can help gain functional insights by revealing spatial, tissue and organ-level intermediate effects. However, the methodology for such analyses has not kept pace with the availability of data, owing in part to the challenges in properly modelling the structured, high-dimensional phenotype matrices and to the relatively low sample sizes. In this work, we propose a new statistical method for the genetic analysis of structured, high-dimensional phenotypes. Our approach combines the best of several paradigms: (i) linear mixed models for the proven capabilities in genetic association studies; and (ii) rich-capacity neural networks to extract latent (low-dimensional) phenotypes from high-dimensional data. The resulting deep linear mixed model generalizes mixed models for multiple traits [1] and for transformed phenotypes [2], and extends standard approaches in imaging genetics, which often decouple the phenotype extraction step from the genetic modelling. After benchmarking our method using rigorous simulations, we demonstrate it on ~5,000 brain MRI scans of patients from the Alzheimer's Disease Neuroimaging Initiative, revealing known and novel effects of risk loci for neurological and psychiatric disorders on brain structure. Importantly, our method can predict both localized and broad-scale structural changes associated with these variants. Finally, we analyze multiple mental-health-related measures from individuals in UK Biobank, where we extract and characterize a heritable and high-dimensional generalization of the neuroticism phenotype. Overall, our results demonstrate the general utility and broad applicability of our method and provide new functional insights in the genetics of degenerative and psychiatric disorders. [1] Casale FP, Rakitsch B, Lippert C, Stegle O. Efficient set tests for the genetic analysis of correlated traits. *Nature methods*. 2015 Aug;12(8):755. [2] Fusi N, Lippert C, Lawrence ND, Stegle O. Warped linear mixed models for the genetic analysis of transformed phenotypes. *Nature communications*. 2014 Sep 19;5:4890.

1475F

A fast hierarchical model that leverages multi-mapping reads for the quantitation of gene expression from droplet single-cell RNA sequencing data. K. Choi, M.J. Vincent, G.A. Churchill. The Jackson Laboratory, Bar Harbor, ME.

Single-cell RNA sequencing (scRNA-Seq) opens a new perspective on the dynamics of gene expression that can be obscured in whole tissue samples. However, quantifying gene expression from scRNA-Seq data poses multiple analytical challenges caused by amplification of the cellular RNA and the limited size of the RNA pool in individual cells. The current trend in droplet-based scRNA-Seq toward exceedingly larger numbers of cells -- with very low per-cell depth of coverage -- increases the sampling bias and technical variability. For example, high proportion of zero counts has been one of the major issues in quantitation of gene expression. We found the current practice of discarding multi-mapping reads (multi-reads) aggravates the challenge of data sparsity and inflated proportion of zeros. Since the read mappability is different from gene to gene, ignoring multi-reads also distorts the rankings of expression level, critical information for many nonparametric approaches in the downstream analyses of single-cell transcriptome profiling. We developed an expectation-maximization algorithm that leverages all the aligned reads including multi-reads and quantifies expressions in each cell at the isoform and gene level. Our hierarchical algorithm then combines information across cells for each gene and updates expression abundance in each cell by referring to how the gene behaves in other cells. In this model, zeros are not all same zeros as we can interpret them in the context of cell population dynamics. We show our partial pooling approach controls issues due to data sparsity in individual cells while maintaining cell-to-cell heterogeneity in the final counts we provide. We demonstrate that our approach improves the performance on downstream analyses such as identifying subpopulation of cells or detecting differentially expressed genes. We also provide a scoring test to determine whether the count data requires zero-inflated models and show that simple hierarchical models are flexible enough to manage inflated proportion of sampling zeros. Probabilistic allocation of multi-reads effectively controls the number of zeros to a degree that zero-inflated models are not a requirement. We implemented our algorithm as an open-source Python package, scBASE, that is available at <https://github.com/churchill-lab/scbase>.

1476W

RUFUS: Reference free variant detection improves accuracy and sensitivity. A. Farrell, D. Lee, W. Richards, G. Marth. University of Utah, Salt Lake City, UT.

We developed a novel k-mer based variant detection tool, RUFUS, that vastly improves specificity and sensitivity for germline and somatic/tumor *de novo* mutations, and may reveal some of the missing heritability in many genetic diseases. RUFUS is based on direct k-mer comparison, removing the reference from variant detection, and any associated reference bias. This vastly improves the detection of medium sized (20-500bp) insertions/deletions (INDELs) that current methods are unable to reliably detect: small variant detectors (e.g. GATK, FreeBayes) are effective at finding 1-20bp events in read alignments; and structural variant callers (LUMPY, WHAM, etc.) are effective at >500bp events where insert size variations and read coverage anomalies can be confidently detected. As a result, medium length, INDELs have been missed by most sequencing studies. We are currently applying RUFUS to 519 family quartets (mother, father, autistic child, unaffected sibling), a total of 2,076 samples sequenced by the Simon's Foundation Autism Research Initiative to 30x whole genome coverage. This will be the largest *de novo* variation study to date, and combined with RUFUS's sensitivity for all variant types and sizes will provide the most complete picture of *de novo* variation ever constructed. Our preliminary data on a 40 family pilot study has shown that the rate of medium length *de novo* events is twice that of structural events (12 medium-length vs 6 SV events), suggesting that these events may be more common than previously thought. In addition to increased sensitivity for variants of all sizes, RUFUS also shows far higher specificity over mapping based approaches. Previous research has suggested that the human per-nucleotide *de novo* SNV mutation rate is $\sim 1.25 \times 10^{-8}$, or roughly 75 mutations per generation. In our analysis of numerous disease family trio data sets at the University of Utah, RUFUS finds between 77 and 116 *de novo* mutations per child genome, including 74-101 SNV mutations; 98% also seen in mapping-based variant calls. Conversely, traditional mapping based methods call on average 150,000 *de novo* calls per child, dominated by mapping and reference errors, drowning out true variation, and post-processing and genome masking is necessary to improve these, still leaving thousands of *de novo* calls. RUFUS requires no filtering or masking of the genome, enabling true genome wide variant detection of all mutation types, at uniquely high specificity.

1477T

Preparing short read sequencing data for variant filtration and calling with deep learning. S. Friedman, L. Gauthier, Y. Farjoun, E. Banks. Broad Institute of MIT and Harvard, Cambridge, MA.

Initially developed for image processing, Convolutional Neural Networks (CNNs) have recently been applied to genomics, where they have advanced the state-of-the-art in a number of important applications. Several CNN-based algorithms for both variant calling and filtration have shown impressive results. To be processed by a neural network, a site of genetic variation must be encoded as a numeric tensor. The 'one-hot' encoding represents each genomic site as a 4-vector, one row for each DNA base, with a 1 in the row corresponding to the observed base and 0s everywhere else. This is a natural and unbiased encoding for raw DNA sequences, but when the input is short read sequencing data, as output by an Illumina sequencing machine for example, the tensor encoding involves many subtle and consequential choices. Here, we show several of these choices have profound impacts on the performance of CNNs. Specifically, we explore three ways to improve CNN performance, as measured by F1-score, AUC-ROC and AUPRC, via tensor encoding alone. All encoding methods are available in the GATK's new CNNVariant toolset. First, we quantify the importance of local re-alignment of the reads with a de Bruijn graph to resolve long insertions and deletions. Then, we show how read sorting by reference start position, as is often practiced in the literature, is inferior to sorting based on alternate allele support or pileup entropy. If a heterozygous site of variation exist in the region covered by the tensor, then sorting by the base read at this site clearly reveals the support for each of the haplotypes. This is especially important for CNNs that include convolution in the pileup direction, because the weight sharing in those convolutions assumes structure along the pileups. For variant calling applications, at the time of tensor creation, it is unknown if there are any heterozygous sites within the region covered by the tensor; nonetheless, sorting the reads by the pileup column with maximum entropy produces a similar effect. Lastly, we demonstrate a natural and efficient encoding for base qualities. Finally, varying CNN architectures by depth, width, connection-type and capacity we explore the relative importance of tensor encoding versus hyper-parameter optimization. Experiments across many CNNs with vastly different hyper-parameter settings establish the curious fact that in many cases the choice of tensor encoding is more consequential than the CNN architecture itself.

1478F

Medical ontologies for precision medicine and translational research.

M.A. Haendel^{1,2}, J.A. McMurry², N.L. Harris³, R. Relevo², C.J. Mungall⁴, P.N. Robinson⁴, C.G. Chute⁵. 1) Dept of Environmental and Molecular Toxicology, Oregon State University, Corvallis, OR., USA; 2) 1Department of Medical Informatics and Clinical Epidemiology, Oregon Health and Science University, Portland, OR 97239, USA; 3) Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA; 4) 4The Jackson Laboratory, Farmington, CT 06032, USA; 5) 5School of Medicine, School of Public Health, and School of Nursing, Johns Hopkins University, Baltimore, MD 21205, USA.

The quantity and variety of medical information is exploding, both in research labs and in the clinic. To realize the promise of precision medicine, informatics tools are needed to unify and leverage this data for medical decision making. For this, we need systematic ways to represent knowledge across multiple disciplines such that it can be reasoned over--in other words, we need ontologies. Ontologies go beyond most standard clinical taxonomies or controlled vocabularies by defining relationships between concepts to enable computational logical reasoning. However, ontologies and their corresponding machine learning implications are a young field; many of those who could benefit do not yet know what ontologies are or how to use them. We therefore recently developed an introductory review of medical ontologies including the literature from 2008 - 2017. There are many disease and phenotype ontologies, with varying approaches to data representation and a wide range of applications in biomedicine. The one used most often for rare disease diagnosis is the Human Phenotype Ontology (HPO), which encodes phenotypic abnormalities encountered in human disease. The HPO currently includes over 13,400 terms, most of which have layperson synonyms associated with them. By representing specific phenotypes in a structured, hierarchical form, disease ontologies allow patient phenotype profiles to be computationally compared with the typical presentation of known diseases to aid in diagnosis. Resources such as the HPO also enable patients to record their own phenotypic data (using tools such as Phenotypr.com) to facilitate discussions with their physicians and help improve diagnosis and treatment. The creation and application of models of disease, phenotypes, and the environment using interoperable ontologies will be a game changer, with profound impact on patients, diagnostic practices, and healthcare. Clinical judgment will always remain central to medical decision making, but by unburdening physicians from routine information retrieval tasks and by providing a computational foundation for the interpretation of medical data, ontologies are accelerating our progress towards true precision medicine.

1479W

Updating an efficient pipeline for local ancestry inference. *E.M. Jewett, G.D. Poznik, K.F. McManus, K. Bryc, J.L. Mountain, A. Auton, 23andMe Research Team.* 23andMe, Mountain View, CA.

The goal of local ancestry inference is to identify the population of origin at each base in the genome of an individual. At 23andMe, inference is performed by an efficient pipeline known as Ancestry Composition. Ancestry Composition employs support vector machines (SVM) to assign local ancestries to windows along the genome, as well as a hidden Markov model to combine information across window-level SVM calls. The Ancestry Composition pipeline is computationally efficient and has demonstrated low error rates on both simulated and real data. Here, we explore methods for improving the window-level classification component of the pipeline by introducing a method based on the positional Burrows-Wheeler transform (PBWT). This adaptation of Ancestry Composition improves precision and recall, enables window-free local ancestry estimation, and improves computational efficiency.

1480T

Exploring the consistency of the quality scores with machine learning for next-generation sequencing experiments. *M. Oh¹, E. Cosgun².* 1) Virginia Tech, Blacksburg, VA 24061; 2) Microsoft AI & Research, Redmond, WA 98052.

Next-Generation Sequencing (NGS) enables massively parallel processing, allowing lower cost than the other sequencing technologies. In the subsequent analysis with the NGS data, one of the major concern is the reliability of variant calls. Although researchers can utilize raw quality scores of variant calling, they are forced to start the further analysis without any pre-evaluation of the quality scores. Here, we present a machine learning approach for estimating quality scores of variant calls derived from GATK best practice powered by Microsoft Genomics service. Based on three machine learning algorithms, including Multivariate Linear Regression (MLR), Random Forest Regression (RFR), and Neural Network Regression (NNR), data-driven predictive models were trained on variant call format (VCF) files, which contain technical values of GATK annotation module for each variant call. We analyzed correlations between the quality score and these annotations, specifying informative annotations which were used as features to predict variant quality scores. Some annotations that are simple statistics or scaled values originated from another annotation were excluded from the features. To test the predictive models, we simulated twenty-four paired-end Illumina sequencing reads with 30x coverage based on twenty-four artificial human genomes generated by perturbing a human reference genome with an injection of a wide range of variants, including SNPs, small indels, and large structural variants. Also, twenty-four human genome sequencing reads resulted from Illumina paired-end sequencing with at least 30x coverage were secured from Sequence Read Archive (SRA). Using Microsoft Genomics service, VCFs were derived from simulated and real sequencing reads. We trained the regression models on training data obtained by splitting each VCF data into training and test data, evaluating the prediction performance on the test data. We observed that the prediction models learned by RFR outperformed other algorithms in both simulated and real data. The quality scores of variant calls were highly predictable from informative features of GATK Annotation Modules in the simulated human genome VCF data (R^2 : 96.7%, 94.4%, and 89.8% for RFR, MLR, and NNR, respectively). Interestingly, the robustness of the proposed data-driven models was consistently maintained in the real human genome VCF data (R^2 : 97.8% and 96.5% for RFR and MLR, respectively).

1481F

Exploring disease-associated regions across genes and gene families at mbv.broadinstitute.org. E. Pérez-Palma¹, P. May², S. Iqbal³, H. Heyne^{3,4}, P. Nürnberg¹, A. Palotie^{3,4,5}, M. Daly^{3,4,5}, D. La^{1,3,4,6,7}. 1) Cologne Center for Genomics, University of Cologne, Cologne, Northern-Rhine Westphalia, Germany; 2) Luxembourg Centre for Systems Biomedicine, University Luxembourg, Esch-sur-Alzette, Luxembourg; 3) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 4) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA; 5) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; 6) Epilepsy Center, Neurological Institute, Cleveland Clinic, Cleveland, OH; 7) Genomic Medicine Institute, Lerner Research Institute Cleveland Clinic, OH.

Protein key functional regions are more frequently subject to pathogenic variants in patients than to neutral variants in controls. However, patient/control gene level comparisons of missense variant distribution is still underpowered to identify disease-associated regions. Yet, key functional regions are conserved among gene family members. We hypothesize that gene family members display a similar missense variant distribution pattern. Our aim was to leverage gene-family information to increase the statistical power in patient-vs.-control variant burden analyses to identify disease-associated regions and guide pathogenic interpretation. Gene family information was retrieved from the Biomart platform of Ensembl v92 and only the canonical transcript was considered. Protein alignments were performed using the MUSCLE software. "Neutral-control" and "Pathogenic-patient" missense variants were extracted from gnomAD v0.2 and ClinVar/HGMD databases (Feb, 2018), respectively. We implemented a binary approach to map the variants in each gene and gene-alignment. Further, a Fisher's exact test was performed to identify "missense pathogenic regions" (MPRs) defined as amino acids (aa) regions that are at the same time depleted from variation in the general population and enriched with pathogenic variants. We mapped 2,219,811 gnomAD and 97,391 ClinVar missense variants into 9,991 protein sequences belonging to 2,871 gene families. We identified 398 MPRs spanning 4,827 aa in 1,090 genes. In addition, single-gene analysis was able to identify 225 additional significant MPRs involving less than half of the aa ($n = 2,344$) detected with family alignment. For 46.7% (2,254/4,827) of the aa in MPRs no patient variants have been reported yet. To evaluate the utility of MPRs, we tested MPR enrichment in a set of 84 disease-associated genes vs. all other genes for *de novo* missense variants from 11,000 trios with neurodevelopmental disorders. We observed enrichment for variants within MPRs in known disease-associated genes compared to other genes ($p=1.09 \times 10^{-31}$). Our method can evaluate pathogenicity of positions across the linear protein sequence alignment and highlight disease-associated regions that would not have been identified through single-gene approaches. All results were deployed online into a user-friendly platform available at <http://mbv.broadinstitute.org> where MPRs as well as the raw gnomAD/ClinVar missense burden data can be explored at a gene- or gene-family level.

1482W

AVX2 based hardware acceleration in StrandNGS. A. Pal, S. Ramachandru, S.C. Cherukuri. Strand Life Sciences Pvt. Ltd., Bangalore, Karnataka, India.

Algorithms that identify similarity between two nucleotide sequences lie at the heart of aligning Next Generation Sequencing (NGS) data. The standard methods identify the substrings of maximal similarity, as defined by similarity matrix, for a pair of two nucleotide sequences. The problem could be solved by exact dynamic programming (DP) method using Smith-Waterman algorithm which is applied to determine if the read matches around the area anchored by the seed. Given the prominence of this step, coupled with its inherent complexity, many commercial and open-source software solve this problem in varied ways. Additionally, there are recommendations from these and other vendors to couple these software with expensive, specialized hardware. Here, we propose how a now commonplace hardware feature could be used to the optimum and also save time. StrandNGS, our flagship NGS analysis software could accelerate Smith-Waterman algorithm on Intel and AMD heterogeneous architectures. StrandNGS takes advantage of Single Instruction Multiple Data (SIMD) computing through the use of Advanced Vector Extensions 2 (AVX2) extensions on the processor. On a heterogeneous platform, using AVX2 instructions, Smith-Waterman dynamic programming takes 30 mins for execution, whereas, on a normal hardware (without SIMD) it takes 3-4 hours. AVX2 can execute sixteen 16 bit instructions simultaneously, that means 16 different seed alignments can be performed in parallel in a thread. So, if a machine has 16 threads, 256 (16x16) DPs can be performed simultaneously. Here the most advantageous part is that we are making parallel processing using AVX2 which comes as an in-built processor in all new systems including commodity hardwares like Intel(R) Core(TM) i5-4300U, AMD Zen processor etc. StrandNGS integrates SSE2 and AVX2 seamlessly so that if an old machine doesn't support AVX2, it will automatically switch to SSE2. There is no need of either any high configurable machines nor any external plug-ins. Hence, this is more cost effective for the users. We intend to showcase how StrandNGS continues to make best use of cutting-edge technological advancements for a continuous improvement of NGS analyses efficiency and thereby enhancing user experience.

1483T

Predicting pathogenicity of missense variants using deep learning. Y. Shen^{1,5,6}, H. Zhang¹, H. Qi^{1,2}, C. Chen^{1,3}, W.K. Chung⁴. 1) Department of Systems Biology, Columbia University, New York, NY; 2) Department of Applied Mathematics and Applied Physics, Columbia University, New York, NY; 3) Department of Biological Sciences, Columbia University, New York, NY; 4) Departments of Pediatrics and Medicine, Columbia University, New York, NY; 5) Department of Biomedical Informatics, Columbia University, New York, NY; 6) JP Sulzberger Columbia Genome Center, Columbia University, New York, NY.

Accurate predictions of genetic effects of missense variants is critical to the interpretation of genome sequence. Previously published prediction methods have been used to improve power in genetic studies and identify pathogenic variants in clinical genetic testing. However, the performance of these methods is suboptimal, due to several issues, including the complexity of the mechanisms of pathogenic variants, false positives in the training data, and suboptimal usage of large training data sets. Here we report on a new prediction method MVP. MVP uses a deep learning approach to learn protein-context-specific pathogenicity of amino acid changes from large number of curated pathogenic variants. We define the protein context as a window centered on the missense variant. For each residue in the context, we collect information about structural and biophysical properties, evolutionary conservation, and depletion of variation in general population. MVP uses convolutional networks to learn salient features without prior knowledge. Using cancer hotspot mutations as a benchmark, MVP achieved an area under the receiver operating characteristic curve of 0.9, substantially better than other published methods. Applied in genetic analyses of autism and congenital heart disease, MVP had better performance in prioritizing pathogenic *de novo* missense variants than other published methods. To further assess the utility in clinical genetic diagnostic testing, we obtained functional data of ~1800 missense variants (441 deleterious) in *BRCA1* and ~1800 missense variants (258 deleterious) in *PTEN* benchmarks. The area under the precision-recall curve of MVP is 0.76 and 0.81 in *BRCA1* and *PTEN*, respectively, substantially better than previously published methods (0.72 and 0.67, respectively). Setting a threshold to achieve a positive predictive value (PPV) of 0.9, previous methods only correctly classify 20-25% of pathogenic variants, categorizing most of the true pathogenic variants as variants of uncertain significance (VUS). MVP is able to identify 40% of true pathogenic variants with the same PPV threshold, substantially reducing the number of VUS. In summary, we show that MVP significantly improves predictions of pathogenic missense variants by capturing protein context-dependent information using deep learning, and it has the potential to significantly increase the power in genetic studies and enhance the utility of clinical genetic testing.

1484F

MOPower: An R-Shiny application for the simulation and power calculation of multi-omics studies. H. Syed, G. Otto, D. Kelberman, C. Bacchelli, P. Beales. GOSgene, Genetics and Genomic Medicine, UCL GOS Institute of Child Health, University College London, London, United Kingdom.

In recent years there has been an emergence of multi-omics studies, which seeks to integrate information from the genome, transcriptome, epigenome, metabolome and proteome to better understand clinical phenotypes. Power calculators are currently available for the design of genomic and RNA-seq experiments individually for a variety of outcomes (binary, quantitative and survival traits). However, software is not available which encompasses all this information together with data from other high-throughput omics technologies. Doing this is of particular relevance and benefit in rare disease studies where the sample size is small. With the rapid emergence of multi-omics studies and the complexity of information they consider, there is a need for software to perform power calculations over a range of design scenarios and analytical methodologies. We have developed the interactive R-shiny web application MOPower to perform statistical data simulation and power calculations for multi-omics studies with binary and longitudinal outcomes. The application allows simulation of SNP, rare variant, gene expression and methylation data over a range of rare disease study designs (acute and chronic). MOPower offers many different integration and analytical approaches such as joint regression modelling, multi-factor analysis and multivariate projection-based methodology. The tool calculates the power to detect an association with the phenotype of interest over a specified number of replicates. Sample pilot study data can be uploaded to the application to determine the realistic calculation of study design parameters such as fold change, average MAF and mean gene expression, thereby increasing the accuracy of the simulated data. In conclusion, MOPower will aid in the design of multi-omics studies, providing investigators with essential information on optimal sample size and choice of analytical model. With the increasing wealth of omics data from multiple sources, arises a multitude of issues regarding data integration and analysis. Consequently, this application will help with the development of future multi-omics data integration methods through the comparison of existing methodology.

1485W

CNV analysis of whole exome sequencing data in a clinical diagnostic setting. D. Becker¹, M. Ziegler¹, Y. Dincer^{1,2}, R. Brumm¹, K. Pindl¹, J. Schulz¹, K. Mayer¹, S.H. Eck¹, H.G. Klein¹. 1) Center for Human Genetics and Laboratory Diagnostics (AHC), Martinsried, Bavaria, Germany; 2) Chair of Social Pediatrics, Technical University of Munich, Munich, Germany.

The implementation of Next Generation Sequencing (NGS) in a clinical diagnostic setting opens vast opportunities through the ability to sequence all genes contributing to a certain indication simultaneously at a cost and speed that is superior to traditional sequencing approaches. Furthermore, these data allow the identification of small to medium sized copy number variations (CNVs). While the detection of single nucleotide polymorphisms (SNVs) and small insertions and deletions (InDels) is very straight forward, the reliable detection of CNVs from NGS data still poses several problems during data analysis. Here we present an approach for CNV calling based on whole exome NGS data in a clinical diagnostic setting as part of our multicentric Multiple Integration and Data Annotation Study (MIDAS). The application of more than one CNV calling algorithm provides significant improvements in sensitivity and robustness against sample variability in the normalization process. The current pipeline incorporates ExomeDepth, CoNVaDING and XHMM as primary CNV detection algorithms as well as VisCap for the visualization of complex events. Results are verified by qPCR to rule out false positive calls. While it became apparent that calling for targeted panel enrichment NGS data lead to acceptable verification effort through a small number of called events, whole exome (WE) data contains more events and therefore poses a much greater challenge. WE Trios are generally filtered for SNV analysis based on quality of the variants, the effect for the protein sequence and inheritance patterns. In contrast, CNVs neither come with an abundance of quality parameters, information about zygosity or annotations, nor are they easily comparable. Various filtering steps, calculation of frequencies and comparison of events in index and parent samples enable us to prioritize certain results, but not to permanently filter out calls. To gain the ability to reduce the noise and identify false positives, it is paramount to build a database and review the mapping of putatively true positive events. Our solution is based on a combination of our in house MIDAS software in combination with a graphical analysis in the integrated genome viewer (IGV). The IGV browser enables the analyst to compare calls from different samples and databases like Decipher, DGV and ExAC CNVs while MIDAS can calculate frequencies and annotates CNVs with further information like occurrences in an indexes parent.

1486T

LinkedSV: Detection of mosaic structural variants from linked-read sequencing. L. Fang¹, C. Kao², M.V. Gonzalez², F.A. Mafra², R. Pellegrino da Silva², M. Li³, H. Hakonarson², K. Wang¹. 1) Raymond G. Perleman Center for Cellular and Molecular Therapeutics, Children's Hospital of Philadelphia, Philadelphia, PA19104; 2) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA19104; 3) Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, PA19104.

Reliable detection of structural variants (SVs) from short-reads remains challenging, because the breakpoints are often located in repetitive regions. Linked-read sequencing developed by 10X Genomics provides long-range information by adding barcode tags to short-reads that belong to the same "fragment", and thus has the potential to improve SV detection. However, SV detection from linked-read sequencing data is still in the early stage. Currently available methods cannot reliably detect SVs from capture sequencing (e.g. whole-exome sequencing, WES), somatic mosaic SVs that have low variant allele frequencies (that is, only a small fraction of cells contain the SVs), and SVs for which the breakpoints have no coverage or located in repeat regions. To address these challenges, we introduce LinkedSV, a novel open source SV caller for linked-read sequencing. LinkedSV detects SV breakpoints by incorporating two types of evidence. First, fragments with shared barcodes between two genomic locations and the enriched fragment endpoints near breakpoints provide clues that a SV may be present nearby. Second, reads on two sides of a genomic position with different barcodes are likely derived from DNA molecules with a SV breakpoint. While existing methods utilize the first type of evidence, to the best of our knowledge, LinkedSV is the only SV caller that uses the second type of evidence to detect SV breakpoints. Integrating both types of evidence, we designed a statistical model in LinkedSV to detect SVs from whole-genome sequencing (WGS) and WES data. We assessed the performance of LinkedSV on three simulated data sets and one real data set, and compared its performance with all existing linked-read SV callers including Longranger, GROCC-SVs and NAIBR. When the variant allele frequency (VAF) was 10%, LinkedSV had a recall of 0.761, which was 72% higher than the second best SV caller NAIBR. On the simulated linked-read WES data, LinkedSV also had the highest recall (0.79) and highest F1 score (0.86). In addition, LinkedSV is the only SV caller that successfully detected an inversion in the *F8* gene from a clinical WES data set, where the breakpoints were located in intronic and intergenic regions. In summary, LinkedSV greatly improved the detection of mosaic SVs from linked-reads genome or exome sequencing data.

1487F

Development of structural variation calling in clinical whole genome sequencing: A progress report. *A.M. Gross, A. Chawla, S.S. Ajay, N. Nariai, X. Chen, C.T. Saunders, D.R. Bentley, R.J. Taft.* Illumina, San Diego, CA.

Here we report progress on the incorporation of structural variants (SVs, insertions, deletions, inversions, and non-homologous chromosomal breakends) into a clinical whole genome sequencing (cWGS) test for the diagnosis of rare and undiagnosed genetic disease (RUGD). We assess the utility of incorporating paired and split read evidence across a pedigree to generate and genotype SV calls using the Manta variant caller. These SVs are used to supplement and extend SNV/INDEL/CNV calling pipelines, and can be combined with depth-based CNVs to provide breakpoint resolution of CNV boundaries, classify copy number gains (as tandem, inverted or dispersed duplications), and link multiple distinct CNVs to resolve complex rearrangements such as unbalanced translocations and DUP-INV-DUP structures. In addition, SV calls are incorporated as priors in a targeted depth-based CNV calling pipeline, expanding CNV detection limits from 10kb (our depth based caller's cutoff) to 1kb. Finally, SV calls are used to increase the upper bound of standard small INDEL calling to the 50bp-500bp range, which includes single exon deletions. To limit interpretation burden, we implement a number of variant filtering and prioritization steps. To improve specificity, we apply a black-list of variable or low complexity regions, which filters 66% of false positive insertions and deletions, with marginal loss of clinical sensitivity (2% of exons are masked). Rare variation is prioritized via practical heuristics for population frequency assessment using a background population of 2,400 samples. Within a test development setting, we used this pipeline to identify a 276bp single exon X-linked recessive deletion on L1CAM and a 4kb inherited NOTCH3 deletion. In addition we used SV calls to characterize over 25 pathogenic CNVs including resolution of unbalanced translocations and dispersed duplications. While the generation of short-read based SV callsets with clinical specificity remains an ongoing challenge, here we show encouraging progress towards their incorporation into a cWGS test for RUGD.

1488W

Cobalt: Improved detection of germline copy number variants from hybrid-capture NGS data. *B.D. O'Fallon, J. Durtschi, T. Lewis, D. Close, H. Best.* Biocomputing, ARUP Labs, 500 Chipeta Way, Salt Lake City, UT 84108.

Purpose Accurate detection of germline copy number variants (CNVs) from hybrid-capture next generation sequencing (NGS) data remains a significant computational challenge. While many CNV detection methods have been described, several recent reviews have found relatively low sensitivity and reproducibility among these tools, especially for CNVs spanning a small number of capture targets. **Methods** We introduce an improved method for detecting germline CNVs from read depth data and describe two algorithmic advances that increase sensitivity and reduce the number of false positive variants detected. First, we propose a method for estimating target-specific parameters for the emission distributions used to construct the Hidden Markov Model (HMM) for computing CNV probabilities. Secondly, we show that an alternative method of HMM decoding known as "pointwise maximum a posteriori" yields increased sensitivity compared to the more commonly used Viterbi method. Using both simulated and real, orthogonally confirmed CNVs we compared our new method to six other recently developed CNV detection tools. Evaluation samples included 100 human exomes captured with IDT xGen Exome Research Panel v1.0 and sequenced on 4 independent runs of an Illumina HiSeq 4000 instrument, 37 of which contained a previously detected CNV. Reads were aligned with BWA MEM (v0.7.12) and duplicate marked using Picard Tools (v2.9). All callers were individually tuned to maximize overall accuracy as measured by the F₁ statistic. For the simulation study, CNVs 1, 2, 3 and 10 exons in size were introduced into alignment (BAM) files by removing or duplicating reads, and sensitivity and positive predictive value (PPV) were computed for each CNV size class. For real, previously detected CNVs a 50% reciprocal overlap threshold was used to determine correctly called variants. **Results** In the simulation analysis, our method achieved the highest mean sensitivity and PPV of the callers evaluated in each CNV size class. Sensitivity increased from 87% for single-exon deletions to 94% for deletions spanning 10 exons, while PPV ranged from 94%-95%. On exomes containing previously detected CNVs, our method yielded sensitivity indistinguishable from other top-performing callers, but with significantly fewer CNVs called per exome. Our method identified a median of 32 CNVs per sample, while similarly sensitive callers typically identified 300-500 CNVs per sample.

1489T

GATK-SV: Integrating Structural Variation calling in GATK. V. Ruano-Rubio¹, S. Huang¹, T. Sharpe¹, T. Brookings¹, M. Walker¹, C. Whelan¹, X. Zhao^{1,2}, R. Collins^{1,2}, H. Brand^{1,2}, M. Stone^{1,2}, E. Banks^{1,2}, C. Nusbaum^{1,2}, M. Talkowski^{1,2}. 1) Data Science Platform at the Broad Institute of MIT and Harvard, Cambridge, MA; 2) Massachusetts General Hospital, Boston, MA.

The Genome Analysis Toolkit (GATK) software is well known and widely used for calling single-nucleotide polymorphism (SNP) and short insertion-deletion (sINDEL) variants in germline DNA-seq data. The latest major release (GATK4+) includes tools to call additional types of variation including small variants in somatic sequencing data and also read-depth based copy number variation (CNV) calls in either somatic and germline datasets. GATK-SV, the latest addition to GATK, provides a structural variation (SV) detection pipeline that integrates multiple SV evidence sources to GATK's variation calling tool set. GATK-SV's algorithmic pipeline includes several stages starting with (i) an initial phase for coverage-bias detection and normalization (ii) sample QC and batching, as these two factors can greatly affect analyses based on read-depth (RD) evidence. Then we progress to (iii) multi-typed evidence collection, plausible break-point detection and filtering. We gather RD, paired-end (PE), and split-reads (SR) evidence and identify likely regions of variation using machine learning with a boosted tree classifier. In order to determine accurate break points we perform (iv) local de-novo assemblies using reads that map to the candidate regions, their mates and other reads that share the same k-mers. Then (v) we compare the reconstructed haplotypes to the reference to classify variation into different types including insertions, deletions, inversions, tandem or dispersed duplications, and arbitrary translocations. Finally we (vi) genotype each variant using the input reads and provide likelihoods of possible genotype in the input sample. A distinctive feature of GATK-SV with respect to other methods is that it uses the Apache Spark distributed computing engine to process large data sets across parallel compute nodes with low latency hosted on the cloud. We provide tools to run the pipeline in Google cloud Dataproc and will make it available in the Broad Institute Fire-Cloud methods repository. In this presentation we will provide details on our algorithm and the benchmarking results including variation calling accuracy and cost per sample. We tested the GATK-SV prototype on the CHM1/CHM13 lines and found the performance was encouraging. In addition to the required processing speed for a scalable tool, the combination of local assembly, PE/SR, and RD signatures appears to be relatively sensitive and comparable with the most sensitive algorithms available.

1490F

CCGT: A database for sharing Chinese children's genetic testing results. X. Dong, Y. Lu, Q. Ni, H. Wang, B. Wu, L. Yang, B. Liu, W. Zhou. Children's Hospital of Fudan University, Shanghai Key Laboratory of Birth Defects, The Translational Medicine Center of Children Development and Disease of Fudan University, Shanghai 201102, P.R. China.

Background Genetic rare diseases are one of the essential causes of children deaths, often lead to adverse consequences. Rapid and accurate diagnosis could assist therapeutic strategies, reduce adverse consequences and families could be benefit in further fertility plans. Children's Hospital of Fudan University has successfully carried out the Children's Genetic Testing Project, using WES (whole exome-sequencing) and CES (clinical exome-sequencing) to perform genetic testing for children with rare diseases. So far, the project has benefited about 20,000 patients with NGS testing, and nearly 10,000 patients have been diagnosed. Considering the population background, the testing results make a powerful database for rare disease diagnosis in China, offering complementary to existed variant annotation databases such as ClinVar or HGMD. Therefore, we build a server to share the result of these genetic testing with physicians, rare disease researchers and genetic counselors. **Methods & Results** We reviewed 8,742 testing cases with diagnostic report from the last two years, including a total of 14,000 samples (including parents). Variants called from NGS data were first filtered with public population allele frequency and sequencing target region, and then enrolled in the database. The first feature of our database is to calculate and provide local allele frequency separately for WES and CES. Second, for each variant the database would provide summarized clinical phenotype for patients who carried it. Phenotypes were presented in major systems and divided into three groups depending on the diagnostic conclusion of the variant. Variants were labeled with Positive/Likely positive, Related/Uncertain and Additional Findings. The diagnostic conclusion was based on variant's pathogenic level, inheritance model and the match of patient's phenotype. For each system, percentage of patients with related phenotypes among all patients with the same conclusion level would be presented. In this way the patients' privacy was well protected. Finally, if you are interested in further research, you could fill in the online forms for follow-up communication. The current URL is <http://101.230.2.57:8090/>, and users need to register for acquiring information. **Conclusion** This server will be maintained by the translational medicine center of our hospital. We hope that by sharing the above information we can help the diagnosis of genetic rare disease and the interpretation of genetic variant.

1491W

Variant allele frequency discrepancies in population databases. E.

Kaseniit, C. Cropper, D. Previte, D. Muzzey. Counsyl, South San Francisco, CA.

IntroductionThe allele frequency (AF) of a variant in different populations is often used as evidence when classifying the clinical significance of the variant. For instance, arguments against a variant being classified as pathogenic include it being too common in controls or not significantly more frequent in cases compared to controls. Two common sources for these control AFs are gnomAD and ExAC. We compared the AFs between these two datasets for variants in 176 genes associated with serious Mendelian conditions to assess whether AF discrepancies due to sequence context may influence variant interpretation. **Methods**Hard-to-sequence regions were defined from Global Alliance for Genomics and Health resources. The two-tailed Fisher's exact test or the chi-squared test (if all contingency table values >5) was used to calculate p-values of pairwise AF comparisons, and a Bonferroni correction was applied. Only entries with the PASS filter were included from ExAC and gnomAD; gnomAD entries with the lcr or segdup flags were ignored. **Results**The AFs of 932 out of 148,269 variants analyzed were found to be significantly different between ExAC and gnomAD in at least one of six considered subpopulations. These "discrepant variants" were enriched in hard-to-sequence regions, which included 15.1% of discrepant variants but only 4.5% of total variants analyzed ($p < 1e-10$). Most discrepant variant AF comparisons (51.6%) had gnomAD AF $>5\%$. Typically, an AF cut-off of 5% is used to classify a variant as too common to be pathogenic; for 123 (7.7%) discrepant variant AF comparisons, the AF was $>5\%$ in one dataset but $<5\%$ in the other, with a median absolute difference of 3.6% points. Most discrepant variants differed in their AF in one ethnicity, but 30.3% (282) had different AFs in two or more populations. Out of 2,138 variants that were evaluated in each of the six populations, 40 were discrepant variants with significantly different AF in all evaluated populations, and 14 (35%) were in hard-to-sequence regions. In general, ExAC tended to have the higher AF for discrepant variants compared to gnomAD, with a median fold-difference of 1.6. **Conclusions**Our results demonstrate that AF data must be carefully considered during variant classification—even for high-frequency variants—because AF values can differ greatly in common population databases. These differences are enriched in hard-to-sequence areas and may only be revealed via comparison among subpopulations.

1492T

Increasing phenotypic and functional evidence in ClinVar. M.J. Landrum, S. Chitipiralla, M. Benson, G. Brown, C. Chen, B. Gu, J. Hart, D. Hoffman, W. Jang, C. Liu, Z. Maddipatla, R. Maiti, G. Riley, G. Zhou, D.R. Maglott, A. Glodek, J.B. Holmes, B.L. Kattman. National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, USA.

ClinVar is a public archive for submitted reports of human genetic variants and their significance for human health and disease. The database currently holds more than 675,000 submitted interpretations representing 417,000 unique variants. The data in ClinVar have been submitted by clinical testing laboratories, research laboratories, curation groups, OMIM, *GeneReviews*, and expert panels. Most data in ClinVar was submitted by clinical testing laboratories. However, these labs often lack access to certain types of evidence that would aid their ability to interpret variants, such as phenotypic information for patients and functional evidence. ClinVar provides opportunities to submit both of these types of evidence. ClinVar welcomes submissions from clinical providers and patient registries, which are rich sources of phenotypic data for patients who have had genetic testing. Clinical providers may submit their own interpretation of the variant, if they have re-evaluated a variant reported by a testing lab. Both clinical providers and patient registries can submit detailed information about the phenotype of a patient with a variant, along with the date the phenotype was evaluated, the name of the testing laboratory, and the interpretation from the testing laboratory. ClinVar currently holds a small number of interpretations from clinical providers, and more than 500 submitted records that focus on phenotyping only, mostly from GenomeConnect, the ClinGen patient registry. Researchers who have developed functional assays, animal models, or cell culture models are also invited to submit their data to ClinVar. Submissions that focus on functional evidence do not include an interpretation of the clinical significance of the variant. Instead they include a statement about the functional significance of the variant. As evidence for the functional significance, these submissions may include details about the experimental method, including any type of scoring system used, and the results of the experiment for each variant. ClinVar currently has 1,168 variants with functional significance of the variant; 459 variants include details of the experimental methods and results. This presentation emphasizes how clinical providers, patient registries, and researchers can share their phenotype and functional data, and how clinical testing laboratories and others can have integrated access to the data that is critical for classification of variants.

1493F

MGeND: A new Integrated database of clinical and genomic information.

M. Nakatsui, M. Kamada, E. Uchino, R. Kojima, Y. Okuno. Graduate School of Medicine, Kyoto University, Kyoto, Kyoto, Japan.

Recent advances in powerful technologies such as next-generation sequencing have provided an abundance of genomic information. Even in the clinical practice, genomic information of patients are generated and accumulated continuously. Genetic information, such as genotypes or genetic mutations, has a deep connection with disease onset and progress. For instance, cancer is a disease caused by a genetic mutation. Depending on what kind of mutation is occurring in cancer cells, it is possible to discriminate the type of cancer, and select an appropriate anticancer drug. One of the most important projects in clinical field is accumulating relation of genomic information and clinical patient information, that is clinical significance, in an organized form in various disease area from many medical institutions. Some databases that store genetic information with clinical significance, such as Clinvar developed by NIH in United States, have already been released. Also in Japan, disease area specialized database such as Database of Pathogenic Variants (DPV) specialized for intractable diseases are being developed. However, there are no database that reflect the characteristics of the Japanese and do not specialize in the disease area. In this study, we developed a disease-related genomic information database, Medical Genomics Japan Database (MGeND), reflecting the characteristics of the Japanese population by collecting genomic and clinical information from 11 of clinical groups located in Japan. Initially we collect data from 5 disease area, namely "cancer", "rare disease", "infectious disease", "dementia", and "hearing-loss". The development of MGeND realizes data-sharing of genomic and clinical information among clinical groups crossing disease area. The accumulation and share of genetic information with clinical significance will realize a more sophisticated precision medicine.

1494W

Open access 2030: Economics of data sharing in the genomic age. *N.*

Stockham, D. Wall. Stanford University, Stanford, CA.

Introduction The prospect of universal clinical adoption of whole genome sequencing (WGS) is moving swiftly into the realm of probable, sweeping society and science into a new genomic age. However, certain implementation details of this project have not been considered in depth. In particular, three essential problems face the scientific community. First, what are the requisite cost and medical utility conditions necessary to incentivize the universal adoption of WGS in the fields of oncology and newborn screening (NBS), and when will those conditions be met? Second, if widespread WGS is implemented, will the current architectures of data sharing between scientific and medical communities be economically sustainable? Finally, what data sharing architectures will scale economically? These questions are fundamental to the future of human genetics and definite answers are prerequisite for truly population scale genomic science. **Methods** We performed a cost effectiveness analysis (CEA) to establish the minimum additional medical utility of WGS necessary to justify use in oncology and NBS in the United States as a function of WGS price, parameterized by the likelihood of actionable findings from recent studies of NGS at Memorial Sloan Kettering for oncology and twin-studies for NBS. Furthermore, we constructed economic models for ad-hoc, centralized database, and novel compute-in-place (CIP) data sharing architectures, parameterized by number of scientific entities and total data shared. **Results** At \$100 for a whole genome at 30x depth WGS is cost effective for oncology if the expected increase in life exceeds one month, or if it increases probability of remission by more than 1%. For NBS WGS is cost effective if the expected increase in life exceeds two weeks. However, current approaches for data sharing are entirely infeasible. For adhoc and centralized database architectures it would cost on the order of 100 million dollars each year for 10 entities to share %1 of the total United States NBS and oncology data. But, with a CIP architecture, data sharing is free as computation is moved to the data. **Discussion** The \$100 WGS is a stated near-term goal of Illumina, implying WGS will be adopted widely in NBS and oncology before 2025. However, with current architectures this invaluable data will be too expensive to share. The scientific community must implement CIP architecture now, or leave the fruit of human heritage fallow in the field.

1495T

Characterizing individuals with pathogenic variants: An application of the phenotype risk score method. L. Bastarache, J. Bastarache, J. Moseley, J. Hughey, D. Roden, J. Denny. Vanderbilt University Medical Center, Department of Medicine, Nashville, TN.

Background Many patients who have both the genetics and symptoms consistent with a Mendelian disease remain undiagnosed. We test an automated approach to identifying these individuals using the electronic health record (EHR) and extant genetic data (Exome BeadChip). **Methods** We annotated pathogenic & benign alleles in ClinVar with OMIM disease identifiers. We generated star ratings for each variant-disease pair using ClinVar's submission rating algorithm. We restricted our analysis to recessive diseases. In a cohort of 35,842 genotyped de-identified individuals, we identified 1651 pathogenic and 751 benign variants. We calculated Phenotype Risk Scores (PheRS) for target diseases. PheRS is an automated method to score a individual based on their similarity to a disease's clinical profile using the EHR. We generated normalized z-scores (PheRS-z) to compare across diseases & calculated the average PheRS-z for different star ratings. We reviewed charts of individuals who had pathogenic variants with 2+ stars (n=292) to assess if they were diagnosed with or symptomatic for the disease suggested by their genotype. **Results** The average PheRS-z was elevated in homozygotes for pathogenic but not benign alleles (0.38 & 0.03, respectively) & was higher for alleles with more stars (i.e. stronger evidence of pathogenicity). Chart review revealed 50 of 292 homozygotes for recessive pathogenic alleles were clinically diagnosed (mean PheRS-z=2.0). The PheRS-z was elevated among the remaining 142 undiagnosed (mean PheRS-z=0.12). Eight undiagnosed individuals had outlying scores (PheRS-z > 2), with genetics and clinical features consistent with hemochromatosis (*HFE* p.C282Y), Bardet-Biedl syndrome (*BBS1* p.M390R), & Glycogen storage disease IV (*GBE1* c.691+2T>C), among others. Chart review of the undiagnosed individuals confirmed the feature overlap reflected in the elevated PheRS-z, and revealed further features suggestive of undiagnosed Mendelian disease. **Conclusion** PheRS effectively measured the presence of Mendelian disease features in conjunction with genetic data. Many homozygotes for pathogenic alleles did not have elevated PheRS or signs of disease on chart review, which may be due to low or age-related penetrance, genotyping error, or limited phenotypic data. Our work supports the notion that some Mendelian diseases may be under-diagnosed.

Pathogenic variants			
ClinVar rating (stars)	#Variants	#Homozygotes	avg PheRS-z
4	2	2	1.7
3	1	1	1.5
2	37	289	0.6
1	34	147	0.4
0	85	1212	0.3

1496F

Leveraging electronic health records to map the genetic and phenotypic complexity of preterm birth. A. Abraham¹, C. Bejan^{3,4}, J. Capra^{1,2}. 1) Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN, USA; 2) Department of Biological Sciences, Vanderbilt University, Nashville TN, USA; 3) Department of Biomedical Informatics, Vanderbilt University, Nashville TN, USA; 4) Vanderbilt Center for Precision Medicine, Vanderbilt University, Nashville TN, USA.

Background: Despite advances in neonatal medicine, preterm birth (PTB) causes significant morbidity and mortality across the world. In the United States, ~10% of births each year are preterm and up to 34% of infant deaths are attributable to prematurity. Premature babies are at high risk for both immediate and long-term complications. PTB heritability has been estimated to be as high as 40%. However, our understanding of the genetic underpinnings of PTB remains limited. A variety of different phenotypic factors – infection, placental pathologies, maternal comorbidities, etc. – can increase the risk of PTB. The extent to which these factors influence the genetic basis for PTB remains unknown. We hypothesize that failure to account for this phenotypic complexity has hampered previous genetic studies of PTB and that characterizing it will yield insights into PTB's genetic architecture. **Objective:** This project aims to classify pregnancies into preterm, term, and postterm through a phenotyping algorithm and investigate the shared genetic liability of PTB and comorbid diseases. **Methods:** Using counts of ICD and CPT codes, we identify preterm, term, and postterm deliveries within the electronic health record (EHR). Next, machine learning tools will extract features from the EHR to distinguish spontaneous from medically induced PTB and associated comorbidities. These algorithms will be trained and validated on non-overlapping samples that have been manually phenotyped. Leveraging Vanderbilt's biobank of genotype data linked with EHRs, we will construct genetic risk score models for subtypes of PTB and compare to a control cohort of term birth. **Results:** Delivery classified using one or more relevant ICD code resulted in 11,266, 7,140, and 10,037 individuals with at least one preterm, term, or postterm delivery respectively. A small fraction of individuals (2,314, 8.8%) had more than one type of delivery. The majority of individual (93.0%) within the preterm cohort has only one preterm delivery. To gain clinical insight into comorbidities, ICD codes were grouped into disease categories (PheCodes v1.2). Preterm deliveries showed substantial enrichment/depletion across 12 out of 17 categories. The top three were 'pregnancy complications' (OR = 1.35, p-value = 5.76x10⁻¹⁰⁰), 'genitourinary' (0.78, 2.18x10⁻³⁹), and traits with no disease grouping, (0.79, 4.59x10⁻³¹). We are now exploring the shared genetic architecture of these comorbidities with PTB.

1497W

Data fingerprints: Fast and simple comparison of semi-structured data, with emphasis on electronic health records. G. Glusman^{1,2}, M. Robinson¹, J. Hadlock¹, J. Yu², A. Khatamian³, J. Matelsky⁴, A.Y. Aravkin², E.W. Deutsch¹, N.D. Price¹, S. Huang¹. 1) Institute for Systems Biology, Seattle, WA; 2) The University of Washington eScience Institute, Seattle, WA; 3) Department of Computational Biology, St. Jude Children's Research Hospital, Memphis, TN; 4) Johns Hopkins University Applied Physics Laboratory, Laurel, MD.

Fast comparison of a large number of semi-structured data records can be valuable for analysis of any knowledge domain. Electronic health records (EHRs) present an important real-world test case because many healthcare questions rely upon finding the similarity between different records. Examples include finding candidates for clinical trials, evaluating treatment response, and investigating cost variation among patients with similar health situations. Currently, approaches to identifying similar records include (1) designing semantically-informed comparisons on structured elements of interest, such as demographics, conditions, or medications, and (2) using semantically and structurally agnostic file comparison. The first requires explicit mapping of domain knowledge to data structure, and the second becomes resource-intensive when scaled to millions of patient records. We present a fast, semantically and structurally agnostic method for analyzing semi-structured data (e.g., JSON or XML formats). Our locality-sensitive hashing strategy summarizes semi-structured data into 'data fingerprints': highly compressed representations which cannot recreate details in the data, yet simplify and greatly accelerate the comparison and clustering of semi-structured data by preserving similarity relationships. Computation on data fingerprints is fast: in one example involving complex simulated medical records (SyntheticMass, in FHIR format), the average time to encode one record was 0.53 seconds, and the average pairwise comparison time was 3.75 microseconds. Both processes are trivially parallelizable. Applications include detection of duplicates, clustering and classification of semi-structured data, which support larger goals including summarizing large and complex data sets, analyzing cohort structure, quality assessment, evaluating methods for generating simulated patient data, and data mining. Beyond EHRs, our approach is applicable to any domain in which semi-structured data are commonly used. While our novel methodology is currently designed for semi-structured data, it can be used to simplify and accelerate analyses of fully structured data, and (based on initial experiments) can be extended to work with fully unstructured data. Project page: http://db.systemsbiology.net/gestalt/data_fingerprints This work is supported by NIH 1U54EB020406 and NCATS award OT3TR002026.

1498T

Evaluating the potential of screening exome sequencing for newborn screening. A.N. Adhikari¹, Y.Q. Wang¹, R.J. Currier², R.C. Gallagher², R.L. Nussbaum³, Y. Zou¹, U. Sunderam⁴, J.T. Shieh², F. Chen², M. Kvale², S.D. Mooney⁵, R. Srinivasan⁴, B.A. Koenig², P.-Y. Kwok², J.M. Puck², S.E. Brenner¹, *The NBSseq Project*. 1) University of California, Berkeley, CA, USA; 2) University of California, San Francisco, CA, USA; 3) Invitae, San Francisco, CA, USA; 4) Tata Consultancy Services, Hyderabad, TS, India; 5) University of Washington, Seattle, WA, USA.

Exome screening shows Genetics aids public health But just sequence fails The NBSseq project sequenced and analyzed exomes of most newborns affected with any of the 48 inborn errors of metabolism (IEMs) screened by tandem mass spectrometry (MS/MS) amongst the 4.4 million newborns born in California from mid-2005 to 2013. The sequences were obtained under an IRB-approved protocol from 1216 de-identified dried blood spots archived by the California Dept. of Public Health. Using a 180 exome validation set, we developed an interpretation pipeline sensitive for all screened IEMs, yielding few false positives. Final analysis was restricted to 78 genes causing these IEMs, and incorporated curated pathogenic and benign variants, and variant predictions. The pipeline was then run once on the 1016 exome test set for evaluation. Sensitivity for predicting disease from variants varied across IEMs (eg, PKU:93%,3-MCC:71%,citruillinemia type1:77%) with an overall weighted sensitivity of 88% and specificity of 95%. Current newborn screening by MS/MS is 99.3% sensitive and 99.7% specific; therefore, exome sequencing alone cannot currently replace MS/MS screening for these disorders. A pipeline using only curated variants achieved 99.6% specificity and 58% sensitivity. For other disorders with no available screen, a curated variant pipeline could potentially identify a portion of the affected individuals, with few false positives. Of 103 affected individuals not predicted to have disease by our pipeline, 52 had no rare missense variants and 51 contained one heterozygous deleterious variant in a pertinent gene. In two isovaleric acidemia (IVA) patients, whole genome sequencing confidently revealed multi-exon deletions in the causative *IVD* gene, whereas exomes did not. Review of the cases prompted revision of the diagnosed disorder in 63 cases. An individual clinically diagnosed with IVA carried rare homozygous, missense variants (M389V) in *ACADSB*, known to cause 2-MBCD deficiency. Clinical review and demographic data supported this alternative diagnosis. Another individual had positive MCAD/VLCAD MS/MS screens and was clinically diagnosed as an unaffected carrier upon follow-up. The child carried a pair of missense variants (A84T/R175H) in *ETFDH*, suggesting an additional possible diagnosis of late-onset multiple acyl-CoA dehydrogenase deficiency. These results indicate that following a positive MS/MS screen, reflex DNA sequencing could aid precise and timely clinical resolution.

1499F

ExomeLoupe: A secure genetic variant viewer tool. C. Egas^{1,2}, B. Diéguez¹, F. Santos¹, M.J. Simões¹, H.J.C. Froufe¹. 1) Biocant - Biotechnology Innovation Center, Cantanhede, Portugal; 2) Center for Neuroscience and Cell Biology, Coimbra, Portugal.

One of the applications of NGS technologies is to sequence whole exomes and genomes to identify causal genetic variants in Mendelian diseases and provide useful information for clinical diagnosis. These technologies generate massive amounts of data that have to be processed into meaningful genetic information to be interpreted by clinicians and researchers. Compliance with the new European General Data Protection Regulation (GDPR), which considers human genetic information as personal and sensitive data, requires implementation of security measures to process, store and share genetic data. According to this regulation, measures such as pseudonymization and encryption should be implemented to manage security and mitigate risks for undesirable access to the information. There is a lack of desktop tools that represent a true end solution for visualization and prioritization of genetic variants that not only help researchers to identify the most relevant variants but also allow genetic data storage and sharing in a protected manner according to the GDPR. In this context, we present ExomeLoupe, an intuitive and user-friendly Windows platform to help with the prioritization and interpretation of genetic variants in a secure environment. ExomeLoupe requires as input a GEMINI database and offers the user the possibility to encrypt the database using SQLCipher according to the Advanced Encryption Standard 256 bit (AES256). Once encrypted, ExomeLoupe interacts directly with the database without the need for complete decryption, displaying variant annotation and searching for variants according to the filters defined by the user. Furthermore, ExomeLoupe can be used to encrypt files produced during data processing through PyCrypto, enabling users to securely store and share this sensitive genetic information complying with the GDPR. ExomeLoupe is the first software to our knowledge that provides the user the tools to work in an encrypted secure way. This work was funded by the In2Genome project, ref. CENTRO-01-0247-FEDER-017800, CENTRO 2020, Portugal 2020 and European Union, through the European Regional Development Fund (ERDF) and by GenomePT project, ref. POCI-01-0145-FEDER-022184, co-funded by FCT/MEC (PIDDAC) and ERDF through COMPETE – Operational Programme Competitiveness Factors (POFC).

1500W

Time-course expression QTL-atlas of 301 pregnant Japanese women in Maternity Log Study. M. Nagasaki^{1,2}, R. Yamashita¹, T.F. Shibata¹, J. Kawashima^{1,3}, Y. Tsunemoto^{1,2}, D. Ochi^{1,2}, T. Yamauchi^{1,2}, S. Kumatani^{1,2}, Y. Harada^{1,3}, T. Mimori^{1,3}, F. Katsuoka^{1,3}, N. Minegishi^{1,3}, O. Tanabe^{1,3}, S. Hiyama², J. Sugawara^{1,3}, Maternity Log Study Group. 1) Tohoku Medical Megabank Organization, Tohoku University, Tokyo, Japan; 2) NTT DOCOMO, INC., Yokosuka, Japan; 3) Tohoku University Graduate school of Medicine, Sendai Japan.

In Japan, about 200,000 people are affected each year by pregnancy-related diseases including hypertensive disorders of pregnancy, gestational diabetes, and premature delivery. Pregnancy-related diseases are multifactorial disorders caused by a complex interaction of genetic factors and environmental factors such as lifestyle and living environments. Maternity Log Study (MLOG) is an add-on cohort project on a prospective cohort study by Tohoku Medical Megabank Organization in Japan. MLOG is designed to elucidate the associations between genomic and environmental factors and pregnancy complications using continuous lifestyle monitoring combined with multi-omics data on the genome, transcriptome, metabolome, and microbiome. MLOG has completed recruitment of 301 individuals and constructed a whole-genome reference panel of them by using the HiSeq sequencer with an average depth of 25x. We also collected 902 peripheral blood samples from them during pregnancy and after delivery with PAXgene® RNA tubes, and performed RNA-seq using the HiSeq sequencer. By combining these datasets, we have conducted time-course expression QTL (eQTL) analysis as well as differentially expressed eQTL analysis to construct a time-course eQTL-atlas of 301 pregnant Japanese women. Here, we report novel variants related to gene expression during pregnancy, and annotate those eQTL variants using public database, e.g. GTEx, GWASCatalog, and our in-house results of genome-wide association studies for omics and lifelog data in MLOG. The time-course eQTL-atlas will contribute to global research communities for genetic studies on pregnancy.

1501T

Pipeline for trio analysis in StrandNGS. A. Choudhury, M. Toufiq, A. Pal, S. Ramachandrala. Strand Life Sciences Private Limited, Bengaluru, India.

Whole genome and exome sequencing is now commonly employed as a clinical diagnostic test to identify disease-causing mutations in the affected children of seemingly unaffected parents. While the technology of sequencing itself has evolved to become simplistic and cheaper, analyses and interpretation of this data remains the bottle-neck causing delays in reporting thereby delaying the process of improved decision making. We designed a simplistic pipeline that helps in faster detection of the disease causing variants with an accuracy that is on par with the GATK pipelines. The Trio analysis pipeline in StrandNGS encompasses raw data import, pre-alignment QC, alignment, filtering reads (based on mapping quality and alignment score) and duplicate reads, SNP pre-processing (Local re-alignment and Base Quality Score Recalibration), detection of SNPs (including low frequency SNPs), and annotating the variants. For narrowing down causative mutations in the proband with unaffected parents, the software could identify SNPs that belong to either of these 3 categories: (1) homozygous recessive SNP (2) de novo SNP (3) compound heterozygote. Given that the StrandNGS platform has a robust and systematic framework for SNP prioritization, identification of the causative mutation in case of any genetic disorder is facilitated through features like (a) Find significant SNPs, to compare similar cases and identify any SNPs that recur across cases (b) SNP effect analysis, to find all the protein and non-protein effects of a given SNP and annotate it with HGVS nomenclature at genomic, transcriptomic, and proteomic levels (c) Find damaging non-synonymous (NS) variants, to identify NS variants that are damaging using popular functional prediction algorithms (d) Check for known variants, to scan the SNP list for variants that are known to cause the suspected disorder (e) Annotate SNP list to add any known metadata about a given SNP. All the above mentioned steps can be configured into pipelines with parameters of choice and run in an automated fashion for quick TAT (Turnaround Time). We intend to showcase this pipeline using data of a Wilms tumor case from 1000 Genomes Project. While there are a number of open-source pipelines that can be used to perform one or more of the above mentioned steps, to the best of our knowledge, StrandNGS is the only desktop software that provides such a comprehensive workflow that can be configured into a pipeline, all on a single platform.

1502F

Alvar: An artificial intelligent variant classifier for clinical interpretation of genetic variants and prioritization of variants of unknown significance (VUS). J. Luo, T. Zhou, Y. Zi, X. Li, Y. Wu, Y. Deng, A. Li, Y. Shi. TopGene Tech Co., Ltd., Guangzhou, Guangdong Province, P.R.China.

Clinical interpretation of sequence variants has been a difficult task in genetic medicine. Although the updated guidelines published by American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG-AMP) in 2015 provides a gold standard for germline variant interpretations, some problems remain unresolved. For example, variants of unknown significance (VUS) are commonly seen in classification which are not referable for any clinical intervention. They are classified as VUS due to the lack of conclusive evidence. Prioritizing VUS with high probability to fall in the benign or pathogenic category is the basis for further research in these variants, which can eventually result in their reclassification. Therefore, high-throughput methods for VUS pathogenicity estimation are in urgent need. To address this, we developed Alvar (Artificial-Intelligent variant classifier), a neural network model which is capable of classifying variants into either pathogenic/likely pathogenic (P/LP) or benign/likely benign (B/LB) category. Alvar's prediction capacity was verified on known data sets classified by human experts. It also showed superior performance over 7 other popular variant classification tools. More importantly, Alvar demonstrated great potential in VUS reclassification by correctly predicting 43 out of 44 former VUS. We then applied Alvar on 15,670 VUS (classified by the Clinvar expert panel). 79% of these variants were prioritized towards either P/LP or B/LB category by Alvar with at least 95% confidence. Alvar is available as an open-source program at Github <https://github.com/OpenBioWiki/Alvar>. An online version is available at <http://www.bioopenwiki.com/Alvar>. Alvar demonstrates that neural network is suitable for sequence variant classification and a promising solution to the VUS problem.

1503W

A fast and interpretable new approach to fine mapping of genetic associations. G. Wang¹, A. Sarkar¹, P. Carbonetto¹, M. Stephens^{1,2}. 1) Department of Human Genetics, University of Chicago, Chicago, IL; 2) Department of Statistics, University of Chicago, Chicago, IL.

Despite near two decades of work in the area, fine mapping of genetic associations continues to pose major statistical and computational challenges, even when all potential causal variants are typed. Several tools have been developed, many based on Bayesian multiple regression (BMR). BMR has several appealing features compared with other approaches -- for example, it can assess uncertainty about which genetic variants explain an association, taking account of Linkage Disequilibrium (LD). And it can incorporate annotation information on variants through a modified prior distribution. However, BMR is computationally challenging. Some methods enumerate all possible models involving a small number of causal variants, but this quickly becomes computationally infeasible. Approaches such as MCMC and stochastic search can help, but remain computationally intensive. And it remains a major challenge to summarize and interpret their output, especially with high LD and many causal variants. Here we propose a new approach to fine mapping that is fast and gives new and more interpretable results. Our approach involves several innovations. First we introduce a new version of BMR, which we call the *Sum of Single Effects* (SuSiE) model. This model, while similar to existing widely-used models, has a different structure that makes it possible to fit via a new variational algorithm whose computation scales *linearly* with the number of causal variants. In contrast to existing variational algorithms for BMR, which are unsuited to fine mapping because of the way they deal with SNPs in high LD, our new version is ideally suited to this difficult case. The SuSiE structure also yields, by design, independent confidence sets, each designed to capture *one association signal*, making the results easy to interpret, and ideal for guiding follow-up studies. Although our initial work focuses on the simplest case, the framework is easily extended in many ways, including joint fine mapping of multiple phenotypes, incorporating functional annotations, and inference from summary data of association studies. We believe that the simplicity and power of this new approach has the potential to transform ongoing fine mapping efforts. This talk introduces SuSiE, illustrates its benefits compared with existing methods, and show several examples of fine mapping on real data.

1504T

Evaluation of variant scoring methods in non-coding region in the context of complex traits/diseases. G. Zheng, R. Chen, H. Yang, B. Li. Vanderbilt University, Nashville, TN. 1161 21st Ave S D3300, Nashville, TN 37232, 5th floor.

Introduction: An attractive alternative to Whole-exome sequencing, Whole-genome sequencing (WGS) identifies noncoding variants and mutations associated with complex human traits and diseases. However WGS introduces the challenge of discriminating functionally relevant variants from a huge number of background variants. To ameliorate this issue several scoring methods utilizing functional genomic data were created to prioritize noncoding variants. However, their performance in terms of predicting the potential as risk variants in complex diseases remains unclear. **Methods:** In this study, we aim to assess the performance of various scoring methods including DVAR, Linsight, Eigan, and CADD, on prioritizing risk variants. Specifically we used LD score regression to estimate heritability of top scoring variants and applied to GWAS summary statistics of Schizophrenia, Alzheimers, and Breast Cancer. Summary statistics were obtained through PGC, Breast Cancer Consortium, and International Genomic of Alzheimer Project. As a proof of concept we utilized DVAR as an example and extracted top scoring variants by making a list of 2 different cutoffs of 0.95 and 0.99 representing the top 95 and 99% of predicted noncoding variants on genome. These lists would then be used in ldsc to calculate the disease heritability explained by these variants and calculate their enrichment compared to background. **Results:** Preliminary results show that top scoring variants explain significantly more heritability than genome background ($P < 0.001$). When comparing the two DVAR cutoffs, the top 99% of noncoding variants explain significantly more heritability than the top 95% score variants in schizophrenia (0.0001 vs 0.005). We observed similar trends in breast cancer and Alzheimer GWAS summary statistics. Further investigation will be done for other scoring methods such as CADD, Eigan, and Linsight to answer questions such as whether all scores perform well in all diseases, or they show some specific patterns in different disease or if certain scoring methods perform better for specific category of variants. Ldsc can be found at <https://github.com/bulik/ldsc>.

1505F

How well can we create phased, diploid, human genomes? An assessment of FALCON-Unzip phasing using a human trio. *A. Fungtammasan, B. Hannigan.* DNAnexus, Mountain view, CA.

Long read sequencing technology has allowed researchers to create de novo assemblies with impressive continuity[1,2]. This advancement has dramatically increased the number of reference genomes available and hints at the possibility of a future where personal genomes are assembled rather than resequenced. In 2016 Pacific Biosciences released the FALCON-Unzip framework, which can provide long, phased haplotype contigs from de novo assemblies. This phased genome algorithm enhances the accuracy of highly heterozygous organisms and allows researchers to explore questions that require haplotype information such as allele-specific expression and regulation. However, validation of this technique has been limited to small genomes or inbred individuals[3]. As a roadmap to personal genome assembly and phasing, we assess the phasing accuracy of FALCON-Unzip in humans using publicly available data for the Ashkenazi trio from the Genome in a Bottle Consortium[4]. To assess the accuracy of the Unzip algorithm, we assembled the son's genome using FALCON and FALCON Unzip, genotyped publicly available short read data for the mother and the father, and observed the inheritance pattern of the parents' SNPs along the son's phased genome. We found that 72.8% of haplotype contigs share SNPs with only one parent suggesting that these contigs are correctly phased. Most mis-phased SNPs are random but present in high frequency toward the end of haplotype contigs. Approximately, 20.7% of mis-phased haplotype contigs contain clusters of mis-phased SNPs, suggesting that haplotypes were mis-joined by FALCON-Unzip. Mis-joined boundaries in those contigs are located in areas of low SNP density. This research demonstrates that the FALCON-Unzip algorithm can be used to create long and accurate haplotypes for humans and identifies problematic regions that could benefit in future improvement.

1506W

Improved dog reference assembly enables paralog-specific identification of tandem repeats in assembled gaps using QuickK-mer. *F. Shen¹, A. Pendleton¹, A. Boyko², J. Kidd¹.* 1) University of Michigan, Ann Arbor, MI; 2) Cornell University, Ithaca, NY.

Dogs are an emerging model system for the study of genome evolution and human disease, with over 300 heritable traits similar in humans. Ongoing studies using short-read re-sequencing data often rely on an incomplete and fragmented reference assembly. The latest version of the dog genome reference, CanFam3.1, contains 19,553 gaps on primary chromosomes as well as 3,000 unplaced contigs with a combined length of 83Mbp. To complement this assembly, we recently completed a de novo assembly of a Great Dane dog based on 50x PacBio long-read sequence data. The assembly results demonstrate a vast improvement with 20-fold increase in continuity and a drastically reduced number of gaps and unplaced contigs. We determine the sequence of most gaps in the CanFam3.1 assembly identify examples of partial or full transcript models which were previously missing. Assembly comparison identified 2595 insertions > 1kb in size totally missing from the canFam3.1 assembly. Analysis shows that these new sequences are enriched for tandem repeats. In order to efficiently genotype novel sequences identified in our assembly, we assessed copy-number in hundreds of sequenced canines using our paralog-specific CNV pipeline, QuickK-mer. We have improved QuickK-mer by eliminating a major portion of control k-mers, while still retaining sufficient GC correction accuracy. Our changes significantly improved its speed and reduced the memory usage. Our results suggest these novel regions are polymorphic with respect to presence and are a part of canine evolution process.

1507T

Caring without sharing: Genome-wide association and mapping on cohorts fragmented across institutional silos. A. Pourshafeie¹, C.D. Bustamante², S. Prabhu². 1) Department of Physics, Stanford University, Stanford, CA; 2) Department of Biomedical Data Science, Stanford University, Stanford, CA.

The recent explosion in genomic and phenotypic data generated by new biomolecular and biometric technologies presents a great opportunity for expanding our understanding of human genetics and disease. Moreover, the available data is becoming more expansive and private in nature. This expansion, in cohort size and per-individual information can be expensive, unsafe and time consuming to generate or maintain at a single centralized facility. Meta-analysis/summary statistics techniques offer a work-around by combining summary statistics across multiple studies. Despite their overall success, these methods face a few serious limitations: (a) the shared summary statistics might be inadequate for some types of inference (e.g. estimating effects from shared two-tailed p-values), (b) subtle differences in model, assumptions and QC can introduce bias (Greco et al. 2013), (c) in each silo, the parameter estimates can be unreliable if the per-silo sample size is small compared to the number of covariates (k). (d) fine tuned population structure control is not possible. Here, we implement a logistic regression based, GWAS pipeline that allows participant data to remain fragmented across institutional silos. We show that Alternating Direction Method of Multipliers (ADMM) (Boyd et al. 2010) can be used to perform generalized-regression in a decentralized fashion. This method generalizes the intuition of averaging the per-silo estimates and iteratively solves the problem while only communicating summary statistics at each iteration. To evaluate this approach in a GWAS pipeline, we simulate phenotypes on POPRES dataset (Nelson, et al. 2008). We then perform a centralized and decentralized GWAS. Specifically, we show that centralized QC performed using Plink (Purcell, Chang 2010) is indistinguishable from decentralized QC. We then perform PCA and show that the results are both mathematically and numerically identical in both paradigm. Lastly, we perform GWAS via logistic regression correcting for the first 5 PC's and show that both effect sizes and the associated p-values are essentially identical between the two paradigms. Taken together, this shows that a GWAS pipeline can be performed in a decentralized fashion.

1508F

Analyzing phenotype imputation effectiveness on summary statistics from the UK Biobank. J. Rotman, E. Eskin. Computer Science Dept, UCLA, Los Angeles, CA.

Genome-wide association studies (GWAS) have been extremely effective in identifying loci associated with complex phenotypes. However, many phenotypes can be difficult to measure, or too expensive to obtain in large quantities. Since the power of a GWAS depends on sample size and accurate phenotype measurement, for many phenotypes performing a GWAS is not practical. One potential approach to this problem is through the use of phenotype imputation. The idea behind this approach is that a subset of the data that contains many phenotypes, including the phenotype of interest, is used to train a model that predicts the phenotype of interest using the other phenotypes. The model is then used to predict or impute the phenotype of interest in data which collects the other phenotypes but not the phenotype of interest. Specifically, the model uses correlation between the pairs of phenotypes and these correlations are calculated in the smaller subset of individuals. A GWAS can then be performed on the imputed phenotype. Phenotype imputation works because of pleiotropy, the phenomenon that a variant can affect more than one trait and the effect sizes of the variant on the pair of traits is correlated. In phenotype imputation, the correlation of many traits are combined with each other which results in an accurate enough prediction of the trait that a GWAS association can be discovered. While phenotype imputation has shown promising results, it is yet to be used on a large enough scale with just summary statistics. This further complicates the problem because summary statistics increase the difficulty of obtaining phenotypic correlation. By utilizing LD score regression to compute phenotypic correlation, phenotype imputation can produce an accurate estimation of SNP association for unknown phenotypes using summary statistics from other correlated phenotypes. In this work we test this method using summary statistic data from the UK Biobank. We demonstrate how our method can be used to obtain significant SNPs at a comparable level to a measured phenotype. Additionally, we show how the impact of these significant SNPs extends beyond the SNPs highly associated with correlated phenotypes.

1509W

Short tandem repeat (STR) profile of pentanucleotide repeats and high-resolution STR typing solution for the Japanese population. *M. Saito¹, S. Hirata¹, K. Kojima¹, K. Misawa¹, T. Mimori¹, O. Gervais¹, Y. Kawai², Y. Hitomi³, K. Tokunaga², M. Nakamura^{3,4}, M. Nagasaki¹.* 1) Division of Biomedical Information Analysis, Department of Integrative Genomics, Tohoku Medical Megabank Organization, Tohoku University, Japan; 2) Department of Human Genetics, Graduate School of Medicine, The University of Tokyo, Japan; 3) Clinical Research Center, National Hospital Organization, Nagasaki Medical Center, Japan; 4) Department of Hepatology, Graduate School of Biomedical Sciences, Nagasaki University, Japan.

Objective: In the field of forensic, DNA typing is one of the main methods for identifying individuals. Microsatellite short tandem repeat (STR) analysis is widely used, for identifying victims in large-scale disasters as well as missing persons in criminal investigations. Since the number of repetitions of an STR is on average more diverse than a single-nucleotide polymorphism among individuals, STR typing can distinguish individuals with fewer genotyping regions. Furthermore, the use of highly polymorphic STRs in regional populations, such as the D6S1043 locus in Asian populations, has enabled higher precision in individual identification. **Result:** In this study, we used the whole-genome data from 1,070 Japanese individuals, generated using massively parallel short-read sequencing of 162 paired-end bases in Tohoku Medical Megabank Organization. For this dataset, we analyzed 106,418 pentanucleotide STR loci in human assembly GRCh37 and detected 8,771 highly polymorphic STR loci (with which at least five types of repeat numbers were observed in the 1,070 individuals). From them, we selected 53 pentanucleotide STR loci suitable both for short-read sequencing and PCR-based technologies, which would be candidates to the actual forensic DNA typing in the Japanese population. For these loci, we designed PCR primers and evaluated performance by measuring the repeat length of each individual with the electrophoresis experimental results. For the performance evaluation, we used the whole-genome sequence data of another 181 Japanese specimens owned by the National Hospital Organization Nagasaki Medical Center and calculated the STR repeat length. Consequently, we have established a reliable pentanucleotide repeat typing method for the Japanese population.

1510T

The Encyclopedia of DNA elements (ENCODE): 2018 data portal update. *E.T. Chan, ENCODE Data Coordination Center.* Stanford University, Palo Alto, CA.

The Encyclopedia of DNA Elements (ENCODE) project aims to identify functional elements in the human genome. The ENCODE consortium, an international collaboration of experimental and computational research groups, has produced over 15,000 high-throughput experimental data sets since 2003. These data include transcriptome and epigenome profiles in multiple human and mouse tissues and cell lines. A data portal (<https://www.encodeproject.org>), developed by the ENCODE Data Coordinating Center (DCC), provides integrated, free access to over 600 terabytes of raw experimental, processed, and analysis data files with corresponding metadata. The portal also provides access to details of experimental protocols, computational procedures, and validations of key reagents (e.g. antibodies and genetically-modified cell lines). Additionally, the portal also provides access to data from other related projects such as Model organism ENCODE (modENCODE), Model organism Encyclopedia of Regulatory Networks (modERN), Genomics of Gene Regulation (GGR), and Roadmap Epigenome Project (Roadmap). Several new features are currently being added to the ENCODE data portal, such as a "Shopping Cart" function on the web user interface to facilitate customized data selection and download. On the back end, the search code has been refactored and the version of Elasticsearch updated to improve search efficiency and versatility. The ENCODE portal server architecture has also been updated and decentralized, improving stability and responsiveness of the application. New datasets are also continually being added as ENCODE labs submit new data and metadata meeting ENCODE standards. Of note, new to the portal will be data aimed at characterizing putative functions of identified ENCODE DNA elements using various genetic modification and perturbation methods (e.g. CRISPR-based screening). Up-to-date details on these developing features will be presented at the time of the meeting.

1511F

GFAKluge: A toolbox for manipulating the graphical fragment assembly format. *E.T. Dawson*^{1,2}, *R. Durbin*^{1,3}. 1) Department of Genetics, University of Cambridge, Cambridge, United Kingdom; 2) Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD, USA; 3) Wellcome Trust Sanger Institute, Hinxton, UK.

Graph models of genomes have become increasingly common in recent years, especially in the field of genome assembly. While FASTA has remained the dominant assembly format for several decades, it does not permit encoding any information about the assembly graph or population variation. The Graphical Fragment / Assembly format (GFA) has been proposed as a community standard for representing assembly graphs, sequence graphs, and variation graphs. To encourage uptake of the format we have developed a set of tools that replicate many of the analyses commonly performed on FASTA, such as calculating genome statistics. The toolkit contains tools for sorting and conversion that facilitate interchange between programs for assembly, read mapping and visualization. We also include a modern C++ API to encourage adoption among developers of high-performance tools. Our code is freely available at <https://github.com/edawson/gfakluge>.

1512W

Towards reproducible, transparent, and systematic benchmarking of omics computational tools. *S. Mangul*¹, *L. Martini*¹, *B. Hill*¹, *A. Lam*¹, *M. Distler*¹, *A. Zelikovsky*², *E. Eskin*¹, *J. Flint*¹. 1) UCLA, Los Angeles, CA; 2) Georgia State University, GA.

Rapid technological advances, such as the development of next-generation sequencing, are driving the need for comprehensive computational tools to analyze the wealth of generated genomic data. Systematic benchmarking has been successful in many research disciplines to help non-computational researchers evaluate the accuracy and applicability of new computational tools. Adopting a standardized benchmarking practice and following established principles for the design of new benchmarking studies could help researchers using omics data to better leverage technological innovation. Here, we summarize established principles for guiding the design of new benchmarking studies. We separately discuss challenges and limitations of benchmarking studies, highlighting domains of computational biology where it is impossible to obtain the accurate gold standard making the benchmarking impossible or extremely limited. We reviewed of 26 benchmarking efforts published across 10 relevant peer-reviewed journals from 2011 to 2017 reveals several aspects of software development unique to computational biology. Our survey of benchmarking tools revealed a large variety of techniques currently used to prepare gold standard data. According to our survey, the most popular technique (10 out of 26 surveyed benchmarking studies) for preparing gold standard data is to use various alternative technologies. Benchmarking studies performed by one research team evaluated on average 18 computational tools, competition-based studies evaluated on average 37 tools. All the data generated by a benchmarking study offer substantial value to the software development and research community — yet these data are often not shared (58% of studies failed to share benchmarking data). We designated a separate discussion of the ephemeral nature of benchmarking studies, in particular, addressing the challenge of asynchronous version changes of the benchmarked method, new tool development and obsoleting the benchmarking study results. We also discuss different strategies to optimize benchmarking, including crowdsourcing and challenge-based benchmarking. We identify and discuss aspects of challenge-based benchmarking relevant to tests performed by individual research groups. Finally, we propose step-by-step instructions for increasing reusability, transparency, and reproducibility of benchmarking studies by using containerization, common data representation, open data, and systematic parameter description..

1513T

Leveraging patterns of purifying selection to identify clinically important non-coding sequence in human whole-genome sequence data. *D. Vitsios, A. Platt, S. Petrovski.* AstraZeneca Centre for Genomics Research, Precision Medicine and Genomics, IMED Biotech Unit, AstraZeneca, 1 Francis Crick Avenue, Cambridge Biomedical Campus, Cambridge CB2 0AA, UK.

Elucidating functionality in human non-coding sequence is a key genomics challenge. The Centre for Genomics Research (CGR), AstraZeneca is developing tools to exploit the full value of whole-genome sequencing (WGS) data to facilitate identifying novel drug targets in the non-coding genome. We previously showed that the intolerance to variation of coding and proximal non-coding sequence is a strong predictor of human disease relevance (Petrovski et al., 2013, 2015). Here, we expand the published method by applying a tiled genome-wide Residual Variation Intolerance Score (gwRVIS) to WGS data from 15,496 individuals (Lek et al., 2016) with lower gwRVIS corresponding to greater intolerance. We verified that gwRVIS distinguishes human disease genes (OMIM-haploinsufficient; median gwRVIS=-0.51) from other protein-coding genes (median gwRVIS=-0.42), both significantly departing from intergenic sequence (median gwRVIS=-0.02; $p < 1 \times 10^{-200}$). Strikingly, gwRVIS identified ultra-conserved non-coding elements as the most intolerant category (median gwRVIS=-0.62; $p < 1 \times 10^{-320}$). ClinVar non-coding pathogenic regions (median gwRVIS=-0.46; $p = 2.6 \times 10^{-89}$) and ubiquitous enhancers (median gwRVIS=-0.47; $p = 3.3 \times 10^{-120}$) were the next most significantly intolerant distributions compared to intergenic sequence. Remarkably, the 20% most intolerant enhancers were significantly enriched for ubiquitous enhancers compared to the 20% most tolerant (2.2-fold increase, $p = 2 \times 10^{-9}$). Our score captures previously inaccessible information to help prioritise genetic variants found in the human non-coding regulatory sequence of known and novel drug targets and biomarkers.

1514F

ncER (non-coding Essential Regulation) score – prioritization of variants for clinical diagnosis in the non-coding genome. *J. di Iulio^{1,3}, A. Wells^{2,3}, P. Mohammad^{1,4}, D. Heckerman³, T. Lappalainen⁴, A. Telenti¹.* 1) The Scripps Research Institute, La Jolla, CA; 2) Stanford University, Stanford, CA; 3) Human Longevity, Inc., San Diego, CA; 4) New York genome Center, New York, NY.

Understanding the consequences of variation in the non-coding genome is crucial. We use machine learning techniques to model non-coding genome biochemical, genetic, 3D chromatin structure, gene regulatory constraint and high-throughput enhancer reporter screen data. We trained a XGBoost model, an implementation of gradient boosted decision trees, to differentiate between pathogenic and control genomic positions using 73 features, including (i) existing non-coding deleteriousness scores (e.g.: CDTS, ncEigen, FATHMM, FunSeq2, GERP, LINSIGHT and ReMM), (ii) genome essentiality (e.g.: pLI, and gene regulatory constraint from ANEVA), (iii) biochemically detected genomic interactions (e.g.: 3D chromatin structure data and histone marks) and (iv) regulatory screens (obtained with STARR-seq). The model was trained on 1225 non-coding pathogenic variants from ClinVar and HGMD, and 8093 control variants. Controls were variants present in the gnomAD dataset at an allelic frequency > 1% and matched for the distance to nearest splice site and genomic elements. 80/10/10 percent of the variants were used respectively as training/dev/test sets. The final model achieved 83% ROC-AUC and 64% PR curve and outperforms any single deleteriousness metrics by at least 9% ROC-AUC and 12% in PR curve. The top 20 most important features in the model included nine deleteriousness metrics, four gene essentiality features, two expression features, three chromatin marks feature, one 3D chromatin structure feature and one regulatory screen feature. Interestingly, although the deleteriousness metrics showed the highest cumulative feature importance, the contribution of the additional features is considerable. Indeed, their incorporation in the model improved the ROC-AUC by 3% (from 80 to 83%) and the PR curve by 8% (from 58 to 64%), indicating the complementary benefit of each additional data source. The use of new information sources on genome structure, human constraint and functional essentiality increases the accuracy of predicting the impact of genetic variations at nucleotide resolution. This new metric – ncER score – can guide functional experimentation and prioritization of variants for clinical diagnosis in the non-coding genome.

1515W

Application of a machine learning framework to prioritise genes from large-scale cohort analyses. *S. Petrovski, D. Vitsios.* AstraZeneca Centre for Genomics Research, Precision Medicine and Genomics, IMED Biotech Unit, AstraZeneca, 1 Francis Crick Avenue, Cambridge Biomedical Campus, Cambridge CB2 0AA, UK.

The tools to identify causal functional variants in known disease genes have greatly improved in recent years through the development of novel scores (e.g. Kircher et al, 2014) and novel machine learning frameworks (e.g. Traynelis et al, 2017). Present day, the genomics community, including our team at the Centre for Genomics Research (CGR), AstraZeneca is generating and analysing large volumes of exome and genome sequence data to better understand the genetic architecture of rare and common complex disorders. Here, we present a multi-dimensional machine learning framework aiming to identify novel disease genes among rare-variant collapsing analyses of all protein-coding genes that rank highly in gene-based analyses but do not achieve experiment-wide significance. We have applied this framework to a number of our cohort analyses harvesting diverse types of information per gene, including cohort-level association statistic, gene expression, human disease literature, mouse phenotypes, proteomic, interactome, genic metrics of human-lineage purifying selection (e.g. Residual Variation Intolerance Score-RVIS and Missense Tolerance Ratio-MTR) and others (52 features in total). Using a collection of already known disease-specific genes, and comparing to the remainder of the currently non-trait specific (unknown) genes we train a Random Forest classifier (R Caret package; $n=500$, $mtry=8$) and focus on unknown genes that based on shared properties with known genes are being classified as potentially novel 'trait-associated genes'. Across different parametrisations selected after model tuning, we often converge on a priority set of < 10 potentially novel trait-associated target genes. We also assess feature importance using the Boruta algorithm to allow for a permutation-based independence testing. Certain features like disease literature annotations, mouse trait-specific/lethal phenotypes, GO annotation and metrics of genic purifying selection like gnomAD-based MTR and RVIS frequently emerge as key features. Our multi-dimensional gene prioritisation framework accepts results from any gene-based burden or collapsing analysis and is agnostic to the disease cohort, provided there is already sufficient knowledge of disease-specific genes. We propose this approach as a support framework for objectively and quantitatively triaging potential novel disease target genes that are well ranked in rare variant collapsing or burden analyses.

1516T

Integrated DNase I hypersensitivity information to supervise cell type correction in DNA methylation analysis. *X. Fu, HW. Deng.* Global Biostatistics and Data Science Department, Tulane University, New Orleans, LA.

In epigenome-wide association studies, to investigate the associations between DNA methylation and phenotypes through tissue samples, cell components could be a potential significant confounding due to the nature that each cell has its distinct methylation pattern. To adjust for this heterogeneity, two major types of algorithms have been developed, including reference-based and reference-free methods. While reference-based methods are now popular and well adapted in current studies, these methods are requiring cell individual methylation profiles are available, thus mostly limited within studies involving whole blood. Reference-free, on the other hand, offers more flexibility to deal with complex tissues, for example, most cases of cancers. However, the performance of existed reference-free methods is still arguable. DNase I hypersensitive site (DHS) is the accessible DNA regions that are sensitive to cleavage by Dnase I enzyme, which feasible the transcriptional activity of the genes. Large percentage of distinct DHSs are specific to each cell type, and most importantly, more stable in each cell type. With DHS profiles largely available through epigenomic projects, this sparse feature is highly proposed to use in methylation prediction studies. In this study, I propose to integrate DHS signals to the first DHS guided Nonnegative Matrix Factorization (NMF) in methylation studies, thus to better identify the underlying variations of cell mixture. Through simulations I demonstrate this method could outperform existed methods in several scenarios in regards of specificity and sensitivity. While applying this method in real data is still underway, it's promising to develop an advantage solution for cell-mixture heterogeneity in methylation studies by integrating other DNA regulation information into reference-free algorithms.

1517F

MuPhenome: A curated catalog of microbiome correlates with clinical phenotypes. S.K. Belman¹, O.K. Mathew², M.R. Walther-Antonio³, P. Jeraldo³, F. Abdu¹, R.D. Vuniki^{1,4}, M.A. Badgeley⁵, K.W. Johnson⁵, B.S. Glucksberg⁶, M. Iyer⁸, J.S. Hirsch¹⁰, A. Lee⁸, R. Saha⁷, L. Subramanian⁴, K. Bock¹⁰, M. Oppenheim¹¹, R. Sowdhamini⁸, I.J. Kullo⁹, P.K. Gregersen⁴, J. Chelico¹, J.T. Dudley⁶, N. Chia³, K. Shameer¹⁰. 1) Center for Research Informatics and Innovation, Northwell Health, New Hyde Park, NY, USA; 2) B3, JKR, Aluva, Kerala, India; 3) Microbiome Program of Centre for Individualized Medicine; Division of Surgery Mayo Clinic, Rochester, MN, USA; 4) Courant Institute of Mathematical Sciences, New York University, New York, NY, USA; 5) Institute for Next Generation Healthcare, Mount Sinai Health System, New York, NY, USA; 6) Feinstein Institute of Medical Research, Northwell Health, Manhasset, NY, USA; 7) Department of Chemical and Biomolecular Engineering, University of Nebraska-Lincoln, Lincoln, Nebraska, USA; 8) National Center for Biological Sciences, Bangalore, India; 9) Division of Cardiovascular Diseases, Mayo Clinic, Rochester, MN, USA; 10) Department of Information Services, Northwell Health, New Hyde Park, NY, USA; 11) Division of Infectious Diseases, Zucker School of Medicine at Hofstra/Northwell, Hempstead NY, USA.

Metagenomic sequencing of clinical samples have expanded our understanding of disease phenotypes and the dynamic nature of microbial flora in different body cavities. Emerging studies have shown that several microbial taxa or species were altered in the setting of immunological, cardiovascular, and gastrointestinal diseases among others. Microbial taxa identified in the setting of clinical phenotypes offer unique challenges to understand the specific functional pathways and metabolic reactions mediated by host-microbiome interactions. The collective study of microbial dysbiosis and corresponding clinical phenotypes by compiling datasets from multiple studies will be an important step toward understanding the role of microbes in health and disease. Further, it is important to perform a systematic exploration of microbial species within taxa to assess various parameters including predictive mapping from taxa to species level, co-metabolism mediated by host-microbe interactions, computational flux balance analyses, protein domain level species interaction, and prioritization using interolog evidences. To achieve this goal, we are crowdsourcing a catalog of clinical microbiome studies with details including disease phenotype or clinical trait, microbial taxa and species implicated in the phenotype and various experimental details. We envisage that the availability of an open access microbiome-phenome catalog, with its curated content, crowdsourcing features, bioinformatics utilities and information related to microbiome and clinical phenotypes, will be a useful resources for the research community interested in exploring clinical utilities of microbiome datasets. MuPhenome would also be vital for developing pharmacomicrobiome studies and microbiome-driven drug development and drug repositioning investigations. The catalog is available from the URL: <http://www.muphenome.org>.

1518W

Testing hypotheses about the microbiome using an ordination-based linear decomposition model. Y. Hu¹, G. Satten². 1) Emory University, Atlanta, GA, 30322, USA; 2) Centers for Disease Control and Prevention, Atlanta, GA, 30333, USA;.

Background: Data from a microbiome study is often analyzed by calculating a matrix of pair-wise distances between observations; ordination using principal components or multi-dimensional scaling of the distance matrix is frequently successful in separating meaningful groups of observations (e.g., cases and controls). Although distance-based analyses like PERMANOVA can be used to test association using only the distance matrix, the connection between data on individual species (or operational taxonomic units, OTUs) and the information in the distance matrix is lost. This makes it hard to know which species contribute to the patterns seen in ordination for the high-dimensional data we gather in a microbiome study. **Methods:** We introduce a novel approach called the linear decomposition model (LDM) that provides a single analysis path that includes distance-based ordination, global tests of any effect of the microbiome, and tests of the effects of individual OTUs with false discovery rate (FDR)-based correction for multiple testing. The LDM accommodates both continuous and discrete variables (e.g., clinical outcomes, environmental factors) as well as interaction terms to be tested either singly or in combination, allows for adjustment of confounding covariates, and uses permutation-based p-values that can control for correlation (e.g., repeated measurements on the same individual). The LDM can also be applied to transformed data, and an 'omnibus' test can easily combine results from analyses conducted on different transformation scales. **Results:** For global testing, our simulations indicate the LDM provides correct type I error, even with substantial confounding and/or correlations, and can have substantially higher power than existing distance-based methods. For testing individual OTUs, our simulations indicate the LDM generally controlled the FDR well. In contrast, DESeq2 had inflated FDR in the presence of confounding and inferior sensitivity with sparse OTU data; MetagenomeSeq generally had the lowest sensitivity. The exibility of the LDM for a variety of microbiome studies is illustrated by the analysis of data from two microbiome studies. **Conclusions:** We present the LDM, a method that integrates global testing of any microbiome effect and detection of differentially abundant OTUs. The LDM is generally more powerful than existing methods, and is capable of handling the confounders and correlated data that frequently occur in modern microbiome studies.

1519T**Meta-analysis of vaginal microbiome data provides new insights into preterm birth.** *I. Kostj^{1,2}, S. Lyalina³, K.S. Pollard^{4,5,6}, A.J. Butte^{1,2}, M. Sirota^{1,2}.*

1) Institute for Computational Health Sciences, University of California, San Francisco, CA 94158, USA; 2) Department of Pediatrics, University of California, San Francisco, CA 94158, USA; 3) Integrative Program in Quantitative Biology, Gladstone Institutes, University of California, San Francisco, San Francisco, CA 94143, USA; 4) Department of Epidemiology & Biostatistics, Institute for Human Genetics, and Quantitative Biology Institute, University of California, San Francisco, San Francisco, CA 94143, USA; 5) Gladstone Institutes, San Francisco, CA 94158, USA; 6) Chan-Zuckerberg Biohub, San Francisco, CA 94158, USA.

Preterm birth (PTB) is defined as the birth of an infant before 37 weeks of gestational age. It is the leading cause of perinatal morbidity and mortality worldwide. In this study, we present a comprehensive meta-analysis of vaginal microbiome in preterm birth. We integrated raw longitudinal 16S ribosomal RNA vaginal microbiome data from five independent studies across over 3,000 samples and were able to gain new insights into the vaginal microbiome state in women who deliver preterm in comparison to those who deliver at term. We found that women who later have a preterm delivery show a higher variance in vaginal microbiome abundance, with the most significant difference observed during the first and second trimesters. Modeling the data longitudinally revealed a number specific microbial genera as associated with PTB, including several that were previously known and two newly identified as associated with PTB by this meta-analysis: *Olsenella* and *Clostridium sensu stricto*. New hypotheses emerging from such an integrative analysis can lead to novel diagnostics to identify women who are at higher risk for PTB and potentially inform new therapeutic interventions.

1520F**Identification of microRNA target sites: In-silico screening on NGS data.**

V. Rallo, A. Angius. Istituto di Ricerca Genetica e Biomedica, Monserrato CA, Cagliari, Italy.

MiRNAs may be biomarkers associated with the progression of several diseases and great efforts have been made to discover miRNAs, identify miRNA targets and infer miRNA functions. Different bioinformatic tools and algorithms enabled to predict miRNAs binding regions together with several databases that provided information about miRNA sequences. WGS, WES and RNAseq datasets can represent an optimal support to correlate variants with miRNA target sites at no additional cost. Our goal is to link variant calling information with a potential gain or loss of miRNAs target sites. So far, there is no simply logical and sequential procedure to do this. Using already existing software, we developed a simple pipeline that allows us to identify the impact of variants on miRNA target sites with a biological role. We created a unique database of miRNA's sequences that collected all existing miRNA databases (miRBase, tarbase, mirSNP, etc) and performed prediction analyses of target sites focused on the most performing existing algorithms (MiRanda, TargetScan). For example, starting from about 23,000 genomic variants that lie in 3'UTR region, a simple input command enabled us to select 4637 lost and 4302 gained target sites. The output is a simple tabular text file contains genomic variants, target site, variant impact, associated miRNAs, loss and gain sites and annotation data coming from different databases. The output result was designed to be used in subsequent standard software to apply filters based on the desired mode of inheritance and on sample's genotype quality and variant frequencies. This method could be particularly useful on WES data when the identification of the causal variant has failed and the involvement of miRNAs is assumed, without additional costs. This new automated sequential pipeline allows to identify, in silico, the variants involved in the miRNA targets from VCF files and to generate results easily interpreted and usable by people without bioinformatic skills.

1521W

The IMPC: Insights into the correlation of gene essentiality and human disease. P. Cacheiro¹, V. Muñoz-Fuentes², H. Haseli³, H. Westerberg³, S. Murray⁴, M.E. Dickinson⁵, A.M. Mallon³, T.F. Meehan², D. Smedley⁴, the IMPC consortium. 1) William Harvey Research Institute, Queen Mary University of London, London, UK; 2) European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, Cambridge, UK; 3) Medical Research Council Harwell, Mammalian Genetics Unit and Mary Lyon Centre, Harwell, Oxfordshire, UK; 4) The Jackson Laboratory, Bar Harbor, ME, USA; 5) Department of Molecular Physiology and Biophysics, Baylor College of Medicine, Houston, TX, USA.

The International Mouse Phenotyping Consortium (IMPC) is building a functional catalogue of the mammalian genome by producing and phenotyping a knockout mouse line for every protein-coding gene with over 5,000 knockout lines screened to date. One-third of these lines result in embryonic or perinatal lethality/subviability and can be compared with orthologous essential genes discovered in high-throughput screens of human cell lines. Whereas the human cell line studies have found that approximately 2,000 genes are essential for cellular life, the IMPC is on track to identify 6,000-7,000 genes necessary for organism development. In the present study we combined the information derived from studying viability at the cellular and organism level to generate different functional bins representing a spectrum of gene essentiality. A total of 3,354 genes with viability data from both cell and mouse reports and meeting strict quality criteria were subdivided into 5 functional bins: cellular essential (essential both in human cells and mice), developmental essential (non-essential in cell lines and essential in mice), subviable, viable with phenotypic abnormalities and viable with no abnormal phenotype detected. Whereas 99% of the mouse non-essential genes (viable) were also non-essential in human cells, only 36% of the genes that were lethal in mouse were also essential in cell lines. The remaining lethal genes -developmental essential- were found to be associated to biological processes involved in morphogenesis and tissue development and showed a very high enrichment for disease genes. The disease genes in each of the functional bins also showed a clear differentiation in terms of age of onset, number of physiological systems affected and mode of inheritance. Additional features such as levels of gene expression or the percentage of genes being part of a protein complex were correlated with the level of essentiality. By combining human cell and mouse viability information we are able to capture different levels of gene essentiality. Rare disease genes are not only unequally distributed among these categories, but they also present differences with respect to the type of disorders they are associated with. Practical implications emerging from these findings would include the prioritization of variants in sequencing studies and the use of machine learning to predict new lethal and disease genes based on the distinctive features they present.

1522T

Statistical methods inferring fetal fractions promise the most powerful noninvasive prenatal screen. M. Dang^{1,5}, H. Xu², Y. Guan^{1,3,4,5}. 1) USDA/ARS Children's Nutrition Research Center, Baylor College of Medicine, Houston, TX; 2) Beijing USCI Medical Laboratory, China; 3) Department of Pediatrics, Baylor College of Medicine, Houston, TX; 4) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 5) Center for Statistical Genetics and Genomics, Department of Biostatistics and Bioinformatics, Duke University, Durham, NC.

Fetal fraction plays a pivotal role in noninvasive prenatal screen (NIPS), whose aim is to detect trisomy in fetus. The fetal fraction can be estimated from sex chromosomes for male fetuses, but not for female fetuses. In addition, such estimates are inaccurate when the fetuses have sex chromosomal trisomy XXY or XYY, or monosomy X. Here we present a method that can reliably estimate fetal fractions for both male and female fetus by examining autosome coverage. First, even at low coverage, there are sizable number of 1000 genomes SNPs that are covered by more than one reads. This can be deduced simply by assuming Poisson distribution at each locus and verified by real data. Second, at those SNPs covered by at least two reads, we can define heterozygosity. Third, the heterozygosity is a function of fetal fraction, which in turn allows fetal fraction to be inferred. With a known fetal fraction, the performance of NIPS can be dramatically improved in the following aspects. First, the traditional Z-test likelihood only accounts for how far the inferred chromosomal dosage is away from the null; the inferred fetal fraction enable us to also account for how far the inferred chromosomal dosage is away from the alternative, which is the most powerful test for aneuploidy. If we treat microdeletion and microduplication as aneuploidy of a small artificial chromosome, our method promises the most powerful NIPT Plus. Second, for samples with extremely low fetal fractions we may declare "no call," which removes a major source for false negative. Third, testing aneuploidy of sex chromosomes, such as XXY, XYY, and X, becomes a simple problem when the fetal fraction is known.

1523F

The pharmacogene variant (PharmVar) consortium and database. A.

Gaedigk^{1,2}, G. Twist³, M. Whirl-Carrillo⁴, K. Sangkuhl¹, S. Casey², T. Klein¹, N. Miller^{2,3}. 1) Division of Clinical Pharmacology, Toxicology & Therapeutic Innovation, Children's Mercy, Kansas City, Kansas City, MO; 2) School of Medicine, University of Missouri-Kansas City, Kansas City, MO; 3) Center for Pediatric Genomic Medicine, Children's Mercy, Kansas City, Kansas City, MO; 4) Department of Biomedical Data Science, Stanford University, Stanford, CA.

Background: Broad implementation of pharmacogenetics (PGx) requires a stable nomenclature for allelic variation in pharmacogenes. Since its inception in 2002, the Human Cytochrome P450 Allele Nomenclature Committee has been the authority for cataloging allelic variation of CYP genes. In October, 2017 the Nomenclature Database was transitioned to the PharmVar Consortium at Children's Mercy Kansas City (www.PharmVar.org). **Methods:** To facilitate this transition, a stakeholder group developed a plan to sustain what was agreed upon as an essential resource. The PharmVar Consortium, a NIH-funded PGRN resource (pgn.org/pharmvar.html), was established to serve as a centralized 'Next-Generation' pharmacogene variation data repository.

Results: The PharmVar database was launched in March 2018 featuring three high priority genes, *CYP2C9*, *CYP2C19* and *CYP2D6* with additional genes to follow in Summer 2018. PharmVar provides consistent mapping of variants across multiple reference sequences including the genomic locus reference sequence from RefSeqGene or Locus Reference Genomic (LRG) sequence if available, a representative sequence from RefSeqTranscript and genome builds GRCh37 and GRCh38. Website features include the ability to dynamically switch coordinates between counting from the sequence start or the ATG start codon, an overview page showing all haplotypes sharing a sequence variant and links to external resources such as dbSNP, NCBI Gene, HGNC and PharmGKB. Haplotype data is available for download in FASTA, VCF and TSV formats. PharmVar is developing a revised haplotype naming system that accommodates the designation of suballeles as well as an evidence code system to clearly represent the strength of evidence supporting each allele/haplotype definition. PharmVar will simplify allele submission by developing online tools for users to deposit data. Submissions will be reviewed by gene expert panels composed of PharmVar community members. Expansion to non-CYP PGx genes are also underway. PharmVar efforts are synchronized with the Pharmacogenomic KnowledgeBase (PharmGKB) and the Clinical Pharmacogenetics Implementation Consortium (CPIC) to provide the PGx community with standardized information for implementation of PGx. **Conclusions:** PharmVar provides a rich resource for the entire PGx community. PharmVar invites anyone with an interest in PGx variation to become a consortium member and actively participate to move pharmacogene nomenclature efforts forward.

1524W

Large-scale characterization of CYP2D6 variation in African Americans using TOPMed whole genome sequencing data. S. Lee¹, M.M. Wheeler¹, J.G. Wilson², A. Correa³, P. Anugu³, D.A. Nickerson^{1,4}, NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium. 1) Department of Genome Sciences, School of Medicine, University of Washington, Seattle, WA, USA; 2) Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS, USA; 3) Department of Medicine, University of Mississippi Medical Center, Jackson, MS, USA; 4) Brotman Baty Institute for Precision Medicine, Seattle, WA, USA.

Genotyping *CYP2D6* is important for precision drug therapy because the enzyme it encodes metabolizes about 25% of drugs, and its activity varies substantially within and between populations. Over 100 haplotypes or star alleles have been defined for *CYP2D6*. These are characterized by single nucleotide variants, insertion-deletion variants, and structural variants that include gene deletions, duplications, and hybridizations with a nearby non-functional but highly homologous paralog *CYP2D7*. Although many star alleles are known to exhibit distinct population frequencies, further investigation of *CYP2D6* variation is warranted in non-European populations because these are underrepresented in the overall estimation of *CYP2D6* genetic diversity. In this study, our goal was to comprehensively survey *CYP2D6* variation in a large African American cohort. Specifically, we performed *CYP2D6* genotype analysis of 3,418 African American samples from the Jackson Heart Study that were whole genome sequenced by the Trans-Omics for Precision Medicine (TOPMed) program. These samples were genotyped using Stargazer, a bioinformatics tool we recently developed for accurate calling of star alleles in various polymorphic pharmacogenes, including *CYP2D6*. From our samples, we found a total of 41 unique haplotypes including the decreased function *CYP2D6**17 (16.6%) and *29 (7.8%) alleles that are commonly found in individuals of African ancestry and the non-functional *CYP2D6**4 allele (3.1%). We also observed extensive structural variation ranging from zero to five gene copies and numerous *CYP2D6*/*CYP2D7* hybrids. That is, 22 out of 41 detected haplotypes had structural variation: 5.4% with a gene deletion (*CYP2D6**5), 7.0% with a gene duplication (*CYP2D6**1x2, *2x2, *4x2, *10x2, *29x2, *34x2, *41x2, *43x2, *4N+*4, *36+*10, *68+*4, *77+*2, and *78+*2), 0.1% with a gene multiplication (*CYP2D6**1x3, *2x3, *4x3, *29x3, and *34x3), and 1.3% with a gene hybridization (*CYP2D6**4N, *36, *66, *68, *76, *77, and *78). Overall, 25.5% of the samples had at least one type of structural variation and based on diplotypes calls, 1.6%, 8.9%, 84.4%, and 3.9% were predicted to be poor, intermediate, normal, and ultrarapid metabolizers, respectively. We are currently expanding our analysis to include other cohorts in the TOPMed program. These results demonstrate the importance of enhancing our understanding of *CYP2D6* genetic diversity in large and diverse datasets in order to achieve precision drug therapy.

1525T

High-throughput CLIA QC pipeline for PGx EHR integration. *M. Saylor¹, N. Rafaels², K. Crooks², T. Brunetti², L. Koivukoski¹, B. Jackson¹, N. Hurskainen¹, P. Sevon¹, T. Kanninen¹, A. Ahonen-Bishopp¹, L. Eronen¹.* 1) BC Platforms, Boston, MA; 2) Colorado Center for Personalized Medicine (CCPM), Aurora, CO.

Genotyping is a powerful method for generating genetic insights for patient care and research. Compared to next-generation sequencing (NGS), genotyping is only a fraction of the cost of NGS and is already in practice facilitating population scale adoption. Additional healthcare costs caused by incorrect medication or dosage are significant, and genotype based PGx reporting has potential for major cost savings as well as improving quality of care. However, in adopting PGx on a large scale, there are challenges in maintaining high-throughput and in ensuring data production and insights from produced data are done following clinical practices (CLIA, HIPAA). We have developed an automated pipeline that enables high-throughput processing of Illumina array data, PGx reporting, and direct delivery of results to Epic. Typically, Illumina array data production depends on GenomeStudio software for QC. This pipeline automatically processes gtc files without the need to use GenomeStudio. Data is processed according to quality thresholds, both on the sample level and on the SNP level, also including visuals of SNP intensity data to assess quality. The pipeline is automated while also allowing lab director level overrides. The software supports approval of data for further processing (either for clinical and/or research use), within a high-security audited software and hardware environment. For PGx reporting, the star allele and metabolizer status for each sample and SNP is calculated by the pipeline automatically using the combination of a custom algorithm and a PGx knowledge base. An individual's predicted drug response is valuable information in patient care, this information is communicated immediately to Epic through HL7 messaging. The pipeline can be installed to a variety of environments, including utilization of object/blob storage provided by most cloud vendors, as well as utilizing on-premise infrastructure with connected calculation cluster and file storage. The pipeline is scalable for running genotype data production for biobanks or healthcare institutions targeting population scale, where daily throughput can be thousands of samples. The pipeline can also be complemented by further automated parallel imputation pipeline research purposes. This process makes it possible to produce genotype data in massive scale to be utilized for clinical reporting, population level research and biobanking.

1526F

PubCaseFinder: A phenotype-driven differential diagnosis system for rare diseases based on case reports. *T. Fujiwara^{1,2}, Y. Yamamoto¹, J.D. Kim¹, T. Takagi².* 1) Database Center for Life Science, Kashiwa-shi, Chiba, Japan; 2) Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa-shi, Chiba, Japan.

Next-generation sequencing (NGS)-based diagnostic tests, which particularly employ clinical exome sequencing (CES), have been attempted for diagnosis of patients with suspected rare diseases. Clinical sensitivity of CES, i.e., the percentage of patients who were diagnosed by CES, is up to 40% in several studies. Unfortunately, other ~60% of patients with suspected rare diseases don't receive a diagnosis, and such patients will likely lose opportunities such as optimization of clinical management and early intervention. Recently, to aid these patients, phenotype-driven differential diagnosis systems such as Phenomizer and FACE2GENE have been implemented. These systems provide a ranked list of rare diseases based on the phenotypic similarity between a patient and rare diseases, and the top-listed diseases represent the most likely differential diagnosis. The performance of these systems is greatly influenced by the quantity and quality of underlying databases of disease-phenotype associations (DPAs). Note that these databases rely on manual curation and show a limited coverage. In the case of DPA database from Orphanet, more than half of the rare diseases (~60.5% of 6,268) are not associated with a phenotype. The development of DPA databases is based on the curation of papers by human experts, which is time-consuming and labor-intensive because of the large volume and rapid growth of life sciences papers. To address the problem, we developed a text mining approach to extend the coverage of DPAs in manually curated databases like Orphanet. By applying the approach to a million case reports from PubMed, we could increase the coverage of DPAs from Orphanet more than 2 times. Based on the extended DPA database, we also developed PubCaseFinder (<https://pubcasefinder.dbcls.jp>), a new phenotype-driven differential diagnosis system. A series of experiments which was conducted using clinical cases from PhenomeCentral showed that the performance of PubCaseFinder could substantially be improved thanks to the extension of the DPA database. Note that automatic text mining techniques are often regarded as assistive tools to help manual curation of databases, due to its potentially high chance of noisy results. However, our experimental results suggest that even before a manual screening process, text mining approach can be substantially helpful. Our proposed approach is deemed practically useful, because manual curation will take enormous time and cost.

1527W

Genetic history of Russian indigenous populations and their genome

medical mining. V. Brukhin¹, D. Zhernakova^{1,2}, S. Malov¹, K.P. Koepfli^{1,3}, A. Zhuk¹, P. Dobrynin^{1,3}, S. Kliver¹, N. Cherkasov¹, G. Tamazian¹, M. Rotkevich¹, K. Krashenninnikova¹, I. Evsyukov¹, A. Gorbunova¹, E. Chernyaeva¹, A. Komissarov¹, D.E. Polev¹, A. Glotov¹, A. Novozhilov¹, S.J. O'Brien¹, V. Osakovsky⁴.

1) Dobzhansky Center for Genome Bioinformatics, St. Petersburg State University, St. Petersburg, Russian Federation; 2) University of Groningen, University Medical Center Groningen, Department of Genetics, Groningen, the Netherlands; 3) 1.Smithsonian Conservation Biology Institute, National Zoological Park, Washington, D.C. USA; 4) Institute of Health, North-Eastern Federal University, Yakutsk, Russian Federation.

Despite a large population size and high ethnic diversity, no centralized reference database of functional and endemic genetic variation has been established to date for Russian human populations. Such data are crucial for studying population history and would be essential for medical genetic purposes. The Genome Russia Project aims at filling this gap by performing a high coverage whole genome sequencing and analysis of people from the Russian Federation. A comparative analysis of whole genome sequencing (WGS) was initiated at the St. Petersburg State University. We assessed calling of SNPs, indels, copy number variations from 264 healthy adults (from 52 indigenous Russian populations), including 60 newly sequenced samples consisting of family trios from three geographic regions: Pskov, Novgorod (Western Russians) and Yakutia (Yakut). People of Russia are shown to carry known and novel genetic variants of adaptive, clinical and functional consequence that in many cases show appreciable occurrence or allele frequency divergence from the neighboring Eurasian populations. Principal component and phylogenetic analyses of overall variation revealed strong geographic partitions among indigenous ethnicities corresponding to the geographic locales where they have lived. Allele frequency spectra identified strong constraints to gene flow that correspond to the geological barriers (e.g. the Ural Mountains and Verkhoyansk mountain range). In conclusion, this study presents a genomic characterization of population-specific variation in Russia with results important for medical genetics as well as for population genetic history studies.

1528T

GRAF-pop: A novel tool to quickly infer subject ancestry from multiple genotype datasets.

Y. Jin, A.A. Schaffer, M. Feolo, J.B. Holmes, B.L. Kattman. National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD.

It is essential to be able to infer subject ancestry from genetic data in genome wide association studies, since unhandled population stratification can result in false associations, and self-reported ancestries may not be accurate. Many of the commonly used methods to infer subject ancestry from genetic data rely on principal components analysis (PCA). However, PCA has two important limitations: (1) PCA results are not comparable across data sets; (2) PCA does not work when subjects have high genotype missing rates. The two limitations have become increasingly prominent since investigators now can access more and more genotypes obtained using different platforms, such as the data sets obtained from the National Center for Biotechnology Information's (NCBI) Database of Phenotypes and Genotypes (dbGaP). We have developed an ancestry inference tool, GRAF-pop, that is robust to missing genotypes, fast and easy to use. The results of GRAF-pop can be generalized across datasets obtained using different platforms. GRAF-pop is implemented as a feature of the GRAF (Genetic Relationship and Fingerprinting) software package, using pre-selected 10,000 well separated, autosomal biallelic SNPs with high minor allele frequencies to infer subject ancestry. It calculates genetic distances from each subject to several reference populations from dbGaP and projects the results to a pre-defined 2D plane without using PCA. Barycentric coordinates are used to estimate the ancestry proportions of each subject based on the relative distances to the reference populations, and to normalize the scores to overcome the problem of missing genotypes. GRAF-pop takes PLINK formatted genotypes as input, and outputs assigned populations and estimated ancestral components in tables and scatter plots. No data pre-processing, e.g., filtering SNPs and samples, removing related subjects, are needed to run the program. Results obtained from different datasets can be combined and plotted onto the same graph. GRAF-pop is fast, with runtime linear to the number of individuals. The current version of GRAF is freely downloadable from the following webpage: <https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/Software.cgi>. This presentation will describe the algorithms and use of GRAF-pop by dbGaP staff and data submitters and users, and summarize the results of ancestry inference of over one million dbGaP subjects. .

1529F

Uncertainty in genetic ancestry: The importance of structure in the ancestral populations. *F.L. Mendez, M.L.A. Isaksson, S. Ruddy, W. Lee.* Helix, LLC, SAN CARLOS, CA.

Understanding how the genetic ancestry of individuals is partitioned among ancestral populations is central to the study of human population history and fundamental for the reconstruction of an individual's ancestry. In diverse and admixed populations, knowledge of the genetic ancestry may also inform medical decisions and aid the interpretation of significance of genetic variants. One of the most popular methods to characterize genetic ancestry explains the genome of individuals as the combination of genetic material originated from discrete population clusters (Structure method). For clusters obtained in unsupervised analysis, populations are typically explained as combinations of multiple clusters. Alternatively, a supervised analysis could be used to describe individuals in terms of externally defined groups; for instance, ancestral populations could represent continental ancestry and individuals from various populations within that continent would be used to create a panel for the corresponding cluster. However, when there is structure within a group, the characteristics of its cluster will depend on the relative representation of its subgroups. This may often result in discrepancies in ancestry estimates when reference panels are very dissimilar in how they weigh subgroups in a cluster. We explored the variation in ancestry estimates and its dependence on the choice of the individuals representing each of six ancestral groups: sub-Saharan Africans, East Asians, South Asians, Europeans, Indigenous Americans, and Ashkenazi Jews. Ancestry estimates exhibit different degrees of sensitivity to the choice of representatives for each group. Estimates of European, South Asian and Ashkenazi Jewish ancestry are considerably more sensitive to the choice of individuals in the reference panels. This is especially true for European ancestry: individuals with southern and southeastern European ancestry are inferred as having Ashkenazi Jewish ancestry when the European panel relies on northern or western Europeans. We use the observed variation in estimated ancestry to provide a measure of the uncertainty associated with inferring ancestry. We further use the relationship between ancestry estimates and their uncertainty to provide confidence bounds for ancestry estimates of new samples. Our research addresses a prominent reason for discrepancy in ancestry estimates: differences in population reference panels; it also provides a framework to assess its extent.

1530W

A framework for analytical validation of SNP arrays for clinical diagnostics. *P.F. Cherukuri^{1,2,3}, J.M. Lubieniecka^{1,2,3}, K.P. Lubieniecki^{1,2,3}, S.J. Ahuja¹, R. Pyatt^{1,2,3}, C.F. Boerkoel¹, L. Carmichael¹.* 1) Imagenetics, Sanford Health, Sioux Falls, SD; 2) Sanford School of Medicine, University of South Dakota, Sioux Falls, SD; 3) Sanford Research Center, Sioux Falls, SD.

Background: High-throughput, low-cost genotyping is used for many applications in precision medicine as well as population genomics. Consortia such as the 1000 Genomes project have made available whole genome sequence data from diverse population groups that serve as reference datasets. Additionally, curation of variants causal for diseases (e.g. ACMG59), risk associations (e.g. Cardiovascular), drug metabolism (e.g. Pharmacogenomics), and other traits, have markedly improved. **Hypothesis:** We hypothesize, therefore, that genotyping arrays can be validated to screen clinically for sequence variants. **Methods:** We used the Illumina Infinium Global Screening Array (GSA), which assays >660,000 variants in a 24-sample format, to test this hypothesis. We divided the analytical validation into theoretical and empirical components. For the theoretical component, we used a Bayesian theoretical framework to calculate the expected genotyping error in the GSA-targeted genotypes based on whole-genome sequence-based variation. For the empirical component, we used orthogonal approaches including parallel whole genome sequencing (WGS) (>30x coverage), TaqMan assays, and Sanger sequencing. To improve genotyping accuracy further, the GSA data of 1,200 individuals were manually curated for quality control (QC) and used to create a reference cluster file. **Results:** Theoretical evaluation of false positives and false negatives for single nucleotide variants predicted that we will find ~ 3 true false negatives per GSA run based on WGS and GSA whole genome discordance rates. This modelling was based on a single individual replicate dataset ($N=23$; NA12878) and assumed (1) a probability of variation given a true homozygous reference of 0.0038, (2) a probability of homozygous reference given a true variant was 0.0043, and (3) a prior risk of 1/1000 for genome variation. Genomic regions with higher variation rate will have higher chance of false negatives (e.g. $2/1000 = \sim 6$ false negatives per GSA run given all probes map to highly variable regions). Empirical comparison of replicate data sets from the GSA to a truth dataset from WGS was used to refine the performance metrics (sensitivity, specificity, accuracy, precision, etc.). Preliminary empirical results show high sensitivity and specificity (>0.99) across all GSA probes passing QC. **Conclusion:** A combined theoretical and empirical approach can analytically validate single nucleotide variant arrays for clinical use.

1531T

Comparison of clinical interpretation of genetic variation in the eMERGEseq cohort using multiple methods. D.R. Crosslin¹, P.H. Byers¹, W.K. Chung², J.J. Connolly³, P.K. Crane¹, M.O. Dorschner¹, D. Fasel⁴, A.S. Gordon¹, T.O. Hall¹, M. Harr⁵, K.M. Howell¹, J.E. Jackson⁴, P. Jia⁵, B.R. Korf⁶, I.J. Kullo⁷, E.B. Larson⁸, K. Leppig⁹, W.H. Liang⁸, N.M. Lindor⁷, T. Manolio⁹, P.C. Mathias¹, A.H. Ramirez⁴, Y. Shen², B.H. Shirts¹, K.V. Slattery¹, P.M.A. Sleiman⁹, I.B. Stanaway¹, C. Weng², Z. Zhao⁵, G.P. Jarvik¹. 1) University of Washington, Seattle, WA; 2) Columbia University, New York, NY; 3) Children's Hospital of Philadelphia, Philadelphia, PA; 4) Vanderbilt University, Nashville, TN; 5) University of Texas Health Science Center at Houston, Houston, TX; 6) University of Alabama at Birmingham, Birmingham, AL; 7) Mayo Clinic, Rochester, MN; 8) Kaiser Permanente Washington Health Research Institute, Seattle, WA; 9) National Human Genome Research Institute, Bethesda, MD.

Introduction: The Electronic Medical Records and Genomics (eMERGE) Network has developed eMERGEseq, a panel of 106 genes comprising the "ACMG 56" actionable genes and site-requested genes, as well as approximately 1500 SNVs in other genes. The goal is to eventually return actionable results and incorporate them into the electronic health record with clinical decision support for 25,000 participants. We propose assessing the correlation between interpretations of pathogenicity of genetic variants identified in the eMERGEseq cohort using three sources, with one considered the "gold standard." **Methods:** Stratifying by cancer genetics, cardiovascular genetics, and neurogenetics, we will compare pathogenicity classifications generated by 3 methods: 1. CLIA results from two clinical labs generating clinical reports for eMERGEseq; the pathogenicity interpretation (pathogenic, likely pathogenic, variant of uncertain significance, likely benign, and benign) in these reports will be considered the "gold standard;" 2. The in silico clinical pathogenicity interpretation tool Intervar; and 3. Crowdsourcing participating investigators for clinical interpretation using molecular annotation and phenotype indication via RedCap survey tools. We will use a Cohen's κ coefficient to assess qualitative correlation among the three groups. **Results:** Response to participation in the crowdsourcing project has been enthusiastic with >50 participants at the time of this proposal, including representation from eMERGE, CSER, AMIA, and NHGRI, among others. We have captured the multiple biography variables for our crowdsourcing participants. For instance, highest degree achieved is as follows: 1. MD (22.2%); 2. PhD (44.4%); 3. MD-PhD (19.4%); 4. MS (11.1%); and 5. BS (2.8%). Indication expertise is as follows: 1. Cancer (8.3%); Cardiovascular (2.8%); 3. Neurology (2.8%); 4. Immunology (2.8%); 5. Combination of two or more (27.8%); 6. Other (25.0%); 7. Not an expert (25.0%); and 8. NA (5.6%). At present, the "gold standard" variants have been collected from the eMERGE sequencing centers, and the eMERGE Coordinating Center has ensemble-called variants in ~15,000 participants. **Conclusion:** As the focus of clinical sequencing research moves from targeted panels to exomes and ultimately genomes, it will be advantageous to assess tools that can assist in the clinical interpretation on a larger scale.

1532F

Automating variant assessment using structural features. A.C. Chamberlin, M.S. Park, I.D. Petrik, H.M. Lu, R.B. Fenwick. Bioinformatics, Ambry Genetics Corp, Aliso Viejo, CA.

Variant assessment is rapidly becoming a significant part of clinical diagnoses in both clinical and somatic diseases. As the number of uncategorized variants continues to grow, automated classification schemes will become more and more valuable. The application and automation of simple classification schemes using sequence, and ternary-structural features will be invaluable in prioritizing variants for assessment, classification and functional studies. Structural biology can provide a variety of features for variant assessment that can be leveraged in parallel with traditional methods such as *in-silico* predictions that rely on sequence primary-structure (Baugh et al., 2016). These features can include but are not exclusive to, structural motif identification, variant location in the domain (buried, surface, or interfacial), as well as information about the local environment of the variant (hydrogen bonding, ligand binding, back-bone conformation, etc), all of which provide invaluable insight into variant pathology (Pesaran et al., 2016; Ponzoni & Bahar, 2018). Here we present an automated method for leveraging structural features for variant assessment, which can be used as a tool to both propose appropriate categories for variants of interest and allow prioritization of variants for reporting and reassessment purposes.

1533W

Leveraging external variant frequency information to increase statistical power in rare variant association studies. *J. Chen, R. Callender, R. Bohlender, Y. Yu, M. Hildebrandt, C. Huff.* MD Anderson Cancer Center, Houston, TX.

Allele frequency summary data from large public databases has been a valuable resource in the search for rare genetic variation influencing human disease. However, such information has proved difficult to fully leverage in sequence-based case-control studies. Here, we introduce a novel approach that uses external allele frequency information from a closely related population to enrich for rare variants likely to be associated with a trait in sequencing case-control studies. This approach reduces the multiple testing burden in single marker tests and increases the effective proportion of causal variants in gene-based tests, both resulting in an increase in statistical power. We benchmark this approach in a simulation study involving a wide variety of risk patterns and observe substantial increases in power in the broad majority of scenarios. For example, using a single marker test for a variant with a minor allele frequency (MAF) of 0.01 and an odds ratio (OR) of 1.6, our approach achieves 80% power compared 20% power using standard techniques ($\alpha = 5 \times 10^{-8}$). In another example, using a gene-based test for variants with MAF < 0.005 and OR of 2, our approach yields a p value of 3.7×10^{-5} , compared to a p value of 0.0026 using standard techniques. Importantly, this approach does not inflate Type I error and is not influenced by cross-platform stratification biases under the null model. We further demonstrate the utility of this method in case-control analyses involving whole-exome sequencing data from 783 breast cancer cases, 272 ovarian cancer cases, 322 melanoma cases, and 3507 shared controls, using ExAC as the external reference panel.

1534T

Benchmarking and visualization of short tandem repeats to enable genome-wide discovery and clinical interpretation. *V.B. Deshpande, E. Dolzhenko, S. Chen, D.R. Bentley, M.A. Eberle.* Illumina Inc., 5200 Illumina Way, San Diego, CA-92122, USA.

Expansions of short tandem repeats (STRs) are responsible for more than 20 genetic diseases and many new pathogenic repeat expansions are likely to be discovered. Recently we presented ExpansionHunter (EH), a tool to detect pathogenic repeat expansions using PCR-free whole genome sequence data. We have applied EH to the known pathogenic repeats and the next step is to enable discovery of novel pathogenic STRs anywhere in the genome. As we extend STR genotyping to loci across the genome, we expect that many STR loci will present unique challenges and it is important to systematically assess the genotyping accuracy at each locus. For this purpose, we developed a benchmarking pipeline which reports the Mendelian consistency rate at each STR locus in 50 trios and the 17-member Platinum Genomes pedigree. Additionally, because STRs are difficult to inspect with tools such as IGV which do not display the inserted sequence, we developed a custom visualization tool that displays a pileup of all read alignments reported by the STR genotyping tool against a custom reference containing the longest repeat alleles. Our visualization tool allows rapid assessment of STR genotypes for method development and will be invaluable for the inclusion of EH into clinical pipelines. As a demonstration of this pipeline, we assessed genotype calls made by EH at 24 STR loci associated with monogenic disorders. When we observed Mendelian conflicts, we used the visualization tool to correct the individual genotype calls and finally created a curated truth set that can be used as a benchmarking resource and for regression testing. To extend this database beyond the known pathogenic repeats, and enable researchers to genotype these STRs in their genomic studies, we are now building a genome-wide database of all STRs in genic regions. We will present a review of the genome-wide STR landscape including an assessment of the performance of these variants using our Mendelian validation pipeline. In conclusion, we provide a framework to assess STR genotyping accuracy for each STR locus, a tool to curate STR predictions through visualization and a curated truth set resource for benchmarking.

1535F

Transcriptome analysis by RNA-sequencing as an adjunct to whole genome analysis in undiagnosed genetic disease. *N. Balanda¹, M. Malicdan^{1,2,3}, D. Adams^{1,2,3}, W. Gahl^{1,2,3}, Undiagnosed Diseases Network.* 1) NIH Undiagnosed Diseases Program, NHGRI, National Institutes of Health, Bethesda, Maryland, USA; 2) Medical Genetics Branch, National Human Genome Research Institute, NIH, Bethesda, Maryland, USA; 3) Office of the Clinical Director, National Human Genome Research Institute, NIH, Bethesda, Maryland, USA.

INTRODUCTION: The NIH Undiagnosed Diseases Program (NIH-UDP) aims to uncover genetic causes for rare, previously undiagnosed diseases. In the 10 years since its conception, the UDP has made extensive use of genome-scale DNA sequencing, in combination with comprehensive clinical and biochemical phenotyping, to identify diagnostic hypotheses and make diagnoses. By these efforts, approximately 25% of the over 1,000 patients seen during this period have received diagnoses. Here, we employ functional genomics and expand the use of transcriptome analysis as an adjunct to DNA-based sequencing, with the aim of increasing the rate of diagnosis in our patients. We describe the progress made in a pilot study using transcriptome to augment the analysis of whole genome sequencing (WGS) in UDP participants. **METHODS:** DNA-based sequencing of patients and family members was done through whole genome sequencing. For a cohort of 130 probands without a current diagnosis, total RNA was extracted from patient-derived dermal fibroblasts, purified for cDNA library preparation, and sequenced using Illumina HiSeq. Data were then aligned to the GRCh37/hg19 human reference and analyzed using a custom built pipeline. Given the highly-heterogenous nature of the UDP cohort, transcriptome analysis was performed for each proband using all 129 other members of the cohort as controls. **RESULTS:** Agnostic reanalysis of WGS, revealed that the majority of these 130 cases harbor non-coding variants of interest with unknown significance. We thus utilized RNAseq to verify the biological consequences of these variants, and positive findings in many of these cases suggest that individual hypotheses positing a relationship between specific DNA variants and splicing, or transcriptional regulation can be supported or refuted with RNASeq data. In addition to validating WGS findings, this method contributes new variants of interest to WGS analysis by detecting association between genomic variants and aberrant splicing/expression profiles. **CONCLUSIONS:** Prioritization of DNA sequence variants for further study is a substantial undertaking for cohorts of individuals with undiagnosed diseases. Functional genomics, including the utilization of RNASeq studies, can provide additional information that is useful for specific DNA-variant-associated hypotheses, and aid in the prioritization of genetic variants. Our results support the routine use of RNASeq analysis in improving the overall diagnostic rate over time.

1536W

QTR-seq: A tool for detecting quantitative trait-associated genes in RNA-seq data. *M. Seo^{1,2}, D. Qiao^{1,2}, W. Qiu^{1,2}, P. Castaldi^{1,2}, CP. Hersh^{1,2}.* 1) Chan-ning Division of Network Medicine, Brigham and Women's Hospital, Boston, MA; 2) Department of Medicine, Harvard Medical School, Boston, MA.

Most of the statistical analysis methods for RNA-sequencing data have been developed to focus on group comparisons such as specific experimental conditions or cases and controls. The statistical methods for quantitative trait association testing are not as well established due to the practical limitations such as limited sample sizes. Herein, we present a systematic and flexible approach for detecting continuous trait-associated genes in RNA-seq data. Using simulation studies, we evaluated the variable parameter settings for detecting quantitative trait-associated genes, including effective sample size, gene filtering strategies, normalization methods, and statistical estimators. We found that robust estimators, such as M, S, and MM estimators that account for outlier effects in both measured traits and gene expression, outperform ordinary least-squares regression in terms of false positive rate. We applied the method to 1,211 blood RNA-seq samples from the Genetic Epidemiology of Chronic Obstructive Pulmonary Disease (COPDGene) study. Since cigarette smoking is a major risk factor for COPD, we tested for gene expression-by-smoking interaction associations with quantitative measures of lung function, which could not be performed in existing statistical methods for RNA-seq data. We found that 339 genes are differentially associated with the forced expiratory volumes in 1 second (FEV1) percent predicted in current compared to former smokers at q-value < 0.05. Finally, we implemented QTR-seq, a flexible R package for detecting quantitative trait-associated genes, which is able to incorporate adjustment for covariates, gene-by-environment interactions, and multivariate outcomes.

1537T

Identification of transcriptional effects from loss of the Y chromosome (LOY) in leukocytes using single cell sequencing. *J. Halvardson^{1,2}, M. Danielsson^{1,2}, J. Mattisson^{1,2}, H. Davies^{1,2}, B. Torhab Moghadam^{1,2}, J.P. Duman-ski^{1,2}, L.A. Forsberg^{1,2}.* 1) Immunology, Genetics and Pathology, IGP, Uppsala, Sweden; 2) Science for Life Laboratory, Uppsala University.

It is well established that one of the most common post-zygotic mutations in men is loss of the Y chromosome (LOY) in blood. Recently several studies have shown that men with LOY have a higher occurrence of non-hematological cancers and Alzheimer's disease than what would have been expected by chance. There are currently no known genetic mechanism explaining this, however, it is reasonable to hypothesise that LOY has an effect on the immune system. In order to identify cell-types affected by LOY and its impact on gene expression we use single-cell RNA sequencing (scRNAseq) of the PBMCs from affected men. In addition to this we use the Illumina Infinium QC Array on DNA from FACS sorted cells from the same men. We use clustering techniques to assign cell-type to each sequenced cell and cells with LOY are identified using the occurrence of RNA transcripts from the Y chromosome. We have currently sequenced ~30 000 cells and identified several genes whose expression are affected by LOY. Amongst these genes both autosomal genes and genes in the pseudoautosomal region are present. The fraction of LOY calculated in our single cell analysis showed a good correlation with the results from DNA arrays from the same men. These results support that LOY in blood has an impact on the gene expression of PBMCs.

1538F

Statistical modeling and differential gene expression analysis based on single cell RNA-seq data. *I. Mukhopadhyay, P.K. Mondal.* Human Genetics Unit, Indian Statistical Institute, Kolkata, West Bengal, India, PIN 700108.

Single Cell RNA-seq (scRNA-seq) has emerged as a promising method to measure dynamics of gene regulation at cell level. Bulk RNA-seq measures average gene expression level over cells but introduces Simpson's paradox that results in qualitative difference in trends between single cell level and bulk transcriptome data. Large amount of scRNA-seq data generated from recent high-throughput technologies encounter inappropriate modelling in differentiating gene expression, in two types of cells from two different experimental conditions. Although many studies use bulk RNA-seq analysis methods for scRNA-seq data, they do not address the special characteristics and problems of analysing scRNA-seq data that shows higher variability than its corresponding bulk data. Moreover, there are large outliers and drop-out events resulting from failure in proper amplification and low amount of initial number of transcript. Further, transcriptional bursting that causes on and off switching in gene expression level in individual cells leads to its bimodality and complexity of the analysis. The probability of drop-out event in an individual cell depends on the absolute value of the expression level and the quality of library for that particular cell. These events are considerably higher in certain cells indicating that different cells have different effects on drop-out probabilities. Also transcriptional profile of a gene depends on ongoing cellular processes and interaction with other cells. No method developed for single cell differential expression is known to address all these issues simultaneously. Taking all these into consideration, we develop a novel statistical method using maximum likelihood approach that also removes subpopulation effect in differential expression analysis. We use latent variables with random effects for each cell that makes the model more robust and realistic in capturing the two modes in the data. We also estimate the probability of drop-out events in two different peaks. We tackle the computation efficiently, which is very challenging due to lack of any compact form of the estimators. Real data analysis based on our method could identify the two modes, wherever present, and model the gene expression, validating our model with high accuracy. We have developed robust and powerful statistical tests to differentiate scRNA-seq data under two different conditions. Extensive simulations confirm our claim.

1539W

Integrative differential expression and gene set enrichment analysis in single cell RNAseq studies. S. Sun, X. Zhou. University of Michigan, Ann Arbor, MI.

Single cell RNA sequencing (scRNAseq) has been widely applied for transcriptomics analysis. One important analytic task in scRNAseq is to identify genes that are differentially expression (DE) between different cell types or cellular states and to perform subsequent gene set enrichment analysis (GSEA) to detect biological pathways that are enriched in the identified DE genes. These two types of analytic tasks -- DE analysis and GSEA -- are often treated as two sequential steps in commonly used analytic pipelines. However, these two tasks are intermingled with each other: while DE results are indispensable for detecting enriched gene sets and pathways, the detected enriched gene sets and pathways also contain invaluable information that can in turn improve the power of DE analysis. Therefore, integrating GSEA and DE analysis into a joint statistical framework can potentially improve the power of both. Here, we develop such a computational method to integrate GSEA and DE analysis. Our method is based on the Bayesian sparse linear mixed model that is commonly used for association analysis, automatically accounts for expression correlation across genes, and with a newly developed computationally efficient algorithm, adapting it to performing joint and scalable GSEA and DE analysis in single cell RNAseq studies. With simulations, we show that, by integrating GSEA with DE, our method dramatically improves the power of DE analysis and the accuracy of GSEA over commonly used existing approaches. We also illustrate the benefits of our new method with applications to four published scRNAseq data sets.

1540T

Genetic regulation across 1,500 human induced pluripotent stem cell lines. C. Smail¹, M.J. Bonder², D. Jakubosky³, L. Frésard⁴, C. Brown⁴, E.N. Smith², K.A. Frazer², O. Stegle², S.B. Montgomery¹, i2QTL Consortium. 1) Stanford University, Stanford, CA, United States; 2) EMBL-EBI, Hinxton, United Kingdom; 3) UCSD, San Diego, CA, United States; 4) University of Pennsylvania, Philadelphia, PA, United States.

Human induced pluripotent stem cells (iPSC) are a powerful system to assay the molecular impact of genetic variants. Recent studies have shown that iPSCs provide broader resolution of genetic effects on gene expression than differentiated cell-types thereby offering new potential to study variants involved in complex and rare diseases in unique cellular and developmental contexts. To facilitate this effort, we have assembled the largest resource of iPSC genetic and transcriptomic data described to-date - the i2QTL resource - comprising genetic and RNA-sequencing data from 1,001 individuals (comprising 1,515 cell lines) with whole genome sequencing for 478 individuals. We find eQTLs in 14,272 genes (60% of tested genes), including 10,510 protein coding genes (FDR<5%). The dramatic increase in power compared to previous studies further reveals 603 trans-eQTLs, including 177 cis-eQTLs with trans-acting regulatory effects and 155 disease variants previously identified through GWAS. By comparison to GTEx eQTLs, we discovered 68% of all cis-eGenes in GTEx are discovered in i2QTL and we increased the number of eQTL/GWAS overlaps by 4%. From the set of known protein coding non-eGenes in GTEx, 208 could be tested in i2QTL; of these, 101 genes (49%) were found to be eGenes in i2QTL. This set of genes with iPSC-specific eQTL is enriched for developmental processes, with 30/101 (30%) genes annotated with a GO term under this category. For example, iPSC-specific eQTLs were identified for *CRX*, which causes Leber Congenital Amaurosis and is characterized by severe retinal dystrophy, as well as *PRDM13* (neurogenesis), *FOXB1* (somitogenesis), and *FOXB2* (anatomical structure morphogenesis), which are all involved in early embryo development. In addition, the combination of RNA-sequencing and WGS in our sample allowed us to identify rare variants underlying expression outliers and compare the relative discovery power between iPSCs and differentiated tissues. Interpretation of rare variants in iPSCs provides a unique opportunity to study their impacts in a developmental cellular context. For example, we observed that for OMIM disease gene sets, we find expression in 90% of OMIM genes overall, and 95% in OMIM neurology genes in iPSCs. Together, the i2QTL resource is unique in its large sample size and in its extensive data, including WGS, allowing for the creation of a comprehensive map of common and rare regulatory variants in a pluripotent cell type.

1541F**Machine learning and network topological approaches to predict the pathogenicity of loss of function mutations.**

D.J. Rhodes¹, P. Cacheiro¹, O. Makinde¹, D. Smedley¹, M. Caulfield¹, D. MacArthur², D. van Heel³, M.R. Barnes¹. 1) William Harvey Research Institute, Queen Mary University of London, London, United Kingdom; 2) Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA; 3) Blizard Institute, Queen Mary University of London, London, United Kingdom.

Loss of function (LoF) mutations result in the partial or complete inactivation of gene products and are frequently implicated in disease. A large number of LoF mutations resulting in no overt deleterious phenotype have been identified through the study of Scandinavian populations and highly consanguineous populations of Pakistani ancestry. These benign LoF mutations are potentially attractive drug targets based on preliminary analysis that highlighted increased development success of genes with known benign LoF mutations. We hypothesise that gene non-essentiality may correlate with more favourable drug target side effect profiles. Resources such as the Genome Aggregation Database (gnomAD) are comprehensively collating data from exome and genome sequencing projects, including LoF mutations, but the observed or predicted pathogenicity of LoF mutations have not yet been systematically examined. To this end, we have curated a comprehensive snapshot of homozygous LoF variants in 6,902 genes, and used this to identify features associated with LoF pathogenicity. Based on data from human, animal model and in vitro studies, 1720 benign LoF, and 1800 deleterious LoF mutations have been identified, with the remaining 3382 classified as undetermined. From this a protein-protein interaction network has been created using the database StringDB. Topological features of the network show clear differentiation between benign and deleterious LoF mutations, with benign LoF genes exhibiting a lower degree of connectivity (median 10 vs 28), reduced centrality and betweenness. Benign LoF genes are also expressed in fewer tissues (median 22 vs 52) in GTEx data. When compared to orthologous mouse gene essentiality data from the International Mouse Phenotype Consortium, only 8% of genes containing putative benign human LoF mutations were found to be essential in mice, consistent with a conservation of core network topology across species. Conversely, a large proportion (41%) of human deleterious LoF genes resulted in viable phenotypes in mice. We are combining these data with other features such as expression data in order to apply machine learning methods to predict LoF pathogenicity, using the python TPOT framework enabling naive feature selection and construction, and model selection to more accurately predict LoF pathogenicity.

1542W**Genome-wide discovery of large repeat expansions in whole-genome sequence data.** *M.A. Eberle¹, E. Dolzhenko¹, D.R. Bentley².* 1) Illumina, San Diego, CA; 2) Illumina UK, Little Chesterford, UK.

Expansions of short tandem repeats (STRs) are responsible for over 30 rare Mendelian disorders and undoubtedly many more pathogenic repeat expansions are yet to be discovered. Detection of repeat expansions is a challenging problem that was only demonstrated to be feasible for high throughput whole-genome sequence data very recently. Existing methods for identifying repeat expansions are either targeted or require realignment of reads to a special reference which places constraints on the nucleotide composition of detectable repeats. A comprehensive search for novel repeat expansions requires an efficient method that (1) does not rely on a pre-annotated set of repeat loci and (2) places as few constraints on the composition of the repeat sequence as possible. We developed a software tool called Expansion Hunter DeNovo that identifies repeat expansions by locating reads produced by stretches of tandem repeats longer than the read length. Compared to other methods for repeat expansion discovery, Expansion Hunter DeNovo is not limited to short repeat units (i.e. less than 6bp) and can reliably identify repeat units up to 20 bp. To demonstrate that this tool can enable discovery of novel pathogenic repeat expansions, we applied it to samples harboring expansions that cause fragile X syndrome (FXS), Friedreich Ataxia (FRDA), Myotonic Dystrophy type 1 (DM1), Huntington's disease (HD) and 150 controls. Using a straightforward case-control analysis we were able to correctly rediscover both the locus and the repeat unit for each of these repeat expansions. Lastly, we identified eight control samples with evidence of long repeats that were not observed in any of the other controls. We estimate that the expanded repeats in these eight samples range between ~670 and ~4,870bp in length. In particular, one of these repeats is consistent with an expansion that causes Myotonic Dystrophy type 2 (DM2). This finding is currently being validated by an external lab. Combined, these results demonstrate that Expansion Hunter DeNovo can facilitate discovery of novel repeat expansions and, more generally, can be used to characterize landscape of long repeats present in a collection of samples.

1543T

Variant calling from RNA-seq data for Mendelian disorders needs both reference genome and reference transcriptome. A. Jazayeri^{1,2}, S. Pajouhanfar¹, L. Youssefian^{1,3}, A.H. Saeidian¹, S. Sotoudeh³, S. Zeinali⁴, H. Vahid-nezhad^{1,4}, J. Uitto¹. 1) Thomas Jefferson University, Philadelphia, PA, USA; 2) Drexel University, Philadelphia, PA; 3) Tehran University of Medical Sciences, Tehran, Iran; 4) Pasteur Institute of Iran, Tehran, Iran.

RNA-Seq is considered as a standard approach for expression analysis, differential expression quantification, and transcriptional isoform discovery. Data analysis includes read mapping, transcriptome reconstruction, and expression quantification. For each step, numerous tools have been developed to cope with the complexities of this process. Although genome/exome sequencing are common in diagnosing Mendelian diseases, more than half of all patients remain undiagnosed by these methods. Recently, variant calling from RNA-Seq data has shown to bridge this gap to some extent. In this study, a next generation sequencing panel composed of 21 genes for a patient diagnosed as epidermolysis bullosa (EB) was developed. The variant calling was performed and two homozygous mutations were detected as disease-causing: a homozygous nonsense mutation c.5422C>T, p.R1808* in *EXPH5*, and a homozygous c.202delA, p.T68Lfs*106 in *COL17A1*. The pathogenicity of both variants was confirmed with different laboratory methods. The latter mutation is located on the coordinate of the last base of an exon of *COL17A1* creating a splice junction variant. Therefore, RNA-Seq was performed to identify the consequences of this mutation. As a parallel effort, it was tried to detect these mutations from RNA-Seq reads mapped to the reference genome. It was revealed that due to splicing right before the second mutation, the most common splice-aware mapping tools were unable in mapping this mutation, as the remaining part of the sequence reads are mapped to the adjacent exon, and this base is considered as an unread nucleotide. However, when the WGS/WES mapping tools were used with reference transcriptome instead of reference genome the variant was successfully detected. On the other hand, using the reference transcriptome for mapping reads resulted in mapping quality equal to zero in the region of mutation in *EXPH5*, meaning that the corresponding reads can be aligned to multiple locations equally well. This implies that this mutation either is ignored by variant callers because of the low mapping quality or the false positive rate increases. Using both (or an optimized combination of) the reference genome and transcriptome seems to be the ideal approach for variant calling from RNA-Seq data, supported by the case discussed in this study showing that using either of them alone can result in missing one of the disease-causing mutations.

1544F

Sensitive detection of low-allele fraction structural variants in clinical cancer samples. E. Lam, A. Pang, T. Anantharaman, J. Wang, T. Wang, X. Zhang, J. Lee, H. Sadowski, A. Hastie, M. Borodkin. Bionano Genomics, San Diego, CA.

Tumors are often comprised of heterogeneous populations of cells, with certain cancer-driving mutations at low allele fractions in early stages of cancer development. Effective detection of such variants is critical for diagnosis and targeted treatment. Short reads from next-generation sequencing platform provide fragmented information and limited power to detect complex structural variants (SVs). Based on specific labeling and mapping of ultra-high molecular weight (UHMW) DNA, Bionano's single-molecule platform has the potential to detect disease-relevant SVs and give a high-resolution view of tumor heterogeneity. We have developed a pipeline that effectively detects structural variants at low allele fractions. The pipeline uses and merges information from a single-molecule analysis and a *de novo* assembly to detect candidate variants. The single-molecule analysis can be done in mere hours; it includes both single-molecule based SV calling and fractional copy number analysis. The *de novo* assembly, while more computationally intensive, allows for more complete characterization of complex rearrangements. The candidate variants are then annotated and further prioritized based on control data and publically available annotations. Preliminary analyses using simulated and well-characterized cancer samples showed high sensitivity for variants of different types at as low as 5% allele fractions. The streamlined pipeline is fully compatible with the new Direct Label and Stain (DLS) chemistry and can be run on cluster environments and on the cloud.

1545W

Creating a graph-based Arab reference genome using whole read overlap assembly for accurate SV genotyping. Y. Mokrab¹, A.C. English², L. Herta², E. Aliyev¹, J. Bruestle², S.N. Shekar², K.A. Fakhro¹. 1) SIDRA Medical and Research Center, Doha, Ar-Rayyan, Qatar; 2) Spiral Genetics, Inc., Seattle, WA, USA.

Recent exploration of ancestral populations' genomes has shown that the widely used Human reference genome (GRCh37/GRCh38) lacks the diversity needed to represent populations of specific ethnicities (Fakhro et. al. 2016. Hum Genome Var. 3:16016). By incorporating Single Nucleotide Variants (SNVs) and indels found by alignment-based variant calling into GRCh37, a population-representative reference can be created which increases the proportion of alignable reads from an individual of the same population. Incorporating structural variants (SVs) would further improve alignment. To more thoroughly capture structural variation in a sequenced individual would in principle require aligning their reads to a continually updated reference graph of all previously called variation. To avoid introducing erroneous variation, such a graph must be made up of accurate breakpoints. Currently, the detection of SVs from short-read data using inference methods can lead to inaccurate breakpoints, unresolved insertion sequences and high false discovery rates (English AC, et. al. 2015. BMC Genomics. 16:286). Using a rapidly assemblable read graph-genome representation, BioGraph has been shown to detect SVs with high sensitivity, low false discovery rate and exact sequence of breakpoints and insertions (Parikh H et al 2016. BMC Genomics. 17:64). In a study of SVs within a family using BioGraph Assembly (BGA), over 97% of the detected SVs were shown to follow mendelian inheritance (Shekar et al, 2016 unpublished). Here, we observe the accuracy of breakpoints using BGA as an assembly based method compared to the Parliament method of consolidating breakpoints across inference SV callers using a cohort of 100 individuals with Qatari ancestry with the aim of further enhancing the Arab reference genome (Fakhro et. al. 2016. Hum Genome Var. 3:16016) with SVs having a high population frequency. We use Truvari to first identify those variants that are common between the Parliament calls and the BioGraph calls. For these common calls, we observe that a greater proportion of the BioGraph Assembly calls have evidence in the data set of the sequence spanning the breakpoint (30mers) compared to the Parliament calls. This suggests that called SVs using assembly is more suitable for characterizing structural variation in a population in a variant graph that can aligned to for the purpose of rapid accurate genotyping of SVs.

1546T

A reference haplotype panel for genome-wide imputation of short tandem repeat variants. S. Saini^{1,2}, M. Gymrek^{1,2}. 1) Computer Science and Engineering, University of California, San Diego, La Jolla, CA; 2) School of Medicine, University of California, San Diego, La Jolla, CA.

Short Tandem Repeats (STRs) are involved in dozens of Mendelian disorders and have been implicated in a variety of complex traits in humans. Available STR datasets are limited to at most a few thousand samples for a given trait since STR genotyping currently requires either expensive next-generation sequencing data or low-throughput capillary electrophoresis techniques. On the other hand SNP genotype array data is available for hundreds of thousands of samples. A framework to impute STRs into existing SNP array datasets would enable testing STR associations across large sample sizes across hundreds of traits. However imputing STRs is challenging due to complex relationship between STRs and SNP haplotypes induced by rapid STR mutation rates and the high rate of recurrent mutation. Here, we leverage 500 quad families sequenced to 30X coverage as part of the Simons Simplex Collection (SSC), to genotype SNPs and STRs and phase STRs onto SNP haplotypes. We validated the accuracy of our calls and phasing results using Sanger sequencing and evaluated imputed haplotypes by comparing to fosmid and long-read data for the NA12878 genome. Imputation using our panel achieved an average of 97% concordance between true and imputed STR genotypes in an external dataset, with most errors concentrated in STR loci with highest heterozygosity. Using simulated data, we show that imputed STRs capture 20% more variation in STR length compared to individual SNPs, greatly increasing the power to detect associations due to underlying STRs. Finally, we demonstrate using GTEx data that imputed STRs can identify novel signals not revealed by SNP analysis alone. We have released a phased SNP-STR haplotype panel for 1000 Genomes Project samples as a community resource to facilitate large scale studies of STR associations in existing SNP datasets. Our framework will enable testing for STR associations with hundreds of traits across massive sample sizes without the need to generate additional data.

1547F

Real-time genomic analysis with IOBIO tools. A. Ward^{1,2}, M. Velinder¹, A. Ekawade¹, C. Miller^{1,2}, T. Di Sera¹, G. Marth^{1,2}. 1) Department of Human Genetics, University of Utah, Salt Lake City, UT; 2) Frameshift Genomics, 129 Newbury Street, 2nd floor, Boston, MA.

Diagnostic exome/genome sequencing is becoming increasingly common, and ultimately, will become a standard clinical tool. Typically, bioinformaticians undertake genomic analysis, since it requires UNIX command line tools and computational expertise. These bioinformaticians, however, do not generally possess the clinician's experience, familiarity with the patient's family history, disease presentation and course that is often critical for arriving at accurate diagnosis. We are developing powerful, intuitive web tools to enable clinicians to analyze their own patients' genomic data at the point of care. This system is built on our IOBIO suite of web tools, designed to engage as many users as possible, including clinicians, in the analysis and interpretation of genomic data. Our tools allow users to rapidly and easily evaluate sequencing data and variant call quality, without requiring an understanding of opaque genomic data file formats. Users can generate lists of genes associated with the disorder or phenotype under study, and visually examine and prioritize candidate variants using multiple annotations and predictions (e.g. ClinVar, population allele frequencies, SIFT, PolyPhen etc.). All these tasks can be carried out in real-time from a standard web browser by accessing and analyzing only the necessary portions of the underlying data, i.e. genes likely to harbor diagnostic variants. A team of clinicians from multiple disciplines (e.g. diagnostic pathologists, subspecialty pediatricians, medical geneticists) are involved in refining these tools, and adapting them for deployment at the point of care. Utilization of these tools on undiagnosed clinic disease cases has led to the identification of previously unrealized quality problems (requiring regeneration of raw sequencing data), as well as positive solutions for previously intractable cases.

1548W

Robust detection of differential splicing and transcription events for RNA-seq data. J. Hu, Y. Hu, M. Li. University of Pennsylvania, Philadelphia, PA.

RNA sequencing offers unprecedented opportunities to study alternative splicing and transcription, a major mechanism for transcriptome diversity. An important application of RNA-seq is to identify genes that are differentially spliced or transcribed under different biological conditions, as these genes may reflect information on disease susceptibility. Most existing methods for differential splicing/transcription analysis only use junction reads and suffer from power loss for genes with many isoforms. To incorporate all mapped reads, we previously developed PennDiff (Hu et al. 2018 Bioinformatics), a statistical method that makes use of information on gene structures and pre-estimated isoform relative abundances. Although PennDiff has shown promising performance, it does not explicitly model estimation uncertainty of exon-inclusion level, a quantity utilized in downstream differential analysis. Empirical studies indicate that the dispersion of exon-inclusion levels is reversely correlated with the abundances of those transcripts that contain the target exons, suggesting that modeling the genome-wide distribution of exon-inclusion levels may improve the performance of PennDiff. Motivated by this, we develop PennDiff2, a successor to our PennDiff method. PennDiff2 uses shrinkage estimators for dispersion and exon-inclusion level difference to improve accuracy of the estimates and interpretability of the tests. Specifically, smooth curves are fitted by regressing the exon-wise dispersion estimates and exon-inclusion level difference estimates onto the abundance of related transcripts across all exons and all samples. Then, PennDiff2 employs an Empirical Bayes procedure to shrink these estimates. Hypothesis testing for exon-based differential splicing/transcription analysis is performed based on the shrinkage estimates. PennDiff2 is also able to detect gene-based differential splicing/transcription by combining all exons within a gene using a variance-component score test. The performance of PennDiff2 is evaluated by both simulations and analysis of RNA sequencing data from the Genetics of Evoked Response to Niacin and Endotoxemia Study.

1549T

Mutation maps as a tool for understanding deep learning models in biology: Branchpoint prediction as an example. *R. Srinivasan, S. Gupte.* Innovation Labs Hyderabad, Tata Consultancy Services Ltd., Hyderabad, TS, India.

Background: Deep Learning (DL) models have been shown to match or exceed traditional Machine Learning (ML) approaches in several prediction tasks in computational biology. However, DL models which learn features automatically, unlike ML derived models which build upon expert provided features, do not afford easy interpretation of factors that are important for successful prediction. To address this issue we developed a method called mutation maps (MutMap) that is useful in instances where the input to the DL method is a set of biological sequences (DNA/RNA/Protein). We apply DL to the problem of predicting the location of a branchpoint in an intron and show that MutMap can reveal crucial factors learned by the DL model in achieving successful predictions. **Methodology:** To illustrate our method we took up the task of predicting the location of the splicing branchpoint in a pre-mRNA sequence. We used a branchpoint dataset published by Mercer et al to create a DL model that is able to identify branchpoints using only the sequence around the branchpoint as input. The DL model used a combination of convolutional and recurrent layers to extract both local and global features automatically. We then created new sequences by mutating every nucleotide position in each of the input sequences to the other 3 possible nucleotides (e.g. A>G, A>C and A>T). These sequences were used to generate a map of nucleotide mutation vs effect of the mutation as predicted by our DL model. **Results:** The DL model we trained compared favorably to other ML based approaches such as Branchpointer and SVM-BPfinder. We obtained a higher F-Score (DL 0.706; Branchpointer 0.673; SVM-BPfinder 0.555) and a higher Area under Precision-Recall Curve (0.748, 0.708 and 0.537 respectively). The mutation maps generated by our model showed that the TNA motif is an important factor in branchpoint prediction. Furthermore, the maps also identify a Poly-pyrimidine tract following the branchpoint as a critical feature. Mutations which disrupt the splice site or create new AGs are also shown to have an impact. **Conclusion:** We describe a general method for extracting features learned by DL models through the creation of mutation maps. These maps help visualize important sequence regions as illustrated by the branchpoint prediction study. This is a promising approach that could be used in other prediction problems in genomics to discover novel insights.

1550F

Integrating predicted transcriptome from multiple tissues improves association detection. *A.N. Barbeira¹, M.D. Pividori², H.E. Wheeler^{2,3}, D.L. Nicolae¹, H.K. Im¹.* 1) Section of Genetic Medicine, Department of Medicine, The University of Chicago, Chicago, IL, 60637; 2) Department of Biology, Loyola University Chicago, Chicago, IL, 60660; 3) Department of Computer Science, Loyola University Chicago, Chicago, IL, 60660.

Integration of genome-wide association studies (GWAS) and expression quantitative trait loci (eQTL) studies is needed to improve our understanding of the biological mechanisms underlying GWAS associations, and our ability to identify therapeutic targets. Gene-level association test methods such as PrediXcan can prioritize candidate targets. However, limited eQTL sample sizes and absence of relevant developmental and disease context restrict our ability to detect associations. Here we propose an efficient statistical method that leverages the substantial sharing of eQTLs across tissues and contexts to improve our ability to identify potential target genes: MultiXcan. MultiXcan integrates evidence across multiple panels while taking into account their correlation. We apply our method to a broad set of complex traits available from the UK Biobank and show that we can detect a larger set of significantly associated genes than using each panel separately. To improve applicability, we developed an extension to work on summary statistics: S-MultiXcan, which we show to be concordant with the individual level version. Results from our analysis as well as software and necessary resources to apply our method are publicly available.

1551W

Detecting structural variants from Oxford Nanopore sequencing using signal-level data. *D. Beyter¹, H. Ingimundardóttir¹, H.P. Eggertsson¹, S. Mehlinger², K. Stefánsson¹, B.V. Halldórsson¹.* 1) deCODE Genetics/Amgen, Inc., Reykjavik, Iceland; 2) Algorithmic Bioinformatics (ABI), Freie Universität Berlin, Berlin, Germany.

Structural variants (SVs) are a major source of genetic variation. Long-read, single molecule sequencing presents an opportunity to greatly advance the detection of SVs. However, the high base calling error rate of this new technology makes it challenging to achieve accurate results that can be used for downstream analysis, such as phenotype associations. We present an SV detection pipeline, using signal-level raw nanopore data (squiggles) to filter a highly permissive initial SV call set. We also use Illumina sequencing data for breakpoint refinement wherever applicable. We tested our pipeline on a set of whole-genome sequenced individuals. Trio analysis suggests high accuracy levels, judging by the similarity of SVs in the proband and one or both of its parents.

1552T

CtsCNV: A copy number variant detection method for clinical targeted sequencing data. *Y. Cai, R. Shraga, M. Akana, B. Patel, M. Jansen, M. Jar-emko, O. Puig.* Phosphorus Inc., 1140 Broadway New York, NY.

Background: While several methods exist to detect Copy Number Variants (CNVs) in targeted sequencing data, they either fail to achieve the accuracy required in a clinical diagnostic setting, or are limited to detection of CNVs within specific genes. The problem of detecting CNVs is challenging because of uneven coverage across targets, bias in coverage based on GC content and target proximity, sample and batch level bias due to small changes in laboratory conditions, as well as small batch size and low read depths often present in clinical genetic testing. **Methods:** We have developed a ctsCNV caller, which computes average depth per sample per interval and includes several normalization procedures to remove variation in depth due to non-biological noise and can be readily used in clinical targeted sequencing (CTS) data. This procedure corrects for sample variability, batch effects, bias in GC content in the sequences, and other technical biases. After normalization (spline normalization, z score, PCA), copy number estimates per interval are computed by comparing each sample's normalized depth per interval to the median normalized depth within a batch. CNVs are called by running the widely used Circular Binary Segmentation algorithm on the interval estimates along with a permutation based significance test. **Results:** Initially, using a custom Roche NimbleGen targeted panel and the Illumina HiSeq 3000, 372 previously characterized DNA samples were sequenced at different depths (100-600X). These samples included 81 known CNVs sized between 200bp to 700Kbp. The performance of ctsCNV was compared to several commonly used callers including Codex,XHMM, Excavator2, and ConVadng. CtsCNV achieved the highest sensitivity of 98.77%. To ensure the analysis was not biased to internal lab conditions, all callers were also tested on 90 lower depth WES samples available from the 1000 Genomes project. CtsCNV achieved the highest sensitivity, specificity, and precision as well. Now, we are able to accurately call CNVs ranging from ~300bp (median depth 1000X) to ~4Mbp Y chromosome micro-deletion (median depth 250X) in sequencing batches with as low as four samples. Our results demonstrate that ctsCNV achieves accurate CNV detection on targeted sequencing data, and it is suitable for clinical diagnosis. In comparison to other callers, ctsCNV represents considerable improvement to the current state of the art. Upon publication, ctsCNV will be freely available to download.

1553F

Learning and interpreting gene regulatory architectures in a deep neural network framework. *L. Chen¹, J. Capra^{1,2}*. 1) Department of Biological Sciences, Vanderbilt University, Nashville TN, USA; 2) Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN, USA.

Determining the effects of non-coding mutations on phenotypes remains challenging due to the complexity of mammalian gene regulatory elements. There are several models of how combinatorial binding of transcription factors (TFs) drives enhancer activity, including the billboard, TF collective, and enhanceosome models. However, we have limited knowledge of the prevalence of these architectures across enhancers. Deep convolutional neural networks (CNNs) have achieved state-of-the-art performance in predicting gene regulatory sequences, but they have provided limited insight into the regulatory architecture due to the difficulty of interpreting the biological meaning of the complex patterns they learn. To explore the ability of CNNs to capture the regulatory architecture of enhancers, we performed several hypothesis-driven simulations. First, we trained CNNs to distinguish simulated DNA sequences with different regulatory architectures embedded in them from two negative sets: (i) random genomic sequences and (ii) sequences with equal number of randomly distributed TFs. The CNNs could accurately distinguish positives from both negative sets (auROC >0.99). Next, we used existing and novel algorithms, including maximally activating patches, neuron clustering, gradient ascent, and saliency maps, to interpret patterns learned by the neurons in CNNs. Models trained in both tasks consistently learned TF motifs in their first convolutional layer. The neurons in the internal layers of the CNNs trained against random sequences had similar activation in the presence of different TFs, suggesting that they simply learn to count occurrences of TFs instead of a specific architecture. The neurons in internal layers of the CNNs trained against TF-matched sequences showed different activation by different regulatory architectures, suggesting that they capture complex patterns. Based on these successes, we are using this approach to quantify the architecture of real enhancers. Our results highlight challenges of extracting insights from CNNs; even a highly accurate model may not learn biologically relevant patterns. The ability of CNNs to capture combinatorial patterns is dependent on the training data and is improved by designing a negative set to match relevant attributes of the positives. Our regulatory architecture simulation and visualization framework facilitate the use of deep learning models to reveal the complex combinatorial patterns in regulatory sequences.

1554W

A high-quality benchmark dataset of SV calls from multiple technologies.

S. Chen¹, M. Bekritsky², P. Krushe², F. Sedlazeck³, E. Dolzhenko¹, B. Moore², D. Bentley², M. Eberle¹. 1) Illumina Inc, San Diego, CA; 2) Illumina Cambridge Ltd., Little Chesterford, UK; 3) Baylor College of Medicine, Houston, TX.

Structural variants (SVs) represent a significant proportion of variability between human genomes and are overrepresented in genetic disease. While there are a variety of software tools that can identify and genotype SVs, the performance of these tools is not well quantified. Further development and optimization of these methods will require a comprehensive genome-wide catalog of high-confidence variants resolved to base pair accuracy. For this purpose, we have developed a validation pipeline that re-genotypes SVs by aligning reads to a variant-aware graph where haplotypes are represented as paths. By aggregating reads across many samples and incorporating small mismatches and gaps in the graph structure, we are able to refine the breakpoints of some input SVs. To QC variants, we test Mendelian consistency in the 17-member Platinum Genome pedigree as well as 50 trio families and also test for Hardy Weinberg Equilibrium (HWE) in 1,000 unrelated individuals of European origin. As a demonstration of our validation pipeline, we assessed ~20,000 SV calls made by Sniffles from PacBio data and ~18,000 SV calls made by Manta from Illumina NovaSeq sequence data, ~6,000 of which overlap. These include deletions, insertions, inversions, and translocations. Because the same variant may be represented differently by the two callers (Sniffles or Manta), we analyzed the two call sets independently and merged the passing variants. Ultimately, ~9,500 merged SVs passed our validation pipeline, including ~1,800 that were identified solely by Sniffles and ~5,000 that were identified solely by Manta. In further analyses of SVs that failed our validation, we determined that more than half of the original calls originated within tandem repeats and would require a different validation method. Additionally, a large proportion of the failed SV calls likely represented true variants with poorly resolved breakpoints or the incorrect/erroneous insertion sequences. We are continuing to refine these variants to breakpoint accuracy as part of our work developing a high quality SV truth set. We believe this truth set and validation methods are a valuable community resource for further development of SV calling methods for clinical genomics research.

1555T**Binomial probabilistic principal component analysis for genotype data.**

A.M. Chiu¹, B. Pasaniuc^{1,2,3}, S. Sankararaman^{1,3,4}. 1) Bioinformatics Interdepartmental Program, University of California Los Angeles, Los Angeles, CA USA; 2) Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA USA; 3) Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA USA; 4) Department of Computer Science, University of California Los Angeles, Los Angeles, CA USA.

Principal Component Analysis (PCA) is one of the most commonly used techniques for inferring population structure as well as for visualizing and interpreting genotype data. However, classical PCA lacks a probabilistic generative model. Lack of a probabilistic generative model makes it challenging to apply PCA in the presence of missing data as well as in quantifying the uncertainty of the inferred principal components. While probabilistic Principal Components Analysis (PPCA) has been proposed to overcome the limitation of PCA, PPCA makes the restrictive assumption that the observed data follows a multivariate Gaussian distribution. This assumption is not appropriate for genotype data since each entry of the genotype vector takes on only one of three possible values (zero, one or two). We propose a novel binomial PPCA model that accounts for the discrete nature of genotype data. By modeling the observed data through a binomial distribution, we are able to define a generative model that better represents genotype data. Through the use of variational inference, we derive an EM algorithm to infer the principal components in our model. The binomial PPCA can further be extended to other genomic datatypes that have different characteristics (e.g. count data from RNA-Seq).

1556F**Benchmarking of StrandNGS expression profiling pipeline.** D. Chohan, S. Ramachandrala, H. Adil, A. Pal, S. Cherukuri, M. Nagarajan. Strand Life Sciences Private Limited, Bengaluru, India.

StrandNGS, our flagship next generation sequencing analysis and data management tool has been providing extensive support for the analysis of RNA-Seq data. We support end-to-end analysis as well as facilitate the flexible import of unaligned, aligned, quantified, and processed data. Once the data is imported into StrandNGS, all downstream analysis options could be applied. Here, using GSE47774, the SEQC dataset consisting of samples with known mixing ratios (A & B were pure samples, C & D were derived by mixing A & B samples in 3:1 & 1:3 ratios respectively), we have benchmarked 4 expression profiling pipelines that are readily available in StrandNGS: 1st pipeline comprises of the import of unaligned data, alignment using COBWeb, the in-built aligner in StrandNGS, filtering, quantification, and statistical analysis. 2nd pipeline comprises of the import of unaligned data, trimming the reads to a length of 75 bp, alignment using COBWeb, filtering, quantification, and statistical analysis. 'Trimmed' expression pipeline saves time in alignment without affecting the results. 3rd pipeline comprises of the import of unaligned data, using HISAT for alignment, filtering, quantification, and statistical analysis. 4th pipeline comprises of the import of pre-aligned data (aligned using STAR), filtering, quantification, and statistical analysis. HISAT and STAR have special hardware requirements. HISAT index building requires ~200 GB of memory for human genome+transcriptome. STAR requires ~30 GB of memory to align human genome+transcriptome. Whereas, StrandNGS does not have any such requirements. Also, steps like index building are handled by StrandNGS so that these need not be done manually which isn't the case with other aligners. Based on the following 4 metrics, we conclude that the expression profiles generated using the 4 pipelines have a very high concordance across them: - Number of genes passing the statistical tests - The titration order of each of the genes across samples - A test to assess the A/B mixing ratio recovery in C and D samples - Pearson correlation coefficients of sample-wise expression profiles across pipelines. Given that STAR and HISAT are the commonly used and well-tested aligners for transcriptomics studies, we conclude that COBWeb, our proprietary aligner performs on par with them. Additionally, the trimmed-read alignment pipeline in StrandNGS helps in saving NGS data aligning, processing, and analysis times without affecting the accuracy of results.

1557W

CookHLA: Accurate HLA imputation. S. Cook^{1,2,3}, X. Jia⁴, Y. Luo^{5,6,7}, P. Bakker^{8,9}, S. Raychaudhuri^{5,6,7,10}, J. Rho^{1,2}, B. Han³. 1) Department of Convergence Medicine, University of Ulsan College of Medicine, Seoul, Korea, Republic of; 2) Asan Institute for Life Sciences, Asan Medical Center; 3) Department of Medicine, Seoul National University College of Medicine; 4) Department of Neurology, University of California–San Francisco, San Francisco, CA USA; 5) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, UK; 6) Division of Genetics and Rheumatology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA; 7) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA; 8) Department of Genetics, Center for Molecular Medicine, University Medical Center Utrecht, Utrecht, The Netherlands; 9) Department of Epidemiology, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands; 10) Broad Institute of MIT and Harvard, Cambridge, MA, USA.

Background: Human leukocyte antigen (HLA) genes in the major histocompatibility complex (MHC) region influence many disease phenotypes. It is our utmost interest to investigate as to which HLA alleles or amino acids are driving the diseases. Because HLA genes are highly polymorphic and the whole MHC region exhibits linkage disequilibrium (LD), fine-mapping disease-causing variations often requires HLA typing in tens of thousands of samples. However, classical HLA genotyping is prohibitively costly to be applied to a large-scale cohort. To solve this challenge, recent studies developed computational approaches to impute HLA from intergenic SNP genotype data. These HLA imputation methods have been widely used to fine-map HLA for many immune phenotypes. **Results:** In this study, we present a novel HLA imputation method, CookHLA. CookHLA is superior to previous approaches in accuracy and efficiency. The limitation of the predecessor approach, SNP2HLA, was that the markers defined to indicate HLA alleles were put in a specific (center) position of the gene. Thus, imputation accuracy degraded in distinguishing alleles that differ in the exon sequences far from the center position due to LD decay. To achieve high accuracy, CookHLA makes duplicate sets of markers and embeds each set in each of the highly polymorphic exons. The predictions are then calculated from the consensus posterior probabilities of the duplicate marker sets. To further improve accuracy, CookHLA adaptively learns population-specific and reference panel-specific genetic distance parameters between the markers. In the simulation experiments in which we cross-validate the European reference panel (N=5,225) from the Type 1 Diabetes Genetics Consortium (T1DGC), CookHLA superbly achieved 96.8% average 4-digit accuracy of six HLA genes, where SNP2HLA only achieved 93.9% accuracy, resulting in 2-fold reduction in error rate. A notable accuracy gain occurred at HLA-DRB1 (92.9% vs. 88.5%), which was encouraging because HLA-DRB1 is an important risk factor for many immune diseases. CookHLA is fast and memory-efficient and allows use of large reference panels. Applying the largest-to-date panel (N=10,689) to impute 3,000 individuals takes only 3 hours and 3Gb memory in a personal computer. CookHLA is freely available at <http://software.buhaman.com/COOKHLA/>.

1558T

A balancing act: Aiming for stability in human gene nomenclature.

P. Denny, B. Braschi, K. Gray, T. Jones, R. Seal, S. Tweedie, B. Yates, E. Bruford. HUGO Gene Nomenclature Committee (HGNC), EMBL-EBI, Hinxton, United Kingdom.

The HUGO Gene Nomenclature Committee (HGNC, www.genenames.org) is the only global organisation designating standardized gene symbols and names to all human genes. As precision genetic medicine is increasingly applied in the clinic, it's critical that everyone knows exactly which gene is being discussed. Stable gene identifiers (IDs) underpin these discussions, but are not as human-friendly as the gene symbols and names which are used in everyday communication, but may change. As part of the Transforming Genetic Medicine Initiative (TGMi), we are aiming to minimise the number of gene symbol changes to achieve symbol stability in the long term. What are the characteristics of a "good", stable gene symbol and name? Symbols should be unique, memorable, short and ideally, pronounceable; names should be explanations of a symbol or function and more descriptive and informative. Some of these characteristics are subjective and not easily assessed in an automated fashion. We have chosen four metrics for an initial semi-automated assessment of our gene symbols: 1) is the symbol a placeholder? This class of symbol, e.g. with the root *C#orf#*, *FAM#* or *KIAA#*, is intended to be temporary because it is neither descriptive nor informative; 2) are the HGNC gene name and UniProt protein name similar? It's ideal if proteins are named the same as the genes that encode them; 3) is the gene a member of an HGNC gene family? This often means the family members will share a root symbol; and 4) has a human gene symbol and name been assigned to its chimpanzee ortholog? In our wider role as coordinators of vertebrate gene nomenclature (<https://vertebrate.genenames.org/>), we have curated human gene symbols and names prior to transfer to chimpanzee, based on consensus orthology and conserved synteny. During this process, where possible placeholder symbols have been replaced, e.g. *C2orf71*, '*chromosome 2 open reading frame 71*' > *PCARE*, '*photoreceptor cilium actin regulator*' and human-centric symbols and names changed, e.g. *CECR1*, '*cat eye syndrome chromosome region, candidate 1*' > *ADA2*, '*adenosine deaminase 2*'. We have been applying these four metrics to the ~3000 genes associated with disease in the TGMi Gene-Disease Map and will present our latest findings.

1559F

Prototyping computational tools for human genetics research in NCBI hackathons. A. Dillman^{1,5}, S. La², A. Goncarenko^{1,5}, F.J. Sedlazeck³, D. Timmons⁴, L. Phan^{1,5}, L. Federer⁵, B. Busby^{1,5}. 1) NCBI, Bethesda MD; 2) Simon Fraser University, Vancouver BC; 3) Baylor College of Medicine, Houston TX; 4) University of Colorado, Boulder CO; 5) National Library of Medicine, Bethesda MD.

Over the past three years, NCBI has run or been involved in 25 data science hackathons. In these hackathons, participants assemble into teams of five or six to work collaboratively for three days on pre-scoped projects of general interest to the bioinformatics community. On average, about 80% of teams produce an alpha or beta working prototype, and approximately ten percent ultimately publish a manuscript describing their work. NCBI hackathons have generated over 150 products, some of which continue to be developed. A selection can be found at <https://ncbi-hackathons.github.io>. In addition to the production aspect, hackathons provide an immersive learning environment and promote networking opportunities. We will discuss four functional prototypes for human genetics research developed in these hackathons. The first allows the user to interrogate a number of SNPs from either fastqs at rest, or streaming from public datasets. The second packages this functionality into automated containers, which simply take a list of SNPs and can be distributed to owners of other repositories. The third takes in a list of SNPs and returns both all co-mentioned terms in PubMed abstracts, as well as generating a D3 graph depicting relationships between SNPs and publications. This is particularly useful for looking at a number of GWAS studies in a given field. Finally, we have made DangerTrack, a public blacklist of regions in the human genome where short variant calls may be inaccurate due to a high frequency of structural variants or other genome instability factors. We continue to use and build upon these tools in hackathons, especially as we partner with community patient-centered hackathons..

1560W

Prediction of regulatory SNPs affecting transcription factor binding using RegulomeDB. S. Dong¹, AP. Boyle^{1,2}. 1) Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI; 2) Department of Human Genetics, University of Michigan, Ann Arbor, MI.

Evidence from Genome Wide Association Studies (GWAS) has provided us with insights into human phenotypes by identifying variation statistically associated with disease. However, it is desirable to extend these studies beyond association to an understanding of biological impact. Unfortunately, determining the function of these variants is a major challenge, especially for single-nucleotide polymorphisms (SNPs) in the non-coding regions of the genome where most of these variants fall. To functionally annotate SNPs, computational tools have been developed by using functional genomics data, which are specifically related to regulatory elements and transcription factor (TF) binding. These computational tools provide powerful ways to narrow down from a huge list of candidates to the causative SNPs leading to human disease. One example of such computational tools is RegulomeDB (<http://www.regulomedb.org>). Here, we show a prediction model to improve the current scoring system in RegulomeDB by applying machine learning methods. We generated our training data by calling allele-specific TF binding SNPs from ChIP-seq data. The final model takes advantage of functional genomics data as well as sequence context in order to more accurately predict function. We show the feasibility of this method with training results in a lymphoblastoid cell line (GM12878). Furthermore, we are collecting training data from other cell types, and we expect our new scoring method to predict novel regulatory SNPs disrupting TF binding in a variety of cell types.

1561T

Efficient curation and real-time querying of clinical and genomic data for 500,000 samples. *H. Edgren, H. Garritzen, M. Roccato, S. Suhonen, H. Lodenius.* MediSapiens Ltd, Helsinki, Uusimaa, Finland.

Recent research using the data from projects such as the UK Biobank has emphasized the power of large, well curated and harmonized sets of clinical and genomic data in making groundbreaking discoveries on the genetic bases of health and disease. This approach, however, presents two significant challenges: the need to curate and harmonize large amounts of clinical and phenotypic data, as well as subsequent management and efficient analysis of this together with the genomic information. To address these challenges, we developed Accurate™ and Genius™, cloud-based and highly elastic solutions for efficient curation and real-time analysis of large genomics cohorts. We present results showing the significant time savings achieved with Accurate™ in data cleanup, standardization and ontology encoding of both structured and natural language data. Results include both specific benchmarking of the machine-learning natural language analytics on the MIMIC III data set, as well as of the whole data curation process on a large heterogeneous clinical dataset. We also present methods for and results on achieving efficient and low latency complex queries on clinical and genomic data of over 500.000 samples with Genius™, at low fixed costs. In summary, we here present highly intuitive and efficient solutions for curation and analysis of large sets of clinical and genomic data for biomedical research and healthcare, emphasizing enrichment of data using ontologies, fine grained access control, full audit trail of all curation operations and scalability to millions of samples.

1562F

MegaQC: Continuous monitoring of bioinformatics metrics. *P.A. Ewels¹, D. Moreno^{1,2}.* 1) Department of Biochemistry and Biophysics, Science for Life Laboratory, Stockholm University, Stockholm 106 91, Sweden; 2) Inivata Limited, Granta Park, Cambridge, CB21 6GS, United Kingdom.

Large-scale data analysis is becoming increasingly commonplace in both research and clinical practice. To ensure quality, metrics must be tracked and actively monitored to spot trends and identify outliers. Previously, we developed MultiQC - an open-source package for the aggregation and visualization of results from bioinformatics tools (<http://multiqc.info>, Ewels *et al.* Bioinformatics, 2016). MultiQC can parse output from over 60 different bioinformatics programs, generating a HTML report with summaries and plots for all samples in a project and all tools used in the analysis. This enables fast and accurate evaluation of results. MultiQC has been widely adopted by the bioinformatics community and is now a standard step in analysis pipelines all over the world. However, the reports it creates are generated on a per-project basis, making it difficult to track long-term trends and metrics. Here, we present MegaQC (<http://megaqc.info>) - an interactive web tool to collect parsed output from MultiQC runs over time. Installed on a server, it allows fast and easy monitoring of custom trends and interactive investigation of patterns in data. Dynamic filtering and a drag-and-drop interface makes it easy to quickly build custom dashboards with key metrics. Automated alerts can raise the alarm when samples exceed given thresholds. These make MegaQC a perfect tool for clinical genomics groups and sequencing facilities who need to keep a close eye on data quality. MegaQC can visualise data from any of the programs supported by MultiQC, making it in incredibly flexible; statistics can describe data quality, genome coverage, variant counts or any custom input given to MultiQC. The tools require very little configuration and easily complement existing systems. MegaQC has been developed at the Swedish National Genomics Infrastructure, part of SciLifeLab. It is in use across a number of other clinical facilities and research institutes. It promises to democratise large-scale analysis of bioinformatics results, opening up access to previously unused data.

1563W**Updating and expanding the resources from the 1000 Genomes Project in the International Genome Sample Resource (IGSR).** *S. Fairley, A. Gall, E. Haskell, E. Lowy, B. Moore, E. Perry, L. Clarke, P. Flicek.*

European Molecular Biology Laboratory - European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, CB10 1SD. United Kingdom.

The 1000 Genomes Project generated the largest fully public catalogue of human genetic variation and produced a set of valuable reference resources that remain widely used. Through the International Genome Sample Resource (IGSR), we provide continued access and support to users of the 1000 Genomes data resources, while also building on those resources. Resources from the 1000 Genomes Project have been updated to the current reference genome, GRCh38, with sequence data being aligned to the genome and call sets generated using BCFtools, FreeBayes and GATK's Unified Genotyper, leading to the creation of a combined call set, called directly on GRCh38. Data from the 1000 Genomes Project has been added to, with data from technologies including Oxford Nanopore, PacBio and 10X Genomics being added for some samples from 1000 Genomes. Further, we continue to work to add available high coverage Illumina data generated on 1000 Genomes samples. IGSR is also working to add populations that were not represented in 1000 Genomes. Data from the Gambian Genome Variation Project (GGVP), HG-DP-CEPH collection and the Simons Genome Diversity Project (SGDP) have been added and we continue to work with collaborators to increase the number of populations represented in the resources. As the data sets in IGSR grow, we are developing the project website, which includes a data portal, to aid discoverability, navigation and interpretation of the data sets. Further, we continue to collaborate with Ensembl to assist users in browsing the 1000 Genomes data in genomic context and providing tools which enable users to work with the underlying files. IGSR is open to collaboration with groups generating openly consented genomic data, including providing assistance in ethical review, coordination of data and alignment to the reference genome, with a strong desire to add new populations to the resources.

1564T**Detection of structural variants using linked-reads with novel algorithms.**

I.T. Fiddes, A.M. Barrio, S. Williams, P. Marks, S. Garcia, S. Maheshwari, D.M. Church. 10x Genomics, Pleasanton, CA.

Variation in genomes range from single nucleotide variants (SNVs) to whole chromosomal arm aberrations. Structural variants (SVs) encompass both copy number altering and copy number neutral changes, such as inversions and translocations. Structural variation is increasingly being seen as the molecular explanation for complex phenotypes and disease etiology. Standard short-read approaches struggle to identify SVs, cannot access repetitive regions of the genome and fail to resolve the human genome into its two haplotypes. Incorporating haplotype information has been previously shown to improve SV detection sensitivity, especially for heterozygous variants (Huddleston et al., 2016). We have developed a novel data type known as 'Linked-Reads', constructed by partitioning limiting amounts of high molecular weight DNA and producing barcoded libraries. We also developed a handful of algorithms that leverage barcode information to detect deletions in the 50bp-30kbp range as well as detect very large structural events including translocations, inversions, duplications and deletions. These libraries allow for long range haplotype reconstruction, detecting complex SVs on both Whole Genome (WGS) and Whole Exome (WES). Here we demonstrate the power of Linked-Reads to resolve SVs in control samples as well as samples with known, validated events. We performed 30x Linked-Read genome sequencing on a set of 23 samples obtained from Coriell with known large (>30Kb) balanced or unbalanced SVs. 27 of the 29 known events were detected and another event was called as a candidate. Sequence downsampling was performed on a subset to determine the lowest sequence depth required to detect variations. Copy-number variants can be called with as little as 1-2x sequencing depth (5-10Gb) while balanced events require on the order of 10x coverage for variant calls to be made, although specific signal is clearly present at 1-2x sequencing depth. Additionally, we sequenced the Genome In A Bottle consortium (GIAB) reference sample NA24385 to 84.4x coverage, and compared the results to the GIAB v0.5.0 SV truth set. We detected 4,321 deletion events ≥ 50 bp, with 78% precision and 31% recall. This is comparable to 67% precision and 44% recall seen in the merge of seven different short read callers. Future algorithmic improvements can increase the sensitivity and specificity of Linked-Read based SV calling..

1565F

Whole exome sequencing for exon-level copy-number variation in small-sample clinical settings. D.L. Filer^{1,2}, F. Kuo³, C. Bizon², K. Robasky², J.S. Berg¹, J.L. Tilson², J.P. Evans¹, K.C. Wilhelmsen^{1,4}. 1) Department of Genetics, UNC School of Medicine, Chapel Hill, NC; 2) Renaissance Computing Institute, Chapel Hill, NC; 3) Memorial Sloan Kettering Cancer Center, New York, NY; 4) Department of Neurology, UNC School of Medicine, Chapel Hill, NC.

As whole exome sequencing (WES) integrates into clinical practice, we should make every effort to utilize all information generated. Copy-number variation can lead to Mendelian disorders, but small copy-number variants (CNVs) often get overlooked or are obscured by underpowered data collection. Many groups have developed methodology for detecting copy-number variants from WES, but existing methods perform poorly for small CNVs and rely on large sample numbers not always available to clinical laboratories. Here we present sequencing approaches to remove capture variation and a novel depth-based statistical method for accurate detection of CNVs in small-sample clinical settings. First, using data from the NCGENES project, we demonstrate how pooling samples prior to capture greatly reduces sample-to-sample variation. Second, we show simple simulations based on the multinomial distribution replicate the distribution of observed pooled samples. Third, we perform large-scale simulations demonstrating the ability to discriminate small CNVs as a function of read depth and variant size. For each simulation condition we generated 200 pools of 16 samples with 172,000 exons and 215 expected CNVs per sample. We considered 100 conditions, varying sequencing depth (in millions of mapped reads) from 10 to 200 and variant size from single exons to five exons. Simulation results suggest an average depth of 35 million mapped reads per sample can discriminate single-exon variants with $71.5 \pm 3.6\%$ sensitivity with $3.4 \pm 1.3\%$ incorrectly positive calls (MCC of 0.83 ± 0.02). For two-exon variants, the sensitivity increases to $98.3 \pm 0.3\%$ with $1.3 \pm 0.6\%$ incorrectly positive calls (MCC of 0.98 ± 0.003). Taken together, our results provide a framework for researchers and clinical laboratories to assess CNV calls from WES and methodology to accurately and efficiently utilize WES for CNV detection in small-sample clinical settings.

1566W

Benchmark: An unbiased, genetic-data-driven benchmarking strategy for gene and variant prioritization algorithms using LD score regression. R.S. Fine^{1,2,3}, T.H. Pers^{4,5}, J.N. Hirschhorn^{1,2,3}. 1) Department of Genetics, Harvard Medical School, Boston, MA; 2) Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, MA; 3) Broad Institute of MIT and Harvard, Cambridge, MA; 4) Novo Nordisk Foundation Centre for Basic Metabolic Research, University of Copenhagen, Copenhagen, Denmark; 5) Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark.

Genome-wide association studies (GWAS) have successfully identified numerous loci underpinning hundreds of traits, but determining the causal variants and genes within these loci remains challenging. To address this issue, algorithms to prioritize the most likely causal variants and genes have been developed (for example, prioritizing genes with shared protein-protein interactions or variants with similar profiles of epigenetic marks). However, it is not easy to compare, or "benchmark," the performance of these methods and assess which of them produce the most accurate results. The most common benchmarking approach uses "gold standard" genes (i.e. genes with a known link to the trait of interest), which relies on incomplete information and may punish a method that successfully discovers novel biology. We propose a novel benchmarking method inspired by cross-validation, Benchmarker, which avoids relying on known biology by using the GWAS data itself as its own control. In the Benchmarker strategy, we utilize a "leave-one-chromosome-out" strategy, where we prioritize genes or variants on each chromosome using GWAS data for all other chromosomes. Then, we combine the prioritization results across all 22 chromosomes and apply LD score regression to compare the heritability explained by each set. We applied Benchmarker to compare two gene set enrichment analysis methods, DEPICT and MAGMA, using several well-powered GWAS such as height, body-mass index, and schizophrenia. We defined "prioritized genes" as genes within the enriched gene sets found by each method. We found that while both methods clearly performed better than chance, DEPICT and MAGMA generally performed similarly despite prioritizing fairly distinct groups of genes (average percentage overlap = 32%). We also used Benchmarker to compare two different data sets as input to DEPICT, a gene set membership matrix and a tissue expression matrix, assessing their efficacy in prioritization (where prioritization is done on the basis of a gene's similarity to other genes across the input matrix). Similarly, we found that none of these methods demonstrated clear superiority. These results suggest that evaluation of prioritization algorithms likely should be done on a case-by-case basis, as different prioritization strategies may be optimal for different traits.

1567T

Association testing and estimation of variant effects in multiple traits. *L. Gai¹, E. Eskin^{1,2}*. 1) Department of Computer Science. University of California, Los Angeles, Los Angeles, CA; 2) Department of Human Genetics, University of California, Los Angeles, Los Angeles, CA.

Many variants identified by genome-wide association studies (GWAS) have been found to affect multiple traits either directly or through shared pathways, a phenomenon known as pleiotropy. For this reason, analyzing multiple traits at once can increase power to detect and estimate shared variant effects. However, the degree to which traits share variant effects is often not known, and traditional meta-analysis methods can be underpowered compared to univariate analysis when applied to studies in different traits. In addition, recent studies have suggested that pleiotropy is extremely widespread across the human genome and affects a large portion of human phenotypes. This suggests a need for formal models of pleiotropy. Here we present a flexible framework for association testing and modeling of genetic effects in multiple traits from summary statistics. Using simulated data, we show our method increases power when an effect is present in a subset of traits. We then apply our method to two publicly available GWAS data sets, the Northern Finland Birth cohort and UK Biobank, and discover novel loci in a variety of metabolic traits. We also use this framework for joint estimation of effect sizes across multiple traits and improved prediction of polygenic risk scores for quantitative traits.

1568F

A straightforward method to improve variant calling sensitivity for a *MSH2* splicing mutation (c.942+3A>T) using GATK tools. *A. Gerasimova¹, S.H. Rosenthal¹, D.A. Wolfson¹, A. Smolgovsky¹, A. Grupe¹, A. Buller-Burckle¹, F.L. Lacbawan¹, E.J. Moler¹, C.D. Elzinga²*. 1) Quest Diagnostics Inc, 33608 Ortega Hwy, San Juan Capistrano, CA 92675; 2) Quest Diagnostics Inc, 200 Forest St, Marlborough, MA 01752.

Calling variants in low-complexity and homopolymer regions is a well-known problem for next-generation sequencing (NGS) technologies, due to biochemical and analytical limitations. A specific example in clinical NGS is a pathogenic splice-defect mutation located in a 25-bp homopolymer that is part of *MSH2*; this mutation can be missed by some existing pipelines. We present a straightforward method to increase sensitivity of a Genome Analysis Tool Kit (GATK)-based pipeline to call this particular mutation. We found that HaplotypeCaller alone has a tendency to call noisy deletions in this region, and occasionally misses the *MSH2* c.942+3A>T variant. With HaplotypeCaller, the variant was missed in 5 of 16 known positive samples. With UnifiedGenotyper, using default parameters and input sequence, we observed better but still inconsistent results: it was missed in 3 of 16 known positive samples. In these 3 cases, HaplotypeCaller made an accurate call, but neither method was able to provide consistent results in isolation. However, running UnifiedGenotyper and forcing calls at each site (--output_mode EMIT_ALL_SITES) resulted in an accurate call in 16 of 16 known positive samples. In order to maximize GATK sensitivity to this variant (and potentially others with similar characteristics) we suggest using UnifiedGenotyper with --output_mode EMIT_ALL_SITES parameter in addition to HaplotypeCaller. .

1569W

Multi-omics integration of metagenomic and DNA methylation data using sparse canonical correlation analysis. *J. Greenbaum, H.W. Deng.* Center for Bioinformatics and Genomics, Department of Global Biostatistics & Data Science, School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA, USA.

It is well known that both host genetics and the gut microbiome composition each influence the risk of developing many complex diseases, however the relationship between these factors remains largely unexplored. In addition to the existing evidence for associations between SNPs and the microbial taxa, there is a growing appreciation for the role of interactions between the microbiome and the host epigenome. In this analysis we aim to detect associated features between metagenomic relative abundance data and DNA methylation data measured on the same set of subjects. There are several important challenges to integrating these multiple omics layers including heterogeneous data types and high dimensionality. Conventional techniques for integrating multiple data types such as canonical correlation analysis (CCA) are not effective for analyzing omics data with small sample sizes and a large number of biological features. These issues may be addressed by introducing a sparsity penalty into the CCA model (sCCA). This allows us to incorporate feature selection into CCA to detect the correlation between a smaller subset of important variables. We designed a simulation study using a latent variable model approach to assess the performance of sCCA incorporating various different regularization terms such as LASSO and Elastic Net. We found that on average LASSO seems to identify more true positives and produce a sparser solution, while Elastic Net shrinks fewer coefficients to zero but has a more stable performance. The results are encouraging for the utility of sCCA, a technique traditionally used in the analysis of brain imaging data, to correctly select the correlated features between these two data types. Further exploration of the relationship between the gut microbiome and other omics may have important implications for understanding the biological mechanisms underlying various complex diseases.

1570T

Proteome PheWAS browser: Graph-based interactive platform allowing visualisation of thousands of estimates from Mendelian randomization analyses. *V. Haberland¹, J. Zheng², D. Baird³, A. Gutteridge², B. Elsworth¹, J. Dopierala², J. Staley¹, B.B. Sun⁴, A. Butterworth⁴, J. Yarmolinsky¹, C. Laurin¹, V. Walker¹, J. Liu², K. Estrada², M.R. Nelson², R.A. Scott², P. Haycock¹, G.D. Smith¹, G. Hemani¹, T.R. Gaunt¹.* 1) MRC Integrative Epidemiology Unit, Population Health Sciences, University of Bristol, Bristol, BS8 2BN, United Kingdom; 2) GlaxoSmithKline plc., Gunnels Wood Road, Stevenage, Hertfordshire, SG1 2NY, United Kingdom; 3) Translational Genome Sciences, Biogen, 225 Binney Street, Cambridge, MA 02142, United States; 4) MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, CB1 8RN, United Kingdom.

Proteins can be targeted by drugs for disease prevention and treatment, and a major goal is to improve the efficiency of this process, potentially by eliminating those candidate targets destined to fail earlier in the developmental process. Mendelian randomization (MR) has been widely used to estimate the causal effect of an exposure, e.g. specific plasma protein levels, on specific outcomes, e.g. Alzheimer's disease (AD), using genetic variants as instrumental variables. These analyses can be conducted using existing platforms such as MR-Base (mrbase.org), but proving causality can be difficult as MR relies on several assumptions. Visualizing and interpreting vast numbers of results is also challenging, particularly for non-expert users. However, when combined with appropriate sensitivity analyses, MR is potentially valuable for showing which putative associations are unlikely to be causal. To address these challenges, we have developed a novel proteome PheWAS browser (database and Web interface platform), which allows easy access to precomputed two-sample MR results, including SNP associated information (e.g. effect allele frequency) and results from advanced sensitivity analyses in an interactive format. The database includes MR analysis results for levels of 1449 plasma proteins from four recent protein GWAS against 1483 outcomes (including diseases, risk factors, metabolites) and employs the Neo4J graph-based platform at its core, which can smoothly be translated into an intuitive visual representation of causal dependencies. As an example, we demonstrate the MR results for ADAM11 plasma levels instrumented by the trans-SNP (rs7412) in *APOE*. Here, five associations for this protein putatively influence traits related to various lipid measures and also AD risk (epigraphdb.org/pgt/?query=ADAM11) after multiple testing correction ($p < 0.28 \times 10^{-7}$ for 1,778,473 tests). However, sensitivity analyses for causal directionality and genetic colocalisation either discount or place a warning for most of these putative associations, and all of them are also flagged as being instrumented by the trans-SNP. This example demonstrates how easily and efficiently the user can look up and visualise hypothesised associations and provides guidance on filters to highlight potentially unreliable MR results. In the future, we intend to populate this platform with more omics-related MR results and results from additional analytical methods as they continue to be developed.

1571F

Enhanced clinical copy number variant calling with sample-specific sensitivity. S. Hong, G. Hogan, P. Kaleta, J. Li, D. Muzzey, K. Haas. Counsyl Inc., South San Francisco, CA.

Introduction Accurate structural variant calling, especially for copy number variants (CNVs) that can range in size from a single exon to an entire gene, is critical for clinical germline screening. However, intrinsic variation in sequencing data coupled with insufficient quality control (QC) measures can imperil high CNV-calling accuracy. We combined an enhanced hierarchical Bayesian algorithm for CNV identification with an online simulation framework—embedded in our production bioinformatics pipeline—to improve sample-specific CNV-calling accuracy and flag samples that do not achieve required performance. **Methods** CNVs were called by observing sequencing depth, which reflects the copy number, for ~100 base-pair segments tiled along the targeted region of interest (ROI). The likelihood of each observed depth value was modeled by a parameterized negative binomial distribution with expected depth and variance dependent on the particular sample-segment pair. An additional Bernoulli mixture model automatically disregards spurious depth signals resulting from technical artifacts. A Hidden Markov model (HMM) was used with transition coefficients based upon the prevalence and observed average length of CNVs. Trust Region Newton Conjugate Gradient and Expectation Maximization were used to determine all model parameters. Copy number calls were obtained by running the Viterbi algorithm and noting contiguous non-wildtype segments. Deletions and duplications of one, two, or four exons in size across the ROI were modeled from empirical depth values via downsampling and oversampling, respectively. HMM calling was then rerun and the sample-specific sensitivity was determined by weighing the size and type of variant by its clinically established prevalence. **Results** The algorithm achieved perfect concordance, emitting positive calls for 50 orthogonally confirmed CNVs and negative calls for 685 samples not known to have CNVs in the ROI. The online simulation of 2,400 deletions/duplications in every sample revealed >99% sensitivity. The algorithm showed invariance to a range of model hyperparameters, as well as robustness to 5x variation in sample depth. Retrospective ROC analysis of >10,000 recent random production samples showed an order-of-magnitude improvement in the false-negative rate for batches with high sample variability. **Conclusions** We developed a refined algorithm and complementary sensitivity QC metric to ensure clinical-grade calling of important copy number variants.

1572W

Cox regression outperforms logistic regression for detecting genotype-phenotype associations in genetic data linked to electronic health records. J.J. Hughey^{1,2}, S.D. Rhoades¹, Q. Chen^{1,2}. 1) Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN; 2) Department of Biological Sciences, Vanderbilt University, Nashville, TN; 3) Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN.

Linking genetic data with routinely collected clinical data has enabled genome-wide association studies across the human phenome. Nonetheless, although clinical data are generally longitudinal, most approaches for detecting genotype-phenotype associations in such linked data (notably logistic regression) were originally intended for case-control study designs, and thus do not handle the time dimension gracefully. Here we explored the advantages of quantifying associations using Cox (proportional hazards) regression, which can account for the age at which a patient first visited the healthcare system (left truncation) and the age at which a patient either last visited the healthcare system or acquired a particular phenotype (right censoring). We scanned for genotype-phenotype associations using logistic regression and Cox regression on the electronic health records of approximately 30,000 genotyped individuals. We compared the output from each method to known associations from the NHGRI-EBI GWAS Catalog. Our results suggest that, compared to logistic regression, Cox regression has greater power and equivalent Type I error. The difference in power is largely independent of minor allele frequency, but greater for phenotypes with at least 1,000 cases and for SNP associations whose odds ratio is between 0.6 and 1.6. In terms of effect sizes, the hazard ratios estimated by Cox regression are nearly identical to the odds ratios estimated by logistic regression. Our empirical findings are supported by the results of comprehensive simulations. As longitudinal health-related data continue to grow, Cox regression may improve our ability to identify the genetic basis for a wide range of human phenotypes.

1573T

An accurate CNV model for clinical exome sequencing. *S. Kapil, P. Shaw, S. Nair, J.T. George, S.M. Sakthivel Murugan, V. Ramprasad, R. Gupta, R. Gupta.* MedGenome Lab Ltd., Bangalore, India.

Identification of deletions and amplifications (CNV) from targeted sequencing has become one of the key requirements in genetic testing of Mendelian disorders. Recent reports have shown that accurately detecting CNV results in increasing the diagnostic yield. Although several methods have been developed in past, accurate identification of CNV from targeted sequencing at exon level is still a very challenging task. Among different CNV callers, the read depth approaches are widely used. One of the primary requirement for these methods is the collection of reference or control samples. Due to lack of reference/gold standard dataset, the actual performance of the CNV callers for targeted sequencing is largely unknown. In this study, we have (a) developed a simulator for CNV event creation, (b) benchmarked eight different CNV detection tools (Contra, Exome-Depth,XHMM, ExomeCopy, Cn.mops, Excavator, Control-Freec and Convading) and identified the best performing tool, (c) constructed a sensitive gender specific copy number model for our targeted clinical exome panel of 20Mb size covering coding region of ~6900 known disease causing genes sequenced at an average depth of 100X, (d) validation of CNV model on known MLPA tested samples, (e) improved the run-time execution of the CNV detection tool. We first performed the benchmarking of eight different tools using ~500 CNV events generated using our CNV simulator. The CNV event includes deletion/amplification of single exon to four exons, heterozygous deletion in different proportion, 1 to 3 copies amplifications. The reference samples were selected from our clinical samples satisfying the following criteria: (i) known pathogenic SNP/Indel matching with clinical symptoms identified in the sample, (ii) average read depth of the targeted region ~100X, (iii) passed our all quality metrics (alignment, duplicate percentage, average base quality, on-target read percentage). Overall, we obtained 100 reference samples each for male and female. Our analysis revealed that the exome-depth is the most sensitive program with an overall sensitivity of ~ 75%. Further, we found that the model sensitivity exceeds 90% for the target regions with an average depth >= 100X. Validation of 27 samples using MLPA technique achieved sensitivity of 74.07%. Using our pre-calculated depth model, the runtime of exome-depth for a clinical exome sample was reduced from 4 hours to 15 minutes.

1574F

A computational framework for real time analysis and sharing of large next generation sequencing population datasets. *H. Karten¹, M. Haime^{2,3}, T. Tilly^{2,3}, S. Gräf^{2,3}.* 1) Genalix Core BV, the Netherlands; 2) University of Cambridge, UK; 3) NIHR BioResource - Rare Diseases, UK.

Significant progress has been made in collecting large amounts of genetic data combined with patient information. This has led to the generation of several large repositories that are being used as 'reference' data sets. Depending on ethnicity and volume, these datasets are being used for quality control on smaller cohorts and serve as background to find variants that could be associated with rare diseases. These larger cohorts, however, themselves also exist in a context of accelerated increasing knowledge requiring an almost continuous update of variants and their annotations to be able to serve as a knowledge resource. We have created a real time analysis framework that allows for continuous data analysis and mining using AVRO files. With this framework we migrated one of the largest rare disease cohorts in the world comprising more than 13,000 whole genomes of the NIHR BioResource -- Rare Diseases to the latest human genome reference assembly. We deployed our framework on 10 compute nodes of the University of Cambridge High Performance Computing cluster. The framework uses the Genalix NGS secondary analysis suite for sequencing read alignment, population scale genotyping and preparation of the Hadoop AVRO dataset. All genomes are processable within 14 days (read mapping: 12 days; genotyping: 2 days). Subsequently, the variants are exported to AVRO files and annotated in a distributed Hadoop filesystem. The above described process scales linearly with the number of nodes and samples. The otherwise computationally expensive merging of variants is efficiently performed at the population calling stage. Variants at the same location can be represented with the AVRO file format. The data is ready for further analysis in this format and does not need to be imported into a Hadoop HBase database. Our framework significantly advances research groups to keep pace with the accelerating growth of knowledge and to maintain datasets. This will contribute to the research of others.

1575W

Genomic multilateration. K. Kim^{1,2}, H. Baik³, E. Eskin^{4,5}, J. Noh^{1,2}, B. Han⁶.

1) Asan Medical Institute, Seoul, Korea, Republic of Korea; 2) Department of Convergence Medicine, University of Ulsan College of Medicine & Asan Institute for Life Sciences, Asan Medical Center, Seoul 05505, Republic of Korea; 3) Department of Mathematical Sciences, KAIST, Daejeon 34141, Republic Korea; 4) Department of Computer Science, University of California, Los Angeles, Los Angeles, CA 90095, USA; 5) Department of Human Genetics, University of California, Los Angeles, Los Angeles, CA 90095, USA; 6) Department of Biomedical Sciences, Seoul National University, Seoul, Republic of Korea.

In engineering, many applications use distances between multiple objects. Multilateration is one of the localization technique which uses distances from a node at unknown position to reference nodes at known positions to estimate the spatial coordinates of the node. In the application of the radio navigation of an aircraft, the distances from the aircraft to satellites are used for mathematically calculating the position of the aircraft. We present a framework called *genomic multilateration* where we apply the multilateration technique to genomic data. We consider an individual as a node in a multi-dimensional space where each polymorphic locus becomes an axis. In this space, pairwise distance between nodes represents genetic distance between individuals. Individuals from open data become our satellites with known positions. In contrast to the multilateration where one pinpoints the target position using 4 satellites in the 3-dimensional space, the goal of genomic multilateration is not identifying the exact location. Conversely, we intentionally keep the individual's exact position unidentifiable by making the number of dimension exceeds the number of satellites. This unidentifiability allows us to freely share distances to satellites (*distance vectors*) between institutions or studies without disclosing the genetic contents, thereby protecting privacy. Distance vectors can be shared for two purposes. First, distance vectors can detect duplicate individuals between institutions. We developed a statistical framework to capture duplicates using distance vectors between two datasets without transferring the genetic contents of individuals, which turns out to be highly powerful. The second application of using distance vectors is 2-D mapping. We show that distance vectors of an individual to satellites hold enough information to infer one's geographic origin. Previous studies used principal component analysis (PCA) on genotype data to geographically map individuals, but this required sharing of genotype data. With distance vectors only, our estimated projection closely resembles the plot based on the PCA result of the corresponding data and the geographical map of the Europe.

1576T

FALCON-Phase integrates PacBio and HiC data for *de novo* assembly, scaffolding and phasing of a diploid Puerto Rican genome (HG00733). S.B. Kingan¹, Z.N. Kronenberg², G.T. Concepcion¹, S.T. Sullivan², R.J. Hall¹, P. Peluso¹. 1) PacBio, Menlo Park, CA; 2) Phase Genomics, Seattle, WA.

Haplotype-resolved genomes are important for understanding how combinations of variants impact phenotypes. The study of disease, quantitative traits, forensics, and organ donor matching are aided by phased genomes. Phase is commonly resolved using familial data, population-based imputation, or by isolating and sequencing single haplotypes using fosmids, BACs, or haploid tissues. Because these methods can be prohibitively expensive, or samples may not be available, alternative approaches are required. *de novo* genome assembly with PacBio Single Molecule, Real-Time (SMRT) data produces highly contiguous, accurate assemblies. For non-inbred samples, including humans, the separate resolution of haplotypes results in higher base accuracy and more contiguous assembled sequences. Two primary methods exist for phased diploid genome assembly. The first, TrioCanu requires Illumina data from parents and PacBio data from the offspring. The long reads from the child are partitioned into maternal and paternal bins using parent-specific sequences; the separate PacBio read bins are then assembled, generating two fully phased genomes. An alternative approach (FALCON-Unzip) does not require parental information and separates PacBio reads, during genome assembly, using heterozygous SNPs. The length of haplotype phase blocks in FALCON-Unzip is limited by the magnitude and distribution of heterozygosity, the length of sequence reads, and read coverage. Because of this, FALCON-Unzip contigs typically contain haplotype-switch errors between phase blocks, resulting in primary contig of mixed parental origin. We developed FALCON-Phase, which integrates Hi-C data downstream of FALCON-Unzip to resolve phase switches along contigs. We applied the method to a human (Puerto Rican, HG00733) and non-human genome assemblies and evaluated accuracy using samples with trio data. In a cattle genome, we observe >96% accuracy in phasing when compared to TrioCanu assemblies as well as parental SNPs. For a high-quality PacBio assembly (>90-fold Sequel coverage) of a Puerto Rican individual we scaffolded the FALCON-Phase contigs, and re-phased the contigs creating a *de novo* scaffolded, phased diploid assembly with chromosome-scale contiguity.

1577F

Utilizing IGV.js to visualize genomic variation from trio clinical whole genome sequencing. *BR. Lajoie¹, V. Gainullin¹, A. Halpern¹, A. Warren¹, A. Gross¹, E. Dolzhenko¹, N. Johnson¹, F. Mullen¹, A. Chawla¹, V. Rajan¹, SS. Ajay¹, M. Eberle¹, DR. Bentley¹, RJ. Taft².* 1) Illumina Inc., 5200 Illumina Way, San Diego, CA, USA; 2) Illumina Cambridge Ltd., Chesterford Research Park, Little Chesterford, UK.

Clinical whole genome sequencing (cWGS) has the potential to be a comprehensive platform for the genomic evaluation of patients with unresolved genetic disorders. As both the number and types of variants continues to increase, the need for dynamic, flexible and intuitive genome visualizations increases. Here we describe our deployment and customization of the publicly available Integrated Genome Viewer (IGV.js) to support the visualization of SNVs, INDELs, CNVs, SVs and STRs in the context of trio cWGS. IGV.js is a web-based, embeddable interactive genome visualization tool, capable of interacting with modern data object-store architectures (amazon s3, google cloud). IGV.js was integrated as part of the Illumina iResolve research program, a continuously enrolling study of more than 200 persistently unsolved RUGD cases. The iResolve IGV.js instance allows users to visualize genomic variation across a wide range of commonly used genomic file types. Users can visualize both reads and small variants through the existing support for genomic file formats (BAM, VCF and gVCF). To improve upon the existing IGV.js features, we added support for both additional data types and novel plotting methods. Copy number variants (CNVs) and structural variants (SVs) are visualized via color-coded intervals provided via BED files and confidence intervals around the event breakpoints are displayed to emphasize the ambiguity of the call. A tint is applied to events to visually demarcate the filter status. Depth support for CNV events is represented by a genome-wide normalized depth vector visualized as a color-coded point plot. Read support for SV events is depicted by a BAM pileup of anomalously-aligning reads (>1,000bp insert size or mates mapping to different chromosomes). Basepair resolution quality metrics are displayed by 12 color-coded tracks to aid in evaluation of the overall quality of any given genomic interval. The iResolve IGV.js instance also adds support for visualizing population event aggregates of healthy control samples. These aggregate visualizations help to visually inspect frequencies of both common and rare events along with positional variation in the observed population.

1578W

A-Lister: A tool to identify the most important '-omics' entities across multiple sample groups. *S. Listopad¹, T.M. Norden-Krichmar.* University of California, Irvine, Irvine, CA.

A-Lister is a downloadable freeware tool with an intuitive graphical user interface (GUI) intended to help genomics researchers identify the most important (A-list) entities in their '-omics' data. There are several tools available that perform pairwise group comparisons, but there are very few tools with easy to use GUIs that find both similarities and differences across multiple group comparisons. To address this need, we have developed an open source tool called A-Lister that identifies differences and similarities across multiple lists of data. For example, if the input data files contain gene names, A-Lister will output the gene names that were found in the intersection and union between the groups. Besides exact matching of gene names, A-Lister also provides "fuzzy" matching of gene names based on several metrics, in order to expand the list to potentially match gene names in the same gene families. If differential gene expression files containing fold change values are input, then A-Lister can also be configured to match a user-specified pattern of up- or down-regulation between the groups. The tool was originally conceived to aid in the analysis and integration of differential gene expression data from RNA sequencing generated across multiple biospecimen sample types and multiple disease groups by the Southern California Alcoholic Hepatitis Consortium. While A-Lister has been utilized in our work with transcriptomics data, it could easily be expanded to handle many different types of '-omics' data, such as proteomics, methylomics, and metabolomics. A-Lister could also be integrated into an existing bioinformatics pipeline to facilitate data analysis.

1579T

A preliminary assessment of long-read structural variant calling software. J. LoTempio^{1,2}, S. Bhattacharya², H. Barseghyan², E. Delot^{1,2}, E. Vilain^{1,2}. 1) Institute for Biomedical Sciences, George Washington University, Washington, DC; 2) Children's Research Institute, Children's National Medical Center, Washington, DC.

Background: The human genome reference completed in 2003 leveraged short-read capillary sequencing; subsequent drafts, too, have made use of short-reads from sequencing by synthesis. These initial drafts were closely followed by the initial reports in 2004 and 2005 of the feasibility of long-read sequencing technology platforms, specifically single-molecule real-time sequencing and nanopore-based sequencing, developed by Pacific Biosciences and Oxford Nanopore, respectively. Nearly a decade and half later, these technologies are bearing fruit for our understanding of the structure and sequence of previously dark or repetitive regions of the genome. Another benefit to this technology is the high-accuracy with which genomic structural variation (SV) can be assessed. **Objective:** Establish an alignment → mapping → SV calling pipeline and a set of best practices after evaluation of extant software for long-read SV calling across available platforms. **Methods:** We intend to leverage the Genome in a Bottle healthy female sample, NA12878. Sequence data from short- and long-read platforms, as well as next generation mapping data are available on the NCBI FTP site or Amazon Web Services for this individual. At present, there are five software programs designed specifically for long-read SV calling, including PBHoney, Sniffles, NanoSV, SMRT-SV, and SVAnalyzer; the result of each of these programs is dependent on, and tailored to, a specific mapping algorithm. Here, we compare the results from four SV callers on four mapping algorithms, including BWA-MEM, minimap2, NG-MLR, and LAST. **Conclusions:** The use of gap-aware mapping algorithms greatly increases the detection of real SV, as determined by confirmation of common SV across long read and next generation mapping platforms.

1580F

Comparison of indel callers and metrics to evaluate performance of new tools. S. Marwaha^{1,2}, E. Ashley^{1,2}, M.T. Wheeler^{1,2}. 1) Division of Cardiovascular Medicine, School of Medicine, Stanford, CA, USA; 2) Stanford Center for Undiagnosed Diseases, Stanford, CA, USA.

Background: Next-generation sequencing is increasingly being used both clinically and for discovering new gene-disease associations and thus it is critical to have accurate and efficient variant calling tools. Although most of the current tools call SNPs with high accuracy, there is low concordance between different indel callers. **Methods:** We compared four current state of art variant callers - DeepVariant, GATK4 Haplotype Caller (HC), Real Time Genomics (RTG), Samtools using Genome In A Bottle (GIAB) dataset. GIAB provides high-confidence variant calls and regions for several well-characterized human genomes from publicly available cell lines and DNA, which can be used as 'gold truth' for evaluating a variant caller. We saw high discordance between variant evaluation results by different tools. This is due to discrepant representation of same set of variants by different algorithms, for example, indels and complex variants; consecutive SNPs and MNPs. For our analysis, we chose GA4GH benchmarking tool which allows user to select between vcfEval or hap.py, tools that use sophisticated matching algorithms to handle complex call representation correctly. We used Recall (Truth.TP / (Truth.TP + Truth.FN)), Precision (Query.TP / (Query.TP + Query.FP)) and F1-score (2*Recall*Precision / (Recall + Precision)) as metrics to evaluate the performance of each caller. Further, we investigated variants where most callers performed poorly and how the performance varied with different properties of genomic region. **Results:** We found that DeepVariant outperformed other tools with Recall of 93.10% and Precision of 89.96%, closely followed by GATK4 HC. We also discuss how different genomic metrics affect the performance of the various indel callers. For example, even though DeepVariant had highest F1 score of 0.915 for calling all indels, it performed suboptimally in detecting complex indels (Recall of 8.07% and 16.18% for Complex variants of length 1-5 (C1-5) and 6-10 (C6-10) respectively) as compared to the other callers like RTG (Recall of 95.07% and 97.06% for C1-5 and C6-10 respectively). **Conclusion:** Our work highlights the discordance among tools used for benchmarking variant callers and provides useful insights and guidelines for selecting indel callers for clinical and research goals.

Evaluation of Indel Calling Tools			
Tool	Sensitivity (%)	Precision (%)	F1-score
RTG	92.32	77.59	0.843
DeepVariant	93.1	89.96	0.915
GATK4 HC	90.66	89.66	0.902
Samtools	62.58	72.94	0.674

1581W

Create a reproducible paper with FireCloud. *T. Miller, G. Van der Auwera, K. Noblett, I. Rosenberg, Data Sciences Platform (DSP).* Broad Institute, Cambridge, MA.

The lack of portability and reproducibility of analysis methods limits the effectiveness with which biomedical researchers can benefit from the democratization of genomic analysis. Many analysis tools have complex requirements, making them difficult to utilize without advanced computational training. Conversely, it is difficult for the authors of new methods to share their work in a way that is readily reproducible by others. Method developers or consumers can bypass these challenges by using FireCloud, an open-source, freely accessible cloud-based analysis platform developed at the Broad Institute. Three parts of the platform work together to make method sharing and reproducibility of analysis easy: workspaces, batch analysis (pipelines), and interactive analysis with Jupyter Notebooks. A workspace is a stand-alone computational sandbox that provides a secure collaborative place to organize data, run and monitor analysis pipelines, and develop notebooks. Pipeline results can be queried in the notebook with whatever data analysis or visualization module the researcher wants to import. The notebook will record the steps the researcher takes to process raw data to generate a plot or graph that may be used in a paper. In summary, a researcher can share their workspace URL, and anyone with permission can re-run the batch analysis pipeline and re-execute the notebook steps to produce the same results or even apply this analysis on a different dataset.

1582T

Unifying copy number variant calling and imputation from SNP arrays.

T. Mimori¹, Y. Kawai², K. Ueno¹, S. Khor², Y. Hitomi², O. Gervais¹, K. Tokunaga², M. Nagasaki¹. 1) Department of Integrative Genomics, Tohoku Medical Megabank Organization, Tohoku University Organization, Tohoku University; 2) Department of Human Genetics, Graduate School of Medicine, The University of Tokyo.

Copy number variants (CNVs), including deletions and duplications of genomic segments, affect a considerable amount of base pairs in the human genome. As CNVs can cause Mendelian or sporadic traits, or be associated with complex diseases, sensitive and accurate calling of individual CNVs is an important task for genetics and human health studies. With the advent of SNP array technology, the genome-wide detection of large CNVs, which extend across multiple array probes, has become possible. Although high-throughput sequencing (HTS) technology enables more comprehensive detection of CNVs with higher resolution, SNP array is still the primary choice for large sample studies because of its cost advantage. To complement genotypes of unobserved SNPs, genotype imputation has been regularly used in genome-wide association studies. Also, there have been several studies that impute structural variants including CNVs. However, CNV imputation, particularly for duplications, has been a challenging problem partly because of a lack of good tag SNPs for CNVs. To improve CNV prediction accuracy from SNP arrays, it is important to make use of both linkage disequilibrium (LD) structures in reference panels and raw probe signals obtained from SNP arrays. However, traditional genotype imputation methods need reliable tag SNPs outside of CNVs and cannot use raw probe signals at CNV loci. We propose an approach that unifies CNV calling and imputation processes to overcome this limitation. For evaluation, we constructed a population-specific reference panel including CNVs by using high-depth HTS data from 418 Tokyo Healthy Controls (THC) samples. The concordance between reference and predicted copy numbers considerably improved with this approach. In our presentation, we provide the details of this method and describe the performance evaluation results.

1583F

Locus Reference Genomic (LRG): A resource for reporting clinically relevant sequence variants. *J. Morales¹, A. Astashyn², R. Bennett¹, C. Davidson¹, A. Frankish¹, L. Gil¹, V. Joardar¹, M. Kay¹, J. Loveland¹, K. McGarvey², A. McMahon¹, G. Threadgold¹, T. Murphy², F. Cunningham¹.* 1) European Molecular Biology Laboratory - European Bioinformatics Institute (EMBL-, Cambridge, Cambridgeshire, United Kingdom; 2) National Center for Biotechnology Information, National Library of Medicine, Bethesda, MD, U.S.A. 20894.

The correct interpretation of clinically relevant variants is dependent on accurate and unambiguous annotation of genes and transcripts. Central to this is the reference sequence against which a variant is reported. Ongoing changes to the reference assembly, the existence of numerous versioned reference sequences and the use of legacy reporting systems present challenges in the clinic. The Locus Reference Genomic (LRG), a manually curated resource designed specifically for the reporting of clinically relevant variants, addresses these challenges. Each LRG record provides a stable and non-versioned genomic DNA sequence for a region of the human genome, establishing a coordinate system that is independent from upgrades to the reference assembly. The records also contain a minimal set of stable transcripts to be used as reference standards. Only transcripts for which there is currently good biological understanding are selected for inclusion. Together with RefSeq (NCBI) and Ensembl/GENCODE (EMBL-EBI), our goals for 2018/2019 include faster convergence on key high value annotations in order to ensure identity between RefSeq and Ensembl for transcripts included in LRGs and used for variant reporting. LRGs are manually curated and created in consultation with locus-specific experts in the diagnostic and research communities to ensure that each record fulfils the unique demands of each specific locus. We have recently released a new version of the LRG website to enhance the display of records and improve functionality. LRGs have been recommended by HGVS and other key organisations. Their widespread use will ensure consistent variant reporting over time. LRGs are created jointly by the NCBI (and EMBL-EBI (<http://www.ebi.ac.uk>), are available on the LRG website (<http://www.lrg-sequence.org/>) and can be visualised in all genome browsers. This work was supported in part by the intramural research program of the National Library of Medicine, National Institutes of Health.

1584W

Systematic meta-analysis of structural variation and CNV algorithms calling performance using short read sequencing. *J.A. Morales-Rosado^{1,2,3}, F. Vairo^{1,2,3}, C. Wang^{1,2,3}, E.W. Klee^{1,2,3}.* 1) Center for Individualized Medicine, Mayo Clinic, Rochester, MN; 2) Department of Health Sciences Research, Mayo Clinic, Rochester, MN; 3) Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester.

Introduction: An impressive plethora of bioinformatics tools exists for structural variant (SV) and copy-number variation (CNV) detection with a consistent yearly output of a dozen new tools. Therefore: (1) We try to understand heterogeneity of CNV/SV detection performance with respect to size-limit. (2) Evaluate translational potentials/applicability of existing popular tools. **Methods:** We have performed a systematic review with meta-analysis of SV and CNV algorithms to address the strengths and limitations in currently published tools from 2009 – 2017. (1) We selected the 10 most popular tools within 2009-2017 (based on citations per year extracted from the Web of Science database). (2) We summarize citations content and usages of tools. (3) Evaluated performance measurements with respect to size range. (4) Carried out a detailed examination of NA12878, NS12911 and HS1011 call set use as ground-truth for validation methodology. Statistically, we determined the overall performance of the most used algorithms using regression of concordance rates with the used truth-sets when the exact number of calls was reported or could be extrapolated from reported values. **Results:** The top 5 most popular tools currently used are: Pindel, Breakdancer, CNVnator, Delly, and Lumpy. We identified a series of tools that have reached a plateau of use and others that have fallen into disuse and attempted to uncover some of the factors influencing such trends. Variant size-based performance is currently underreported. Variant size-based reports available demonstrate a broad interval, 40-70% false-discovery rate (FDR), among calls made in the 50bp-1kb size-range. There is high heterogeneity in the use of gold-standard calls from unique sources and consistent omission of calls generated by algorithms during literature comparison with other tools. **Conclusion:** Taken as a whole, our large scale systematic meta-analysis of SV and CNV calling tools yield important recommendations to the bioinformatics and clinical genomics areas on several key aspects: (1) Algorithms performance reports need to be standardized for peer-review and clinical users, with attention to variant size effects. (2) Harmonized and consensus ground-truth sets are highly needed to ease future CNV/SV algorithm development and validation.

1585T

Prioritizing SNPs for functional experiments using available annotations as a feature selection problem. A. Mousas^{1,2}, M.A. Cole⁶, L. Pinello^{3,4,5}, D.E. Bauer⁶, G. Lettre^{1,2}. 1) Department of Medicine, University of Montreal, Montréal, Québec, Canada; 2) Montreal Heart Institute, Montréal, Québec, Canada; 3) Molecular Pathology Unit, Massachusetts General Hospital, Charlestown, MA, USA; 4) Cell Circuits and Epigenomics Program, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; 5) Department of Pathology, Harvard Medical School, Boston, Massachusetts, USA; 6) Division of Hematology/Oncology, Boston Children's Hospital, Department of Pediatric Oncology, Dana-Farber Cancer Institute, Harvard Stem Cell Institute, Department of Pediatrics, Harvard Medical School, Boston, Massachusetts, USA.

Background. Genome-wide association studies have identified thousands of variants associated with complex human phenotypes. One outstanding challenge remaining is to prioritize for functional experiments variants that are in strong linkage disequilibrium. We explored this issue through a feature selection problem by analyzing a dataset of 5,592 SNPs functionally tested by CRISPR-Cas9 for their role in red blood cell biology. **Methods.** As features, we considered various relevant annotations such as erythroid enhancers, ATAC-seq peaks, GATA1 binding motifs, and eQTLs. We divided the CRISPR dataset in two distinct and mutually exclusive subsets: the training set (80%) and the test set (20%). We monitored model performance using various metrics including accuracy, area under the receiving (AUC) curve, sensitivity and specificity. We used 10-fold cross-validation in the training set to tune our models, and sub-sampled the held-in data to account for the extreme class imbalance. We tried two different feature selection methods. For the “filter” method, which disentangles the selection criterion from the model's performance, we pre-selected significantly associated predictors using Fisher's exact test and subsequently fitted a logistic regression and a random forest model. We performed recursive feature elimination with a random forest maximizing AUC for the “wrapper” method. After selecting the important features, we created a per-SNP score counting into how many annotations each SNP falls inside (“voting”). The top scoring SNPs were selected and the corresponding loci were identified. **Results.** We were able to extract 4 and 6 important features of total 23 for the filter and wrapper approach, respectively. These top annotations included erythroid enhancers and ATAC-seq peaks. All 4 selected features were included in the 6 annotations picked by the random forest wrapper. The additional features included plastic regions for chromatin states. In summary, 36 and 41 out of the 308 unique loci were identified by our method with a maximum overlap of 67% while model performance reached approximately AUC = 0.6 with the non-linear wrappers reaching the highest. **Conclusions.** The integration of available annotations to identify causal SNPs is an important problem. We have presented a simple yet principled approach to guide future functional experiments that aim to characterize discoveries from human genetic studies.

1586F

Pedigree-based analysis pipeline (PBAP) v.2: An upgraded version with additional capabilities. A.Q. Nato¹, N.H. Chapman¹, M. Saad², H.K. Sohi¹, C.Y.K. Cheung³, A.R.V.R. Horimoto⁴, R.A. Nafikov¹, K. Kunji², E. Ullah², H.D. Nguyen¹, E.M. Wijsman^{1,4,5}. 1) Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA; 2) Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar; 3) Center of Biologics Evaluation and Research, US Food and Drug Administration, Washington DC, USA; 4) Department of Biostatistics, University of Washington, Seattle, WA; 5) Department of Genome Sciences, University of Washington, Seattle, WA.

Availability of next generation sequencing technologies now makes dense genetic data very common. Most genetic analysis software has been developed for population-based studies, but recognizing the importance of rare variants has again made smaller pedigree-based studies common. We formerly developed a pedigree-based analysis pipeline (PBAP) v.1 that allows users to perform several procedures for pedigree-based genetic analysis. These procedures include manipulating files, sub-selecting markers from a dense panel, validating pedigree structures, and sampling inheritance vectors (IVs), which specify the flow of founder alleles in a pedigree. Here, we describe a second version of PBAP with new capabilities that set up files to make use of the IVs for downstream analyses. PBAP v.2 performs the following analyses by preparing files and accessing different programs: a) parametric linkage analysis including options for tailoring marker allele frequencies to admixed populations, b) variance components linkage analysis, c) family-based genotype imputation, and d) genotype-based kinship estimation that allows use of external allele frequencies. Additional tools have also been added to the pipeline that allow the user to: a) calculate pairwise kinship coefficients based on the sampled IVs, b) visualize spacing of the sub-selected panel of markers, c) perform family-based imputation using structural variant data, and d) compute kinship estimates between unsampled individuals with imputed data from possibly different pedigrees to enable association testing in the presence of background relatedness. These additional tools in PBAP v.2 extend its capabilities and give users more options to make maximal use of their data for family-based analyses.

1587W

A novel deep-learning approach to predict phenotype from genotype and its applications to Japanese direct-to-consumer genetic testing. S. Nogawa¹, T. Kuo², K. Saito¹, S. Takahashi¹, J. Sese². 1) Genequest Inc. , 5-29-11. Siba, Minato-ku, Tokyo 108-0014, Japan; 2) National Institute of Advanced Industrial Science and Technology, 2-4-7. Aomi, Koto-ku, Tokyo 135-0064, Japan.

More than 3,000 genome-wide association (GWA) research studies have been published which enable Direct-to-Consumer (DTC) genetic testing companies to provide trait phenotype of subjects from their genomic data. However, GWAS studies generally perform statistical screening of single nucleotide polymorphisms (SNPs) and not for phenotype prediction, the prediction of trait phenotype using genomic data remains one of the most challenging problems in human genetics research. The present study aims to explore the accurate prediction of phenotype while considering the interaction of multiple SNPs with deep-learning (DL) methods. In contrast to numerical analysis using DL for image processing or signal processing tasks, there are more difficulties in applying DL to the accurate and stable analysis of genotype data because the information is discrete, with the presence or absence of variants. In this study, we have developed a new DL model and hyper-parameter tuning method which enables an accurate and stable analysis of genotype data and the prediction of trait phenotype from genome-wide SNP data. Genome-wide SNP data of more than 10,000 Japanese subjects from the Japanese DTC genetic testing company, were used in a phenotype prediction analysis using the new DL model and method. The phenotype inquiry information obtained from subjects on ABO blood group and eye color were used for the evaluation of the prediction as these traits are already known with regards to their association to genetic data. From the results of the DL analysis, the phenotypes were successfully predicted with a high accuracy of more than 95%. These results were validated with a different population using the public genome information, which suggests high accuracy and good versatility in the new DL model and method. Furthermore, the predictions of some complex and challenging traits were evaluated, including taste preferences and proficiency in types of sports, for which the association to genetic data has not been previously identified. The results of this study suggest a great potential for applying our new DL model and method to the accurate prediction of phenotype from genomic data and contribute to solving one of the most challenging problems in human genetics research.

1588T

Unifying primary, secondary, and tertiary genomic analytics in a scalable analytics platform. F.A. Nothaft, H. Davidge, D. Lin, R. Sriharsha. Databricks, Inc, San Francisco, CA.

As the size of genomic cohorts continues to grow, many organizations are turning to the cloud to alleviate their computational burden. While cloud computing promises elasticity and scalability, traditional bioinformatics tools are designed for single node architectures and cannot effectively leverage cloud computing. This impedance mismatch leads bioinformatics researchers to spend lots of time crafting complex and monolithic pipelines which are slow to run and difficult to share. Simultaneously, this architecture prevents bioinformaticians, computational biologists, and translational analysts from accessing the same dataset from a common infrastructure. To address this problem, researchers have turned to scalable systems like Apache Spark, and tools like Hail, GATK4, ADAM, SparkSeq, and VariantSpark. These tools provide high level APIs that simplify implementing algorithms for analyzing large genomic datasets. Since these tools are built on Apache Spark, they can be scaled to thousands of cores, achieving a scale necessary for processing contemporary genomics cohorts. While these tools present a significant advancement over single-node genomics tools, there is a great opportunity to enhance the state-of-the-art. In this talk, we present a unified analytics platform for processing genomics data. This platform achieves a 100x improvement in performance over the current state-of-the-art, which enables a dramatic improvement in processing speeds for batch analyses like variant calling, while unlocking novel tertiary analyses. We will illustrate how these optimizations allow us to reduce the latency of calling variants using the GATK to under one hour for a 30x coverage whole genome sequencing run, while achieving computational cost that is competitive with state-of-the-art variant calling pipelines and addressing the concordance issues in the GATK4's Spark-based Haplotype-Caller implementation. This approach is not restricted to GATK4, and can be used to scale other variant callers (such as FreeBayes) to thousands of cores. We then describe how these performance optimizations are able to unlock tertiary analyses and interactive visualization use cases that query read and variation data across 10,000+ sample cohorts. This unified platform enables scaling traditional genomics analyses to thousands of cores while providing bioinformaticians, computational biologists, and translational scientists access to a uniform view of the datasets within their organization.

1589F

Shellflow: Shell-script like scientific workflow management system.

Y. Okamura¹, S. Tadaka², K. Kinoshita^{1,2,3}. 1) Advanced Research Center for Innovations in Next-Generation Medicine, Tohoku University, Sendai, Miyagi, Japan; 2) Tohoku Medical Megabank Organization, Tohoku University, Sendai, Miyagi, Japan; 3) Graduate School of Information Sciences, Tohoku University, Sendai, Miyagi, Japan.

Current bioinformatics analyses are built on combination with a lot of tools. To manage these complex workflow, a lot of computational workflow management systems, such as Nextflow, Snakemake, Luigi or Cromwell, were proposed. These systems aimed at reproducibility, scalability and reusability. Since they were designed for scaling with cloud services, and reusing parts of sub-workflow in many workflows, the syntaxes of workflow are complex and hard to learn for new users. For example, Common Workflow Language (<https://github.com/common-workflow-language/common-workflow-language>) requires 7 lines to run "echo hello, world". Other workflow management systems also require many lines to run even if we want to run simple command. Here, we propose Shellflow, easy-to-write, shell-script like workflow management system. To run "echo hello, world", shellflow requires only one line because the syntax of shellflow is almost identically with the shell-script. Shellflow automatically recognizes dependency of commands from simple annotation, and submits to local executor, Grid Engine or other job scheduling systems. This software also has all executed command history and rerun commands if data or command arguments are changed. Since Shellflow focused on rapid development of scientific workflow, simpleness and reproducibility, not to scalability and reusability, Shellflow does not support cloud services, and sub-workflow. In conclusion, Shellflow provides easy and simple workflow management, and traceability of works.

1590W

Kibio.science: Unify and visualize life sciences massive data. R. On-garo-Carcy¹, M. Leclercq¹, A. Dessemond¹, H. Ceril¹, M.P. Scott-Boyer¹, F. Belleau¹, O. Perin², A. Droit¹. 1) CHU de Québec Research Center - Université Laval, Quebec city, Quebec, Canada; 2) L'Oréal Research and Innovation.

Biological datasets have grown exponentially in size and complexity. Representations of biomolecules like DNA/RNA/protein sequences and structures and functional information generated by Omics research (e.g. genomics, metagenomics, transcriptomics, proteomics and metabolomics) amount to tremendous quantities of data. These data are currently organized and made available in hundreds of bioinformatics repositories. However, these repositories, or databases, are disparate and most are specialized in certain biological entities or biological fields. Researchers constantly have to navigate from a database to another to perform their analyses. They often have to adapt their data through various handmade scripts and softwares to make their data analyzable. Moreover, publicly available online resources to analyze data generated by last generation technologies, combine them with clinico-pathological and phenotypic data and link them easily with online knowledge are still lacking. To fill this need, we have developed **Kibio**, a biohub dedicated to the biology and bioinformatics sciences communities. Kibio provides a central view to explore, visualize, and analyze the overall content of repositories such as PubMed, ChEMBL, GTEX, KEGG, snpDB or UniProt simultaneously. Kibio is based on Elasticsearch search engine and Siren, a visualization tool, which allows relational searches between databases. Kibio also contains an integration program that has been developed to convert biological databases into Elasticsearch format for optimal use. It offers advanced filtering and dynamic web visualizations to analyze biological knowledge. Simultaneously mining many databases is a strong core-tightened feature, which allows to deepen the exploration of potentially new biological functions and lead to novel interpretations and discoveries. During exploration, each interaction updates automatically all graphs or statistics built in every linked database. All these features and options offered by Kibio require no programming skills and have been designed to be user-friendly, rapid and flexible. In the near future, Kibio will allow users to import their own datasets, connect them to the desired repositories and create their own visualizations. We aim to make Kibio a great resource to explore, visualize, and analyze multidimensional Omics data. Finally, to enhance the analytical power of Kibio, specialized visualizations dedicated to medical or research purpose will be implemented.

1591T

Exploring bias in ATAC-seq experiments with ataqv, an interactive quality control tool for ATAC-seq data. *P. Orchard, Y. Kyono, J. Hensley, J. Kitzman, S. Parker.* University of Michigan, Ann Arbor, MI.

The assay for transposase-accessible chromatin using sequencing (ATAC-seq) has become the preferred method for mapping chromatin accessibility, due to its speed and low input material requirements. However, it can be difficult to evaluate data quality when processing many samples. Here we uniformly processed >1,500 published ATAC-seq samples across 27 studies. We identified considerable heterogeneity in the data even within single studies. To facilitate ATAC-seq quality control (QC) we created ataqv, a software package for quantifying and dynamically visualizing ATAC-seq bias. Ataqv enables comparison of diverse QC metrics across samples, including a new fragment length distribution (FLD) reference metric. Ataqv produces both machine-readable metrics (JSON format) and an interactive human-readable report (HTML format). Chromatin structure differs across classes of regulatory elements, and different chromatin structures yield different local ATAC-seq FLDs. We therefore hypothesized that FLD-influencing technical noise, such as over- or under-transposition, will obscure true biological signals. In order to test this, we performed ATAC-seq on GM12878 cells using seven different Tn5 concentrations with six replicates per concentration (42 libraries total), while holding the number of nuclei constant at 50K. In addition to predictably shifting the FLD towards shorter fragments, Tn5 concentration positively correlated with enrichment of reads around transcription start sites and in peaks, and negatively correlated with the percentage of mitochondrial reads. Using a negative binomial generalized linear model (NB-GLM), we identified 60,392 (of 92,104) peaks showing an increase in signal with increasing Tn5 concentration (5% FDR). Furthermore, peaks positively associated with increasing Tn5 concentration are more enriched in strong enhancer, active promoter, and weak promoter chromatin states, and more depleted in repressed chromatin regions. This suggests that ATAC-seq experiments performed specifically for the purpose of identifying enhancers and promoters will achieve better signal-to-noise with increasing Tn5 concentration. Importantly, employing our FLD reference metric as a covariate in the NB-GLM effectively corrected for the controlled technical bias (different Tn5 concentrations) we introduced. We conclude that ataqv will help control for technical bias and produce more reliable ATAC-seq analysis results.

1592F

Data visualization feature of Analyst Portal-A distributed real-time data query system. *H. Qiu¹, F. Mentch¹, P.M.A. Sleiman^{1,2}, H. Hakonarson^{1,2}.* 1) Children's Hospital of Philadelphia, Philadelphia, PA, USA; 2) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.

Analyst Portal is a distributed and integrative web tool developed in-house that supports complex queries across data modalities including but not limited to Electronic Medical Records (EMR), laboratory workflow information, genotyping and next generation sequencing (NGS) data. Analyst Portal uses distributed database links to query different source databases on the fly, without the need for a centralized data store. Since production 8 years ago, Analyst portal successfully assisted researchers in data exploration, hypothesis generation and discovery making. Data visualization is the presentation of data in a graphical format to help users see and understand data. In recent years, JavaScript HTML5 charting has gained popularity in visualizing data on web pages. We evaluated several popular JavaScript charting libraries. We identified two initial areas in Analyst Portal to implement web charting. First, we created customizable, interactive dashboard for key Laboratory workflows, to improve real time progress tracking and status reports. Second, for patient cohort exploration, after a cohort meeting some criteria is returned after search, the patient demographics (gender, race, age distribution) charts will be shown simultaneously. A patient's specific lab measurement such as white blood cells count, or IgE level is visualized with the reference distribution created in real time from live data. Here we report our current experience with interactive web data visualization, and discuss how it makes Analyst Portal more interactive and intuitive to use.

1593W

GARFIELD-NGS: Genomic variants filtering by deep learning models in NGS. *V. Ravasio¹, M. Ritelli², A. Legati², E. Giacomuzzi².* 1) Department of Molecular and Translational Medicine, University of Brescia, Brescia 25123, Italy; 2) Unit of Molecular Neurogenetics, Fondazione IRCCS Istituto Neurologico 'Carlo Besta', Milan 20126, Italy.

Exome sequencing approach is extensively used in research and diagnostic laboratories to discover pathological variants and study genetic architecture of human diseases. Even if present platforms produce high quality sequencing data, false positives variants remain an issue and can confound subsequent analysis and results interpretation. Here, we propose a new tool named GARFIELD-NGS (Genomic vARiants Filtering by dEep Learning moDEls in NGS), which is based on deep learning models to dissect false and true variants in single sample exome sequencing experiments performed with Illumina or ION platforms. GARFIELD-NGS consists of 4 distinct models tested on NA12878 gold-standard exome variants dataset (NIST v.3.3.2): Illumina INS/DELS, Illumina SNVs, ION INS/DELS, and ION SNVs. AUROC values for each variants category are 0.9269, 0.7998, 0.9464, and 0.9757, respectively. GARFIELD-NGS is robust on low coverage data down to 30X and on Illumina two-colour data, as well. Our tool outperforms previous hard-filters, and calculates for each variant a score from 0.0 to 1.0, allowing application of different thresholds based on desired level of sensitivity and specificity. GARFIELD-NGS processes standard VCF file input using Perl and Java scripts and produces a regular VCF output. Thus, it can be easily integrated in existing analysis pipeline. Availability: GARFIELD-NGS is available at <https://github.com/gedoardo83/GARFIELD-NGS>.

1594T

Development of a clinical summary tool that reduces clinical note review time and standardizes phenotype terminology. *J.S. Salvo, W. He, F. Xia.* Baylor Genetics, Houston, TX.

Clinical laboratory testing requires accurate evaluation of physician notes to interpret patient phenotypes. This has become especially important with whole exome sequencing (WES), as it can be used in the molecular diagnosis of a wide range of diseases. As WES testing continues to become more prevalent, a need arises for both efficient clinical note interpretation, and standardization of phenotype terminology. We have developed a computer application, the Clinical Summary Clerk (CSC), that improves the rate and accuracy of clinical note interpretation using autocomplete entry fields and phenotype synonym matching. By using the terminology provided by the Human Phenotype Ontology (HPO), the CSC combines synonyms and standardizes patient phenotypes. Phenotype language is important for patient database consistency and using HPO terms improves querying ability by reducing the number of missed patients due to unknown synonymous or misspelled phenotypes. Additionally, variant prioritization software, such as Codified and Emedgene, often require a list of HPO terms to describe patient phenotypes. The CSC helps provide these terms without any additional time input, as list files are automatically generated. With the reduced time for HPO term or synonym lookup, quick phenotype list editing, and automated file generation, the CSC can reduce time spent summarizing clinical notes by more than 30%. While human interaction is still required for clinical interpretation, it is important to have tools like the CSC to reduce human error and improve efficiency.

1595F

New features facilitating search and analysis of models of human disease at Mouse Genome Informatics and The Alliance of Genome Resources. *C.L. Smith, The MGI Curation and Software Team.* Mouse Genome Informatics, The Jackson Laboratory, Bar Harbor, ME.

Mouse Genome Informatics (MGI) is an internationally recognized community knowledgebase for the laboratory mouse. MGI integrates mouse genotype-phenotype datasets from biomedical publications and large-scale projects with genomic, mutation, expression, functional and human disease model data to accelerate correlative discoveries and inform genetic disease etiology and therapeutics. New MGI functionality reinforces the semantic acquisition, annotation, integration and display of phenotype and disease model information, and facilitates comparison of disease-related data across all model organism databases of the Alliance of Genome Resources (www.alliancegenome.org). Recent enhancements include a new MGI Human Phenotype Ontology browser, incorporation of human disease-phenotype relationships from Orphanet, and new mouse embryonic lethal phenotype data from the DMDD consortium. Importantly, MGI has adopted the Disease Ontology (DO), a hierarchical classification of standardized disease concepts with cross-references to OMIM, MeSH, SNOMED and other clinical terminologies. that is also used by all members of the Alliance of Genome Resources. MGI disease model annotations have been translated from OMIM terms to DO terms to optimize searches, and grouping and display of disease classes in relevant detail pages. The MGI Quick Search tool, Human-Mouse Disease Connection (HMDC), and advanced query forms now support searches by DO terms or IDs. Furthermore, the HMDC grid groups phenotype and disease data by top-level terms to enhance visualization of these data. Finally, an improved DO Browser allows users to traverse the ontology tree or graphical view, examine a selected term's definition, IDs and hierarchical relationships, identify all genes and mouse models annotated to that disease or any of its subtypes, and access the unified Alliance multi-species disease report and related data from other species.

1596W

Reproducible workflows for genomics data analyses. *W. Souza^{1,3}, B.*

Carvalho^{2,3}, J. Lopes-Cendes^{1,3}. 1) Department of Medical Genetics, School of Medical Sciences, University of Campinas, UNICAMP, Campinas, SP, Brazil; 2) Department of Statistics, Institute of Mathematics, Statistics and Computer Science, University of Campinas, UNICAMP, Campinas, SP, Brazil; 3) Brazilian Institute of Neuroscience and Neurotechnology (BRINN), Campinas, SP, Brazil.

New sequencing technologies have improved our abilities to collect biological information for large-scale population studies. The availability of public datasets and the rise of federated genomic databases have allowed researchers to find rare genomic variants. An increasing computing power is required to produce meaningful scientific contributions from high-dimensional genomic data. Cloud services have become affordable and popular among bioinformatics groups for computational-intensive tasks. The combination of large-scale datasets and cloud environments has resulted in new challenges to make genomic research more reproducible, which means the ability of other researchers to achieve the same results of the original analysis. Task-specific bundles of software called containers have been used to guarantee that tools will always give the same results regardless of the computing environment. Domain-specific languages for describing tools and workflows have helped documentation and automation of bioinformatics data analysis protocols. To make our genomic data analyses more reproducible, we have changed the methods we used to describe bioinformatics tools and workflows, and how we execute large-scale data processing tasks. We created several Docker container images for bioinformatics tools. We also wrote description files for those tools using the Workflow Description Language (WDL), which is platform-independent and follows open standards. Docker images and WDL files were submitted to Dockstore, a web service that provides bioinformatics tools and workflows. Described tools and workflows have been used by our Biostatistics and Computational Biology Laboratory (BCBLab) for performing RNA-seq, small RNA-seq, whole-genome bisulfite sequencing and exome sequencing data analyses. Validation tests showed that our workflows always generate the same results when executed in different computing environments. All tools and workflows are reusable reducing the effort to perform new analyses. Processing time was reduced since the entire analysis is performed automatically using optimized software executors capable of dynamically manage available computing resources. Disk space required to store genomic data was minimized because intermediary analysis files can be removed after execution and recovered whenever they are required. Docker and WDL files are available at <https://github.com/labcb/rnr>.

1597T

ReQTL: An allele-level measure of variation-expression genomic relationships. L.F. Spurr¹, N. Alomran¹, P. Slowinski², M. Li³, P. Bousounis³, Q. Zhang³, J. Sein⁴, K. Tsaneva-Atanasova^{2,4}, A. Horvath^{1,3,5}. 1) McCormick Genomics and Proteomics Center, School of Medicine and Health Sciences, The George Washington University, 20037, Washington, DC, USA; 2) Department of Mathematics & Living Systems Institute, University of Exeter, Stocker Road, Exeter, EX4 4QD, UK; 3) Department of Biochemistry and Molecular Medicine, School of Medicine and Health Sciences, George Washington University, 20037, Washington, DC, USA; 4) EPSRC Centre for Predictive Modelling in Healthcare, University of Exeter, Exeter, EX4 4QJ, UK; 5) Department of Pharmacology and Physiology, School of Medicine and Health Sciences, The George Washington University, 20037, Washington, DC, USA.

Through association of genotypes with gene expression levels, expression quantitative trait locus (eQTL) analyses have been instrumental in mapping of thousands of single nucleotide variants (SNVs) that affect the gene expression. We propose an eQTL modification, termed **ReQTL**, **RNA-eQTL**, which substitutes the DNA allele count with a continuous metric derived from the expressed variant allele frequency (VAF) at expressed SNV loci in the transcriptome ($VAF_{RNA} = n_{var} / (n_{var} + n_{ref})$), where n_{var} and n_{ref} are the variant and reference read counts estimated from the RNA sequencing data, respectively. In comparison to DNA allele count, VAF_{RNA} supports several distinct types of assessments. First, VAF_{RNA} provides phenotype- (as compared to genotype-) level information and is thus a close proxy for the actual allele content in the studied system. In addition, VAF_{RNA} can reveal allele-level expression regulation, including imprinting. Moreover, VAF_{RNA} can reflect expression regulation mediated by RNA-binding molecules and post-transcriptional events such as RNA-editing. Finally, as compared to DNA alleles, correlations of RNA alleles with gene expression provides several technical advantages. Because VAF_{RNA} constitutes a continuous measure (as compared to the DNA allele count), it offers increased measurability. Furthermore, since VAF_{RNA} and gene expression levels can be retrieved from a single source of transcriptome sequencing data, ReQTL analyses naturally control for measurement differences between samples. Related to the above, ReQTLs do not require matched DNA data and are thus directly applicable on the vast amount of existing transcriptome sequencing data across species and conditions. Our presented results on subsets of several GTEx tissue types show that ReQTL analyses: (1) are computationally feasible and can identify variation-expression relationships solely from RNA-sequencing data, (2) identify more than half of the eQTL-identifiable variants, and (3) identify additional SNV loci which we found to be enriched in RNA-editing sites. Given the quickly growing accessibility of RNA-sequencing data, ReQTLs have considerable potential to facilitate the discovery of novel molecular interactions. ReQTL toolkit is available from: <https://github.com/HorvathLab/ReQTL>.

1598F

The ENCODE pipeline architecture: Reproducible and portable analysis tools for ChIP-seq, RNA-seq, DNase-seq, ATAC-seq, HiC, ChIA-PET, and whole-genome bisulfite experiments. J.S. Strattan¹, U.K. Baymuradov¹, N.C. Durand², I. Gabdank¹, K. Graham¹, B.C. Hitz¹, O.A. Jolanki¹, J.W. Lee¹, E.T. Chan¹, J.A. Hilton¹, K. Onate¹, E.L. Aiden², A. Kundaje¹, J.M. Cherry¹, ENCODE Data Coordinating Center. 1) Stanford University School of Medicine, Department of Genetics, Stanford, CA; 2) Baylor College of Medicine, Department of Genetics, Houston, TX.

Consortia like ENCODE need uniform bioinformatics pipelines to avoid technical artifacts that can confound subsequent integrative analyses of experiments done in different labs or at different times. Such artifacts can be introduced when primary data are analyzed using different software versions, parameters, or run platforms. The ENCODE Data Coordinating Center (DCC) have developed a general containerization and deployment architecture for complex bioinformatics pipelines that produces portable, internally-validating compute environments that run identically in cloud, HPC, and local platforms for the analysis of core ENCODE experiments. Docker, Workflow Description Language (WDL), and continuous integration technologies are used to compose commonly-used bioinformatics tools and purpose-built code into portable, tested workflows. These analysis pipelines produce the uniform ground-level annotations ENCODE applies to the epigenome and transcriptome from ChIP-seq, RNA-seq, DNase-seq, ATAC-seq, HiC, ChIA-PET, and whole-genome bisulfite experiments. By virtue of their portability, both the pipelines and the pipeline architecture itself are of value to other consortia and investigators who need to compare their results to ENCODE's or deploy their own tools in portable ways. Through collaborations with the Roadmap Epigenomics project and the International Human Epigenomics Consortium (IHEC), the ENCODE pipelines are used to produce data and annotations that are compatible across consortia. The ENCODE DCC pipelines and architecture are free and open-source and are documented at <https://encode-dcc.github.io/wdl-pipelines/>. The DCC codebase is at <https://github.com/ENCODE-DCC>. ENCODE analyses are distributed through the ENCODE Portal at <https://www.encodeproject.org/>.

1599W

Wikidata for biomedical knowledge integration and curation. G.S. Stupp¹, A. Waagmeester², T.E. Putman³, L.M. Schrimp³, A.I. Su¹. 1) The Scripps Research Institute, San Diego, CA; 2) Micelio, Belgium; 3) University of Maryland, Baltimore, MD; 4) Oregon Health and Science University, Portland, OR.

The sum total of biomedical knowledge is accumulating at an explosive rate. Unfortunately, however, the entirety of that knowledge is not easily accessible. In most cases, biomedical knowledge is locked away in free-text research articles, which are very difficult to use for querying and computation. In some cases, that knowledge has been deposited in structured databases, but even then the fragmented landscape of such databases is a barrier to knowledge integration. Wikidata is an openly-accessible, community-maintained knowledge base that is editable by anyone and has a large, active user base. We have seeded Wikidata with data on key biomedical entities, including genes, proteins, diseases, drugs, genetic variants, and microbes. To ensure source databases are properly credited, we have implemented a standardized model for referencing and attribution. These data, combined with other data sets imported by the broader Wikidata community, enable powerful integrative queries that span multiple domain areas via the Wikidata SPARQL endpoint. In addition, we have designed a system to monitor, filter, and prioritize changes made by Wikidata contributors to biomedical items. The identified changes are coordinated with the primary resource's curation team through GitHub tickets, reviewed by a curator and integrated into the primary resource where appropriate. Through this process, in collaboration with the Disease Ontology, we have established a new crowdsourcing model for targeted curation and integration of knowledge into Wikidata and the DO. Furthermore, we have developed web applications (Wikigenomes and Chlambase) that facilitate the consumption and curation of genomic data by the entire scientific community, based on Wikidata, with the goal of aggregating published knowledge into a free and open database. WikiGenomes empowers the individual genomic researcher to contribute their expertise to the curation effort and integrates the knowledge into Wikidata, enabling it to be accessed by anyone without restriction. Wikidata can serve as a foundation for other bioinformatics applications and analyses and the code developed to execute this project is available as Open Source software. This suite of code includes modules for populating Wikidata, for automatically synchronizing with source databases, and for creating domain-specific applications to engage specific user communities.

1600T

Dissecting components of genetic associations across 2,138 phenotypes in the UK Biobank. Y. Tanigawa¹, J. Li², J. Justesen^{1,2}, H. Horn^{3,4}, M. Aguirre¹, C.M. DeBoever^{1,5}, M. Morri², C. Chang⁶, B. Narasimhan^{1,7}, G. Bejerano^{8,9,10}, K. Lage^{3,4,11}, T.J. Hastie^{1,7}, E. Ingelsson², M.A. Rivas¹. 1) Department of Biomedical Data Science, Stanford University, Palo Alto, CA; 2) Division of Cardiovascular Medicine, Department of Medicine, Stanford University School of Medicine, Stanford, CA; 3) Department of Surgery, Massachusetts General Hospital, Harvard Medical School, Boston, MA; 4) Broad Institute of MIT and Harvard, Cambridge, MA; 5) Department of Genetics, Stanford University, Stanford, CA; 6) Grail, Inc., 1525 O'Brien Drive, Menlo Park, CA; 7) Department of Statistics, Stanford University, Stanford, CA; 8) Department of Developmental Biology, Stanford University, Stanford, CA; 9) Department of Computer Science, Stanford University, Stanford, CA; 10) Department of Pediatrics, Stanford University School of Medicine, Stanford University, Stanford, CA; 11) Institute for Biological Psychiatry, Mental Health Center Sct. Hans, University of Copenhagen, Roskilde, Denmark.

Population-scale biobanks combined with genomic and health data are critical for generating effective therapeutic hypotheses and understanding the genetic role of disease predisposition. Pervasive pleiotropy and polygenicity, the presence of genetic effects of a variant across multiple phenotypes and multiple variants across a single phenotype, test our understanding of the genetics of human diseases. To address this, we apply an unsupervised statistical learning approach: matRix Decomposition of Genetic Effects (RIDGE), to matrices of summary statistics derived from genome-wide association analysis across 2,138 phenotypes in 337,199 White British individuals in the UK Biobank study. We systematically identify key components and quantify the contribution of variants, genes, and phenotypes. For example, we identify the first component is mainly driven by body mass measurements accounting for the "healthy part" (33%) including whole-body fat-free mass and genetic variants in *FTO* and *DLEU1* loci (1.9% and 0.64%) whereas the second component is driven by fat-related measurements (62%) and genetic variants in *FTO* and *ADCY3* loci (1.5% and 0.31%). To investigate the genetic architecture of complex traits, we apply RIDGE to body mass index (BMI), myocardial infarction, and gallstones. We find the top three key components for these traits combined explain 69%, 36%, and 41% of genetic effects, respectively. For each key component, we apply a variant of NetSig protein-protein interaction and GREAT genomic region analysis for mouse MGI phenotype ontology and demonstrate significant relevant enrichment. For example, we find an interacting network of proteins including *DLGAP1*, 2, and 3 (p-value < 2.5x10⁻⁵ each), brachypodia (MP:0002772, abnormally short feet, p-value = 1.3x10⁻²³), and abnormal pancreas topology (MP:0013861, p-value = 8.8x10⁻¹²) for the top key component for BMI. Finally, we identify three putative loss of function variants, rs114285050 (*GPR151*), rs150090666 (*PDE3B*), and rs137891647 (*LIPT1*), which contribute substantially to the components of genetic effects of obesity through RIDGE and PheWAS analysis. We demonstrate their effects on adipocyte formation by siRNA-mediated gene silencing followed by expression quantification of adipogenic markers with RT-qPCR. Our approach to dissect the components of genetic effects across human phenotypes will accelerate biomedical hypothesis generation by providing insights on latent structures previously unexplored.

1601F

OptimIR: An alignment workflow that integrates genetic variations in miRNA sequencing data. F. Thibord^{1,2}, C. Perret^{1,2}, M. Roux^{1,2}, P. Suchon^{3,4}, M. Ibrahim^{3,4}, L. Goumid⁵, M. Germain^{1,2}, J-F. Deleuze⁵, P-E. Morange^{3,4}, D-A. Trégouët^{1,2}, GENMED consortium. 1) Sorbonne Universités, Université Pierre et Marie Curie (UPMC Univ Paris 06), Institut National pour la Santé et la Recherche Médicale (INSERM), Unité Mixte de Recherche en Santé (UMR_S) 1166, Team Genomics & Pathophysiology of Cardiovascular Diseases, 75013 Paris, France; 2) Institute for Cardiometabolism and Nutrition (ICAN), 75013 Paris, France; 3) Laboratory of Haematology, La Timone Hospital, Marseille, France; 4) Institut National pour la Santé et la Recherche Médicale (INSERM), Unité Mixte de Recherche en Santé(UMR_S) 1062, Nutrition Obesity and Risk of Thrombosis, Center for CardioVascular and Nutrition research (C2VN), Aix-Marseille University, Marseille, France; 5) Centre National de Recherche en Génomique Humaine (CNRGH), Direction de la Recherche Fondamentale, CEA, Institut de Biologie François Jacob, Evry, France.

MiRNAs belong to a class of small (~22nt) non coding RNAs known to regulate post transcriptional expression and to associate with several human diseases. Because of its high detection sensibility and higher quantification accuracy, miRNA profiling by Next-Generation Sequencing (miRSeq) has gained recent interest to identify disease-associated miRNAs. A crucial step in the bioinformatics analysis of miRSeq data consists in aligning sequenced reads to a reference library of mature miRNAs. This step may be challenging because of the variability of mature miRNAs sequences due to post transcriptional editing mechanisms and genetic variations. We here propose a new bioinformatics workflow, OPTIMIR (for pOlymorPhism in Tegratlon for MiRNA data alignment), for miRSeq data alignment aimed at improving the accuracy of miRNA detection and quantification by simultaneously integrating the information on post-transcriptional editing and genetic variations data available in reference databases. In the context of miRSeq profiling applied to large cohorts of individuals, it is difficult to validate post-transcriptional editing induced variations. Conversely, the impact of genetic variations on miRSeq alignment can be assessed in individuals that have also been genotyped for DNA array. OPTIMIR was evaluated on a sample of 391 participants of the MARTHA study that have been profiled for plasma miRNAs and typed for genetic variants using Illumina genotyping array. In this sample, we identified 82 genetic variants located in miRNA sequence. The application of OPTIMIR to MARTHA samples revealed a 99.12% concordance between imputed homozygous genotypes and miRNA sequence deduced genotypes. Sanger sequencing validation of the discordant situations (8 events, i.e ~0.88%) identified one wrongly imputed genotype and suggested improper miRSeq data alignment for the 7 other cases. For heterozygous variants, our analysis revealed that both alleles are not equally expressed. Most of the miRNAs harboring a polymorphism in their sequence preferentially expressed one of the two alleles, not necessarily the most frequent one. Very few miRNAs had a balanced expression of both isoforms. In conclusion, using OPTIMIR, we showed that the detection/quantification of miRNAs by miRSeq are hardly impacted by the presence of genetic variations in the miRNAs sequences. In addition, we suggested that allele specific-expression of miRNAs could also hold as previously reported for gene expression.

1602W

TogoVar: A comprehensive Japanese genetic variation database including 183,884 SNP-array and 125 whole exome sequence based data. L. Toyo-oka¹, N. Mitsuhashi¹, M. Kawashima¹, T. Fujiwara², S. Katayama², S. Kawashima², Y. Tateishi¹, T. Takagi^{2,3}. 1) National Bioscience Database Center, Japan Science and Technology Agency, Tokyo, Japan; 2) Database Center for Life Science, Joint Support-Center for Data Science Research, Research Organization of Information and Systems, Japan; 3) Department of Biological Sciences, Graduate School of Science, the University of Tokyo.

To conduct a genetic analysis, it is essential to have amount of well-qualified genetic variation data in the targeted population. One of the successful contributions to the human genetics is the Exome Aggregation Consortium (ExAC) which has been providing quality controlled allele frequency information in several populations. However, for Japanese population, the Human Genomic Variation Database (HGVD) and the Integrative Japanese Genome Variation Database (iJGVD) are mainly used because the number of Japanese population in the ExAC is limited. On the purpose of providing the amount of allele frequency data of Japanese and compare with that of different populations, we launched the TogoVar database (<https://togovar.biosciencedbc.jp/>). Two new Japanese allele frequency datasets are included to the database in addition to HGVD, iJGVD, ExAC and ClinVar data; (1) JGA-SNP, consisted of 183,884 samples genotyped by SNP-array method and (2) JGA-NGS, consisted of 125 samples produced by whole exome sequencing method. Both of these datasets are constructed from the data deposited to the NBDC Human Database, which is a repository for sharing Japanese genomic data in cooperation with the DNA Data Bank of Japan since 2013. Because about half of variants derived from JGA-SNP, JGA-NGS, HGVD, iJGVD, ExAC and ClinVar did not have the corresponding rs number, another identifier independent from genomic assembly was required. Thus we assign a new identifier named TogoVar ID to all the genetic variants. The total number of assigned genetic variants is up to 67 million. After excluding the assigned genetic variants with low allele frequency due to avoiding individual identifiability, more than 19 million unique variants are included in the public TogoVar dataset. The TogoVar web site provides a query interface for searching variants by a genomic position or region, a TogoVar or dbSNP ID, a gene name, or a disease name with auto-completion. Then user can narrow down the result set by filtering variants by datasets, allele frequencies, variant types and clinical significances. Detailed information of each variant is shown in the variant report page with the corresponding allele frequency data by population, genetic information annotated by Variant Effect Predictor (VEP), citations of relevant literature from PubMed and clinical significance from ClinVar.

1603T

ViVa (Visualization of VArIants):VCF (Variant Call Format) file visualization tool. A. Uzun^{1,2,3}, G. Tollefson¹, J. Schuster^{1,2}, A. Ragavendran⁴, I. Restrepo⁴, F. Gelin⁴, P. Stey⁴, J. Padbury^{1,2,3}. 1) Dept. of Pediatrics, Brown University Warren Alpert Medical School, Providence, RI; 2) Dept. of Pediatrics, Women and Infants Hospital, Providence, RI; 3) Center for Computational and Molecular Biology, Brown University, Providence, RI; 4) Data Science Practice, Brown University, Providence, RI.

ViVa is a software tool designed to visualize data from VCF (Variant Call Format) files. A VCF file is a text file format that contains information about the genome. The file contains data about sequenced samples represented in columns and gene variant positions represented in rows. It is difficult to draw conclusions from this densely filled data structure without writing computer scripts. Because whole genome sequencing is becoming increasingly accessible to researchers and clinicians due to rapidly lowering costs of Next Generation Sequencing technologies, the need to easily extract insights from VCF files is growing in demand. We present a tool that flexibly generates publication quality graphics for the purposes of variant analysis and quality control of sequencing experiments. We offer users flexibility, efficiency, and ease of use compared with similar existing tools. In order to achieve this, we use the Julia programming language: a high-level, high-performance, dynamic programming language for numerical computing. Users can view trends in VCF fields such as sequencing read quality or genotype values across all samples and variant positions to identify regions of interest in an experiment. ViVa is designed to visualize millions of variant positions on the genome-wide to chromosome-region scales and is supported by rigorous filtering options. In addition to generating publication quality, scalable visualizations of data, users can manipulate and export filtered VCF files for downstream analysis. ViVa is comprised of three elements: a command line executable for bioinformaticians, an interactive user interface for clinicians and non-programmers, and a Julia language package for developers to collaborate. Additional plotting features include generation of line plots of read depth across both samples and variant positions. For instance, users can visualize read depth across all samples at a particular gene variant position. This allows analysts to remove samples with poor sequencing quality from queries involving discrete sets of genes. Similarly, users can visualize average read depth across all variant positions to identify poorly captured regions that may be excluded from further analysis.

1604F

Bayesian hierarchical modeling of clustered or longitudinal RNA sequencing experiments. B.E. Vestal¹, C.M. Moore², T. Fingerlin¹, L. Saba³, K. Kechris³. 1) Center for Genes, Environment and Health, National Jewish Health, Denver, CO; 2) Department of Biostatistics, National Jewish Health, Denver, CO; 3) Department of Biostatistics and Informatics, University of Colorado Denver, Denver, CO.

The continued increase in accessibility to RNA sequencing (RNA-Seq) technology has led to more complicated study designs that demand analysis methods beyond the scope of what current methods were designed to handle. For example, studies that take repeated measurements of subjects over time would typically require Generalized Linear Mixed Models (GLMMs) or Generalized Estimating Equation methods for robust analysis. The most popular analysis tools for RNA-Seq data, edgeR and DESeq2, are designed for use on studies that only include fixed effects, and thus do not contain the machinery to account for the correlation between repeated measurements/clustered samples or other random effects. One solution that has been suggested is to include individual subjects as fixed effects when using edgeR or DESeq2 when there are repeated measurements on individuals, but this strategy can lead to complications with estimating specific parameters (e.g. a difference between two groups is not estimable in this scenario). In this work, we propose using a Bayesian hierarchical model for analyzing RNA-Seq data that will naturally allow for the inclusion of random effects. Since RNA-Seq data is quantified as counts, we use the commonly chosen Negative Binomial (NB) distribution with a log link to model the observed data. Model parameters are estimated using Markov Chain Monte Carlo (MCMC) methodologies with a Weighted Least Squares proposal distribution for better mixing of regression parameters (available in the MCMSeq R package). Additionally, by including the NB dispersion parameter in the MCMC sampling, the regression coefficients and their credible intervals take into account the fact that the dispersion is also being estimated. Both edgeR and DESeq2 treat this value as fixed and known when testing, and this has been posited to contribute to hypothesis tests on regression coefficients not attaining their nominal type I error rate for these methods. We compare the MCMC results to edgeR, DESeq2, and traditional generalized linear mixed models (GLMMs) in terms of power, type I error rates, false discover rate (FDR), and mean squared error (MSE) of regression coefficients for data sets simulated under a GLMM framework (paired observations per individual). Preliminary results show that the Bayesian model better controls type I errors and FDR, while edgeR, DESeq2, and the traditional GLMM fits show inflated false positive rates and inflated FDRs.

1605W

Structural variation detection and validation pipelines for whole genome sequencing association studies and diagnostic testing. H.Z. Wang¹, B.B. Currall^{1,2,3}, M.R. Stone¹, X. Zhao^{1,2,3}, R.L. Collins^{1,3,4}, C. Antolik¹, J.T. Glessner⁵, M.J. Waterman⁶, M.E. Talkowski^{1,2,3,7}, H. Brand^{1,2,3}. 1) Center for Genomic Medicine and Department of Neurology, Massachusetts General Hospital, Boston, MA; 2) Department of Neurology, Harvard Medical School, Boston, MA; 3) Program in Medical and Population Genetics and Stanley Center for Psychiatric Research, Broad Institute, Cambridge, MA; 4) Program in Bioinformatics and Integrative Genomics, Division of Medical Sciences, Harvard Medical School, Boston, MA; 5) The Center for Applied Genomics, Abramson Research Center, Children's Hospital of Philadelphia, PA; 6) Department of Biology, Eastern Nazarene College, Quincy, MA; 7) Departments of Pathology and Psychiatry, Massachusetts General Hospital, Boston, MA.

Detection of structural variation (SV) from whole-genome sequencing (WGS) remains highly inconsistent across studies with variable reporting of counts, SV classes, functional annotations, and validation methods. Numerous signatures of SV can be captured from WGS: anomalous paired-end (PE) reads signifying a rearrangement, split reads (SR) crossing a breakpoint, and changes in read depth (RD) indicating alterations in DNA dosage. Few algorithms jointly consider all signatures, hampering accurate variant discovery. We have developed a multi-stage SV detection and classification pipeline (<https://github.com/talkowski-lab/SV-Adjudicator>), which incorporates an array of established PE/SR and RD algorithms to discover an extensive list of candidate SV loci, then adjudicates each predicted variant with a joint statistical analysis across all samples. Candidate copy number variant (CNV) intervals are verified by testing for statistically significance differences in RD between samples predicted to harbor a CNV and those individuals that are diploid for the segment. Putative SV breakpoints are also assessed with an equivalent test of PE/SR evidence. The refined SVs from each algorithm are then clustered and stitched together to resolve both balanced and complex SV events based on 20 previously identified signatures. This method routinely identifies ~6000 SVs per individual and 0.2 *de novo* SVs outside of repetitive sequences and multi-allelic CNV copy gains. We have further developed a high throughput validation approach (BPP3) that can test single nucleotide variants (SNVs), indels, and SV rearrangements detected from WGS. PCR primers are designed via a custom script that integrates Primer3, *in silico* PCR and BLAT to implement strand and orientation aware primers. Our script automatically chooses the appropriate validation strategy according to the type and size of variant, then iteratively adjusts constraints to arrive at optimal experimental conditions, and QCs the resulting primer set to yield the best possible primer pair (<https://github.com/talkowski-lab/bpp3>). PCR products are ligated with Illumina adapters, pooled and then paired-end sequenced at 250 bp on an Illumina MiSeq. Sequencing reads are aligned to with BWA-MEM and variants are screened with bam-readcount, Vardict, blat, and bamstat, followed by manual confirmation in IGV. All PCR reactions are performed on 96-well plates to permit efficient scaling to hundreds of variants.

1606T

SeQuilLa: An elastic, fast and scalable SQL-oriented platform for processing and analyzing genomic data. M. Wiewiórka, A. Szmurlo, T. Gambin. Institute of Computer Science, Warsaw, Poland.

Efficient processing of massive amount of genomic data generated by biological assays became possible thanks to the application of big data technologies in bioinformatics pipelines. Traditionally, next generation sequencing (NGS) secondary analysis pipelines, have been designed and performed in a file-centric manner, resulting in multiple drawbacks, including but not limited to inefficient data processing and re-analysis, lack of fine-grained security, and inconsistent data representation. Our solution combines distributed computing engine based on extended Apache Spark Catalyst query optimizer with a Structured Query Language (SQL) interface for handling large-scale processing and analysing of NGS datasets in a consistent tabular form. It facilitates the following aspects of genomics data handling: Definition of Apache Spark compatible datasources that map the existing file formats such as BAM, ADAM or VCF into easy to query unified table abstractions. Data access to both single and multiple samples is optimized using partition and column pruning strategies. It makes combining results from many stages of bioinformatics pipelines both straightforward and efficient. Genomic interval queries which are crucial to many bioinformatics analyses were optimized through a distributed implementation of interval forests that can be expressed in a form of regular SQL range joins which are automatically transformed into the most efficient physical execution plan using input datasets statistics. Coverage computations are also included in a SQL syntax using table-valued functions and optimized by a distributed implementation that eliminates the unnecessary data shuffling between computing nodes. User defined functions including shift, resize, flank that streamline genomic records operations and parameters maxGap and minOverlap for additional overlap configuration. Straightforward interoperability with existing analytics tools using a widely-adopted JDBC standard. To test the efficiency of our platform, we performed multiple tests on both a single node and on a cluster. In particular, we demonstrated that our range join strategy is significantly (~22x) faster than Apache Spark one and outperforms other state-of-the-art tools for genomic intervals processing. The project documentation and source code is available at <http://biodatageeks.org/sequilla/>.

1607F

A high-throughput computational quality control pipeline for exome and genome data for rare disease studies. *M. Wilson^{1,2}, G. Tiao^{1,2}, K. Laricchia^{1,2}, L. Francioli^{1,2}, K. Karczewski^{1,2}, B. Weisburd^{1,2}, D. MacArthur^{1,2}.* 1) The Broad Institute of MIT and Harvard, Cambridge, MA; 2) Massachusetts General Hospital, Boston, MA, United States.

Every year over 500,000 babies are born with a Mendelian disorder, but current diagnostic approaches provide a diagnosis for less than half of all patients with genetic disease. To increase diagnostic rates for patients with rare Mendelian disorders, we have developed the software platform *seqr*, which allows analysts to efficiently and systematically search for candidate disease-causing variants in patients' sequencing data. By providing multiple resources in an easy-to-use interface, *seqr* enables users to search variant calls by filtering on variant type, functional class, allele frequency, allele count, inheritance mode, deleteriousness predictions, and genomic location to discover variants possibly related to a Mendelian disease. Standard analysis procedures, such as identifying familial inheritance patterns, prioritizing candidate genes, filtering variant calls, and querying online databases and papers typically take days to complete but can be accomplished in minutes within the *seqr* interface, allowing analysts to review and identify more candidate disease-causing variants with less effort. Successful searches for causal disease variants in *seqr* depend on the quality of the sequencing data and the accuracy of sample annotations, including labeling of familial relationships and individual metadata. To ensure data quality and to increase the probability of variant discovery and diagnosis, we have built a computational quality control pipeline for *seqr* using Hail (<https://github.com/hail-is/hail>), an open-source, scalable framework for exploring and analyzing genomic data. The quality control pipeline is written entirely in Python using the Hail library and capitalizes on Hail's ability to handle large genomic datasets for efficient processing. The pipeline flags samples with elevated rates of contamination, chimeric reads, and low coverage. Sex, exome capture platform, and ancestry are all inferred and checked against available sample metadata supplied by clinicians. The pipeline also provides a direct estimate of consanguinity and reconstructs family pedigrees computationally, allowing analysts to check family structures against reported pedigrees. By providing analysts with quality sequencing data and accurate sample metadata, we are actively working to increase the diagnostic rates of rare Mendelian disease.

1608W

Determining sanger confirmation necessity of NGS variants using a data-driven model. *C. Wu¹, K.B. Pechter¹, E.J. Romasko¹, X. Zhao¹, M.A. Diaz-Miranda¹, C.U. Silva², K. Cao¹, M.C. Dulik^{1,3}, A.N. Abou Tayoun^{1,3,4}, M. Sarmady^{1,3}.* 1) Division of Genomic Diagnostics, The Childrens Hospital of Philadelphia, Philadelphia, PA, USA; 2) GenoMed Inc., Instituto De Medicina Molecular, University of Lisbon, Lisbon, Portugal; 3) Department of Pathology and Laboratory Medicine, The University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA; 4) Al Jalila Childrens Specialty Hospital, Dubai, UAE.

Next generation sequencing (NGS) is increasingly the preferred technology for sequence-based clinical genetic tests. The quality of NGS calls may vary widely, with a small fraction of the detected variants needing orthogonal confirmation despite the majority of the calls being accurate and reliable. Orthogonal confirmation can be costly and increase turnaround time for tests. Therefore, many clinical labs are moving towards a quality control strategy to identify only the 'low quality' variants that require orthogonal confirmation. Here, we report the development of our bioinformatics method to differentiate between these two groups of variant calls: 1) not requiring orthogonal confirmation (high confidence) and 2) requiring additional testing (low confidence). Since the vast majority of the detected mutations are single nucleotide variants (SNVs), we derived a set of rules incorporating several features of the characteristics and call quality to determine if an SNV was detected at high or low confidence. Here, a true positive is defined as a variant called in NGS and confirmed by Sanger sequencing, whereas a false positive is a variant called in NGS yet rejected by Sanger. The challenge is to classify all of the false positive calls as 'low confidence', while categorizing most true positive calls as 'high confidence'. We studied 474 SNVs detected by clinical NGS exome capture, including 460 SNVs that were Sanger confirmed and 14 Sanger rejected, and extracted sequencing and quality attributes of each of the variants. Then we performed feature selection based on information entropy to remove features that did not separate the two classes well. Finally, we derived a set of rules to determine the confidence of a detected SNV. As a result, we label a SNV as 'low confidence' if any of the criteria is met: Quality by Depth (QD) < 11; Phred-scaled p-value using Fisher's exact test (FS) > 5; Allele Ratio (AF): Homozygous < 67% and Heterozygous < 33%. Applying this set of rules, 281 out of the 460 true positives are labeled as 'high confidence' (61.1%) while all 14 false positives are labeled as 'low confidence' (100%). To further investigate its clinical utility, we performed additional validation on 'borderline' variants and consolidated the set of criteria. In summary, we present a bioinformatics approach to accurately identify SNVs requiring additional confirmation using quality metrics for clinical implementation to reduce the need for confirmatory testing.

1609T

Performance evaluation of assembly-based structural variation discovery in the human genome. C. Xiao¹, H. Hong², *SEQC2 Germline Working Group*. 1) NCBI/NIH, Bethesda, MD 20894; 2) NCTR/FDA, 3900 NCTR Road, Jefferson, AR 72079.

Structural variations (SVs) are widely known to contribute to genetic diversity of human populations, affect biological functions, and cause various human disorders. However, accurately identifying SVs with correct sizes and locations in the human genome remains a challenge due to the complexities of the human genome, limitations of sequencing technologies, and drawbacks of analysis methods. Recent advancement of next-generation sequencing technologies has significantly reduced the sequencing cost, while substantially increased the lengths of the sequencing reads. Therefore, using *de novo* assembly based approaches for discovering a full spectrum of SVs in human genome becomes appealing. While various assembly methods have been developed and proposed for general use by the community, the relative efficiency and predictive accuracy of SVs calling based on these assembly methods have not been fully evaluated. In this study, we applied several popular *de novo* assembly tools to the sequencing read data that were generated using multiple sequencing technologies with technical replicates for NA12878/HG001 and a Chinese Quartet. Assemblies and SVs callsets were generated for each of the five samples, and repeatability in the SVs of the technical replicates and reproducibility across sequencing sites were evaluated. Preliminary results showed that there were substantial variabilities between SVs callsets based on various assemblies within sequencing platform. We also observed considerable variabilities between assembly-based SVs callsets and alignment-based SVs callsets. The sources of these variabilities will be further discussed. These results allow better understanding of the impacts of *de novo* assembly methods on SVs calling, thus providing a better insight to precision medicine.

1610F

CRISPRO identifies functional protein coding sequences based on genome editing dense mutagenesis. Q. Yao^{1,2,3,4}, V.A.C. Schoonenberg^{1,2,3}, M.A. Cole^{1,2,3}, C. Macias-Trevino^{1,2,3}, F. Sher^{1,2,3}, P.G. Schupp^{1,2,3}, M.C. Canver^{1,2,3}, T. Maeda⁵, L. Pinello^{3,4}, D.E. Bauer^{1,2,3}. 1) Boston Children's Hospital, Boston, MA; 2) Dana-Farber Cancer Institute; 3) Harvard Medical School, Boston, MA; 4) Massachusetts General Hospital, Boston, MA; 5) Kyushu University Hospital, Fukuoka, Japan.

CRISPR/Cas9 pooled screening permits parallel evaluation of comprehensive guide RNA libraries to systematically perturb protein coding sequences *in situ* and correlate with functional readouts. For the analysis and visualization of the resulting datasets we have developed CRISPRO, a computational pipeline that maps functional scores associated with guide RNAs to genome, transcript, and protein coordinates and structure. No available tool has similar functionality. The ensuing genotype-phenotype linear and 3D maps raise hypotheses about structure-function relationships at discrete protein regions. Machine learning based on CRISPRO features improves prediction of guide RNA efficacy. The CRISPRO tool is freely available at gitlab.com/bauerlab/crispro.

1611W

Cell type clustering tools evaluated in six single cell chromatin accessibility experiments. *E. Urrutia¹, L. Chen², Y. Jiang¹*. 1) University of North Carolina, Chapel Hill, NC; 2) Auburn University, Auburn, AL.

Single-cell assay of transposase-accessible chromatin followed by sequencing (scATAC-seq) is an emerging new technology for the study of gene regulation with single-cell resolution. Unlike the conventional ChIP-seq, DNase-seq and ATAC-seq technologies, scATAC-seq measures regulatory element activities within each individual cell, which circumvents the averaging artifacts associated with traditional bulk population data, yielding new insights into the cellular heterogeneity. The data from scATAC-seq are unique – sparse, binary, noisy with biases and artifacts, and highly variable even within cell types. As such, neither methods developed for bulk ATAC-seq nor single-cell RNA-seq are appropriate. Here, we propose Destin, a bioinformatic and statistical framework for comprehensive scATAC-seq data analysis. We evaluate performance of Destin using scATAC-seq data of 19,000 cells from six experiments. The experiments show diversity, ranging from 500 to 3,000 cells, from 100,000 to 150,000 accessible regions, and from 2 to 12 cell types. The experiments also span multiple technologies, including one based on physical isolation of single cells using microfluidics, and one by a two-step combinatorial indexing strategy that avoids single-cell reaction volumes. Destin is benchmarked against seven existing methods using the known cell identities presented by the original authors and is shown to have the best performance across all datasets. Specifically, we show that proper normalization for sequencing depth accounted for a 25-50% improvement in clustering in every scATAC-seq experiment. Furthermore, Destin incorporates weights for chromatin accessible regions based on previous annotations and bulk epigenomic profiles, which (i) give higher weights to long-range regulatory elements (i.e. enhancers, repressors/silencers and insulators, while down-weighting accessible promoter regions) and (ii) up-weight rare accessible regions and down-weight common accessible regions as determined by previous bulk chromatin-accessibility experiments. We show that incorporating the above weights to our recommended best practice resulted in clustering accuracy at least as strong as unweighted in every scATACseq experiment and accounted for up to a 15% improvement. .

1612T

Differences in GWAS findings driven by using ethnically-matched and mixed imputation reference panels. *M. Kals^{1,2}, K. Läll^{1,2}, T.T. Sikka¹, A. Paltie^{3,4}, A. Metspalu^{1,5}, P. Palta^{1,3}, R. Mägi¹, SISu Project*. 1) Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu, Estonia; 2) Institute of Mathematics and Statistics, University of Tartu, Tartu, Estonia; 3) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; 4) Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; 5) Department of Biotechnology, Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia.

Over last decade genome-wide association studies (GWAS) have been most frequent tool to identify genes involved in complex diseases. Input of GWA studies requires high-quality genetic imputation and imputation accuracy depends on underlying imputation reference panel (IRP). Although genotype imputation can be performed relatively well for common and low-frequency variants with publicly available ethnically more heterogeneous IRPs like the 1000 Genomes Project (1000GP) and the Haplotype Reference Consortium (HRC), the imputation of rare variants has been shown to be significantly more accurate when the used reference panel and target dataset are ethnically matched or if population-specific reference panel is used for genotype imputation. Notwithstanding, the genome-wide downstream differences and consequences of using international mixed and ethnically matched reference panels have not been yet comprehensively explored. In this study we determined and quantified these differences by performing several comparative evaluations of the following biologically motivated analysis scenarios by using genome-wide genotype array data of ~52,000 Estonians imputed with both ethnically-matched and 1000GP imputation reference panel. Firstly, a variant-wise GWAS of confidently imputed (INFO > 0.8) variants was performed on seven complex diseases (bipolar disorder, Crohn's disease, coronary artery disease, hypertension, rheumatoid arthritis, type 1 diabetes, and type 2 diabetes). A logistic regression model using Firth's bias reduction method was applied to model genotype dosages. Although several previously reported common variant associations were replicated in both imputed datasets, no major differences were observed in the results of variant-wise GWAS or in fine-mapping effort of associated loci. Secondly, a gene-wise GWAS was performed using SKAT-O test to determine the joint contribution of rare (MAF < 1%) nonsynonymous variants. In this analysis, significant differences were observed in favor of ethnically-matched imputation, where 10-fold statistically significant genes were discovered as compared to the 1000GP-based imputation. Our results clearly demonstrate that although there are no major differences in common variant imputation and consequent GWAS findings, the usage of ethnically-matched imputation reference panel can enable superior imputation of rare variants, facilitating significantly higher number of findings in gene-based rare variant GWAS scans.

1613F

Clinotator: Analyzing ClinVar variation reports to prioritize reclassification efforts. R.R. Butler^{1,2}, P.V. Gejman^{1,2}. 1) Genomic Health Initiative, NorthShore University HealthSystem, Evanston, IL; 2) Department of Psychiatry and Behavioral Neuroscience, Pritzker School of Medicine, University of Chicago, Chicago, IL.

While ClinVar has become an indispensable resource for clinical variant interpretation, its sophisticated structure provides it with a daunting learning curve. Often the sheer depth of types of information provided can make it difficult to analyze variant information with high throughput. Clinotator is a fast and lightweight tool to extract important aspects of criteria-based clinical assertions; it uses that information to generate several metrics to assess the strength and consistency of the evidence supporting the variant clinical significance. Clinical assertions are weighted by significance type, age of submission and submitter expertise category to filter outdated or incomplete assertions that otherwise confound interpretation. This can be accomplished in batches: either lists of Variation IDs or dbSNP rsIDs, or with vcf files that are additionally annotated. Using sample sets ranging from 15,000–50,000 variants, we slice out problem variants in minutes without extensive computational effort (using only a personal computer) and corroborate recently reported trends of discordance hiding amongst the curated masses. With the rapidly growing body of variant evidence, most submitters and researchers have limited resources to devote to variant curation. Clinotator provides efficient, systematic prioritization of discordant variants in need of reclassification. The hope is that this tool can inform ClinVar curation and encourage submitters to keep their clinical assertions current by focusing their efforts. Additionally, researchers can utilize new metrics to analyze variants of interest in pursuit of new insights into pathogenicity.

1614W

Prediction of missense variant function polarity using machine learning.

A. Ghosh^{1,2,3}, A. Navarini^{1,2,3}. 1) Department of Dermatology, University Hospital Zurich, Zurich, Switzerland; 2) Faculty of Medicine, University of Zurich, Zurich, Switzerland; 3) Competence Center Personalized Medicine, University of Zurich and ETH Zurich, Zurich, Switzerland.

Introduction: Functional interpretation is crucial when facing on average 20,000 missense variants per human exome, as the great majority are not associated with any underlying disease. *In silico* bioinformatics tools can predict the deleteriousness of variants or grade their functional impact by assigning scores such as SIFT, PolyPhen2 and CADD. In addition, consensus scores such as Condel aim to further refine the predictions by aggregating weighted single scores. For functional experiments seeking to confirm biological relevance, an as yet unanswered question is whether a missense variant in question results in loss or gain of function at the protein level. **Objectives:** To develop a new classifier for damaging genetic variants allowing to predict their gain or loss of function polarity. **Material and Methods:** 1,543 loss of function (LOF) and 518 gain of function (GOF) variants with supporting evidence in terms of literature texts, clinical reports and research projects were extracted from ClinVar. All damaging variants as annotated by Condel and CADD were selected to develop the classifier. Boosting algorithm was preferred over other machine learning algorithms because of its additive model approach. **Results:** Our results showed two peaks of mean Condel scores across GOF and LOF variants ($p < 0.01$) indicating a non-random difference. The classifier adapted a weighted boosting approach to address the variant class imbalance and therefore ROC curve was used to assess its performance. The discriminative power of the tool was found to be very high with an area under the ROC curve of >0.82 . **Discussion:** Function polarity prediction was found to be possible using available prediction algorithms' data in this new classifier. Even though no prediction has perfect accuracy, these results may help to guide biological experiments on the clinical relevance of damaging genetic variants. The algorithm, named BioPol, is available online. A weakness of our approach is that the great majority of available missense variants have not been investigated with biological experiments.

1615T

DeepSweep: A novel neural model to localize signals of positive selection. S.J. Gosai^{1,2}, J.J. Vitti^{2,3}, S.K. Reilly^{2,3}, P.C. Sabeti^{2,3}. 1) Department of Biological and Biomedical Sciences, Harvard University, Boston, Massachusetts 02115, USA; 2) Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; 3) Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts, USA.

As humans migrated out of Africa over the past 200,000 years and encountered novel environments, diets, and pathogens, distinct sub-populations have been exposed to various selective pressures. The emergency of whole-genome sequences from diverse human populations has allowed us to search for the adaptive human variants that are a response to these selective pressures. However, only a handful of putative adaptive variants (PAVs) have been convincingly linked to evolutionarily advantageous traits. Progress has been hindered by the challenges of pinpointing PAVs amongst numerous variants in linkage disequilibrium (LD), meaning humans neutral variants almost certainly drastically outnumber adaptive ones. Methods to precisely identify variants under selection by minimizing false discovery rate would drastically reduce these challenges. Here, we present deepSweep, a deep neural network used to localize signals of selection in humans to individual variants. Previous methods are able to localize signal to large loci approximately 100kb in size with tens to hundreds of variants. However our model achieves >70% top-1 and >95% top-10 accuracy when searching for individual adaptive variants with derived allele frequency (DAF) >20% in simulated hard-sweeps. Moreover, our model is able to detect simulated adaptive variants with DAF as low as 10%, albeit at reduced sensitivity and selectivity. Our model is trained to identify drivers of both “soft- sweeps” and “hard- sweeps” and can be directly applied to data generated from any population. To our knowledge, this is the first deep-learning approach to scans for adaptive variation in humans. We used this tool in conjunction with Composite of Multiple Signals (CMS) method for Bayesian identification of regions under positive natural selection to scan the 1000 Genomes Phase 3 dataset. Using a two tiered approach, we identify 650 regions with evidence of positive evolutionary selection and identify ~65000 candidate variants for further analysis.

1616F

The Monarch HIPPO: Deriving insight from the medical literature by fuzzy semantic searches over diseases and phenotypes. T. Groza¹, A. Muaz¹, A. Azarkhish¹, P.N. Robinson², C.J. Mungall³, N.L. Harris³, J.A. McMurry⁴, M.A. Haendel⁵. 1) Garvan Institute of Medical Research, Darlinghurst, NSW, Australia; 2) The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA; 3) Lawrence Berkeley National Laboratory, Berkeley, CA, USA; 4) Department of Medical Informatics and Clinical Epidemiology, Oregon Health and Science University, Portland, OR, USA; 5) Dept of Environmental and Molecular Toxicology, Oregon State University, Corvallis, OR, USA.

It takes a clinician substantial time to search for publications relevant to a given patient case, and such searches are often unsuccessful. For example, using PubMed to find all publications that might mention genes responsible for “macrocephaly” and “developmental delay” requires tedious review of numerous abstracts in order to surface those that actually mention a gene relevant to a patient’s diagnosis. Most available full-text search applications provide basic string matching of full text; they usually also leverage what little curated meta-data exists (e.g., MeSH terms). However, to further improve literature search it is necessary to mine specific entities from the text and to leverage what is already known about how they are related to each other. We have developed a high-throughput pipeline to extract clinically relevant knowledge from the literature and present it to users in novel ways. The new application, HIPPO (hippo.monarchinitiative.org) enables researchers to perform fine-grained searches for abstracts mentioning concepts of interest: phenotypes (from Human Phenotype Ontology), diseases (MONDO) and genes (HGNC), and soon for drugs (DrugBank), variants (in HGVS notation) and signs and symptoms (SNOMED-CT). Because these entities are also structured in a hierarchy with synonyms, queries can find all relevant records. e.g., “brain cancer” will automatically retrieve any records that mention “glioblastoma” even if they do not mention “brain” or “cancer”. Similarly, a search for FANCD1 also retrieves abstracts for BRCA1 as these are synonymous. Moreover, the recognized entities in HIPPO are linked out to the corresponding page in the main Monarch application so that users can explore the landscape of published associations between an entity of interest and other diseases, genes, and phenotypes. We are crowdsourcing manual curation of each text-mined entity via a simple “thumbs-up/thumbs-down” voting integrated within the user interface. Finally, the entire pipeline is re-run on a weekly basis, pulling not only the latest abstracts, but also the latest lists of genes, phenotypes, diseases etc. We have also introduced a query subscription service for users to track phenotypes and other related entities of interest. HIPPO can help clinicians find what they need to help diagnose patients while staying updated with the latest developments for their query of interest.

1617W

ForestQC: Variant quality control based on random forest model. J. Li¹, S.K. Service², B. Jew², L. Zhan⁴, S. Hwang², G. Coppola², N.B. Freimer^{1,2}, J.H. Sul¹. 1) Department of Human Genetics, University of California, Los Angeles, Los Angeles, CA; 2) Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles, Los Angeles, CA; 3) Bioinformatics Interdepartmental Program, University of California, Los Angeles, Los Angeles, CA; 4) Molecular Biology Institute, University of California, Los Angeles, Los Angeles, CA.

Next generation sequencing (NGS) technology enables discovery of nearly all genetic variants present in a genome. A subset of these variants, however, may have poor sequencing quality due to biases and errors in sequencing technologies, alignment biases caused by inconsistent mappability of reads, or inaccurate and discordant variant calling algorithms. For example, concordance among different variant callers was 57.4% in SNVs detection and 26.8% in indels discovery. In genetic studies that analyze a large number of sequenced individuals, it is critical to detect and remove those variants with poor quality as they may cause spurious findings. In previous studies, either filtering approaches or machine learning approaches are used for variant quality control (QC). In this paper, we developed a statistical approach for performing QC on variants identified from NGS data by combining a filtering approach and a machine learning approach. Our method uses information on sequencing quality such as sequencing depth, genotyping quality, and GC contents to predict whether a certain variant is likely to contain errors. We applied our method to two different whole-genome sequencing (WGS) datasets whose study designs include family and unrelated individuals. The first WGS dataset consists of 449 individuals from a large extended pedigrees ascertained for bipolar disorder (BP). It has about 25 million SNVs and 4 million indels, whose average coverage is about 40X. The second WGS dataset is based on a sequencing study for progressive supranuclear palsy (PSP). It includes 495 unrelated subjects who were sequenced at average coverage of 29X, generating about 33 million SNVs and 5 million indels. Results indicate that our method outperforms widely used methods for variant QC such as GATK-VQSR by considerably lowering Mendelian error rate, missing rate and genotype discordance rate and improving Ti/Tv ratios of variants to be included in the analysis. Our approach is very efficient that it takes 20-30 minutes to finish processing the two WGS datasets, which is 15 to 20 times faster than GATK-VQSR. Hence, it is scalable to large-scale sequencing datasets. We conclude that combining a machine learning algorithm trained with sequencing quality information and the filtering approach is an effective approach to perform QC on genetic variants detected from sequencing.

1618T

DECA: Scalable XHMM exome copy-number variant calling with ADAM and Apache Spark. M. Linderman¹, D. Chia¹, F. Wallace¹, F. Nothhaft^{2,3}. 1) Department of Computer Science, Middlebury College, Middlebury, VT; 2) AMPLab, University of California, Berkeley, Berkeley, CA; 3) Databricks, Inc., San Francisco, CA.

Background:XHMM is a widely used tool for copy-number variant (CNV) discovery from whole exome sequencing (WES) data, but can require hours to days to run for large cohorts. A more scalable implementation would reduce the need for specialized computational resources and enable increased exploration of the configuration parameter space to obtain the best possible results. **Methods and Results:**DECA is a horizontally scalable implementation of the XHMM algorithm using the ADAM framework and Apache Spark that incorporates novel algorithmic optimizations to eliminate unneeded computation. DECA automatically parallelizes XHMM on both multi-core shared memory computers and large shared-nothing Spark clusters. We performed CNV discovery from the read-depth matrix for the s=2535 1000 Genomes Phase 3 exomes in 31 minutes on a 16-core workstation (10.6x speedup vs. XHMM), 23.2 minutes using 40 executor cores on a Spark cluster (10.3x speedup vs. XHMM), and 18.45 minutes using 48 executor cores on Databricks on the Amazon cloud (approximate cost of \$0.53). We performed CNV discovery directly from the BAM files in 292 minutes using 640 executor cores on a 56-node Spark cluster. Of 16512 XHMM CNV calls in a subset of 500 samples, 99.92% have an overlapping DECA call with same copy number and 93% have identical breakpoints. The breakpoints of most partially overlapping calls differed by 1 target (65.7%) with a mean difference of 3.3 targets. We do not expect identical results due to differences in the numerical behavior of multiplying small probabilities. We describe lessons learned, including the importance of library choice - the XHMM PCA step can be sped up 3.9x by switching to an optimized LAPACK library - and algorithmic optimizations. DECA implements an iterative algorithm for PCA normalization that produces the same results as XHMM, but typically reduces the amount of memory required for any one computer by 10x thus increasing the size of the cohort than can be analyzed on commodity hardware. **Conclusions:**ADAM and Apache Spark are a performant and productive platform for implementing large-scale genome analyses that are scalable by default, but efficiently utilizing large clusters can require algorithmic optimizations and careful setting of Spark's many configuration parameters.

1619F**metro and hudson: R packages to leverage the utility of Manhattan plots.**

A. Lucas¹, A. Verma¹, M.A. Hall², M.D. Ritchie¹. 1) Department of Genetics, Institute for Biomedical Sciences, University of Pennsylvania, Philadelphia, PA; 2) College of Agricultural Sciences, The Pennsylvania State University, University Park, PA.

As genome-, phenome-, and exposome-wide analysis studies (GWAS, PheWAS, EWAS) increase in size and complexity, there is need to further develop visualization techniques for the presentation of results. We have developed a pair of ggplot2-based R packages, metro and hudson, which augment basic Manhattan plots with the aim of improving result communication and interpretation. The metro package allows users to quickly create static, interactive, or animated Manhattan plots for GWAS, PheWAS, and EWAS results. Though the package was designed to cater to SNP (GWAS/PheWAS) and exposure (EWAS) data, it will accept any data that can be mapped to chromosome and position, such as genes, while EWAS plots are suitable for analyses with many results organized into categories, such as tissues or pathways. Further, while Manhattan plots traditionally visualize p-values, our flexible framework allows the user to plot other data, such as beta or intensity values instead. Within metro, adding threshold lines, highlighting points by either name or numeric value, and assigning colors or shapes to groups are common to all plotting functions. Static images allow for additional text annotation while the animated plotting functions, which output GIF files, enable the user to map an additional layer of information to an animation frame. For example, a user can see how a GWAS landscape changes based on population or the model assumed when encoding SNPs. This is potentially more impactful than comparing two static images as it highlights nuance and could reduce change blindness. Interactive plots, which can be shared as an HTML file, making them particularly useful for collaboration and translation of results, allow the user to include meta-information viewed by hovering over the data points as well as link the points to websites or databases, such as dbSNP or the NHGRI-EBI GWAS Catalog, upon mouse click. The counterpart to metro, hudson, can be used to create mirrored Manhattan plots for the position by position comparison of results between two datasets, e.g. discovery and replication, or two features of a dataset, e.g. subsets of SNPs by direction of effect. The Hudson plots are static images characterized by a shared x-axis and diverging y-axis, and contain the same annotation features as described in the metro package. Together these packages enhance traditional Manhattan plots providing for comprehensive visualization of increasingly larger scaled and more detailed analyses.

1620W**A bioinformatics pipeline for next-generation sequencing of the killer cell immunoglobulin-like receptor gene complex.** W.M. Marin¹, P.J. Norman², R. Dandekar¹, J.A. Hollenbach¹. 1) Neurology, UC San Francisco, San Francisco, CA; 2) SOM-MED, UC Denver, Denver, CO.

The killer cell immunoglobulin-like receptor (KIR) complex (Chr19) encodes receptors expressed on the surface of natural killer (NK) cells, which may have either activating or inhibitory effects on NK cell activity, resulting in a finely tuned reactivity. The KIR region is marked by highly homologous loci, extensive structural variation, and high levels of polymorphism, all of which serve to significantly confound genotype determination. Variation in KIR gene-content has been implicated in numerous immune-mediated and infectious diseases, but more detailed interrogation of these loci has been thwarted by the complexity of the region. To facilitate study of KIR variation, we have developed a novel bioinformatics pipeline that allows for high-throughput, accurate determination of high-resolution KIR genotypes from next-generation sequencing data. A gene content determination method identifies sequence motifs that are unique to individual KIR loci, then searches over the sequencing data of a sample for these unique motifs to determine which KIR loci are present. Many KIR loci are copy number variable, and this method allows for the determination of locus presence and copy number, which is vital for accurate read mapping and genotype calling. A read filtration method resolves highly-homologous sequence through a series of read mapping steps that separate reads from homologous loci. The first step is a permissive positive filter, to isolate reads that map to specific KIR loci, which is followed by a stringent negative filter, to discard reads that map strongly to other KIR loci, and the remaining reads are used for final mapping. The negative alignment step reduces genotypic ambiguity through multiple rounds of read mapping to highly divergent haplotype references, which are constructed to most completely represent the variation found within the KIR loci. Combining the results of multiple rounds of mapping allows the identification of variants that might be missed by a single mapping, but more importantly, it allows the exclusion of alleles that contain variants that were not found in any round of mapping. Together, these methods solve the issues preventing the accurate determination of KIR genotypes. The pipeline implementing these methods is open source and has been validated and developed for large-scale analysis of targeted, as well as whole-genome and -exome sequencing data, enabling population level studies investigating the role of KIR genes in disease.

1621T

Accurate fetal DNA variant calling in the presence of maternal cell contamination. *E. Nabieva, Y. Kapushev, A. Chernomorov, S. Mishra Sharma, D. Vinogradov, G. Bazykin, D. Yarotsky.* Skolkovo Institute of Science and Technology, Moscow, Russian Federation.

High-throughput sequencing of fetal DNA is a promising and increasingly common method for the discovery of all (or all coding) genetic variants in the fetus, either as part of prenatal screening or diagnosis, or for genetic diagnosis of spontaneous abortions. In many cases and depending on the tissue and the laboratory technique used, the fetal DNA (from chorionic villi, amniotic fluid, or abortive tissue) can be contaminated with maternal cells, resulting in the mixture of fetal and maternal DNA. This maternal cell contamination (MCC) undermines the assumption, made by traditional variant callers, that each allele in a heterozygous site is covered, on average, by 50% of the reads, and therefore can lead to erroneous genotype calls. We present a panel of methods for reducing the genotyping error in the presence of MCC. All three methods start with the output of GATK HaplotypeCaller on the (contaminated) fetal sample and sequencing data for either one or both of the parents, and additionally rely on information about the MCC fraction (which itself can be readily estimated from the HTS data). The first of these methods uses an explicit formula based on simple probabilistic assumptions to “recalibrate” the fetal genotype calls produced by MCC-unaware HaplotypeCaller. The second method employs machine learning to recalibrate the HaplotypeCaller genotype calls in the presence of MCC; it trains on variant calls for fetal samples with and without simulated MCC, the latter being used as true labels. The third method adapts the model of GATK PhaseByTransmission to calculate the genotype posteriors in the presence of MCC using fetal and parental read data. We use simulated contaminated fetal data to show that all three methods lead to substantially improved accuracy when compared with the original MCC-unaware HaplotypeCaller calls, as judged by the concordance with genotype calls made on uncontaminated child data. We discuss the drawbacks and advantages of each of the presented methods.

1622F

Confounding variable analysis of transcriptomics sequencing samples. *M. Nagarajan, S. Ramachandrala, M.W. Saifi, R. Bettadapura, D. Chohan.* Strand Life Sciences Private Limited, 5th Floor, Kirlskar Business Park, Bellary Road, Hebbal, Bangalore 560024 INDIA.

The purpose of RNA analysis is often to reveal sources of differential expression between various biological groups. It relies on the dogma that phenotype could be inferred from the patterns of gene expression. However, sequencing analysis, starting from the lab and moving on up to bioinformatics and interpretation, despite years of standardization, is notoriously difficult to be made error free. While random errors are unpreventable and acceptable, systematic errors are to be avoided. Some systematic errors are known and preventable, while others creep in, especially, in large data sets. Biologists often use simple batch-correction methods but are largely unaware of more mathematical approaches to find underlying systematic errors and correct for them. To address this problem of known and unknown confounders that affect expression matrices, we have developed an intuitive framework for Confounding Variable Analysis (CVA) in StrandNGS, our comprehensive NGS data analysis and management software. Any gene list could be used as an input and on providing biological variables and known (if any) technical variables, confounders could be identified and corrected. The spread of the data before and after the correction can be visualized. The sample behavior is captured using Principal Component Analysis (PCA) plots while the gene expression distributions are shown as box-whisker plots. An f-test is applied to find differentially expressed genes and genes whose expression values are affected/not-affected by the identified confounders. This corrected expression matrix is available for all downstream analyses. We intend to illustrate the utility of this feature in StrandNGS using 3 data sets: GSE87194 - Set of post-mortem brain samples from Schizophrenic patients and age-matched controls. The paper reported batch-effects and the batch information was not published. We could find the batches and correct for them using CVA. GSE77288 - Set of pan-cancer blood platelet samples and controls. While most tumor samples were from older patients, all control samples were from healthy young individuals. Using CVA, the confounding effect of age could be corrected. GSE22521 - Set of primate prefrontal cortex samples collected from multiple species after post-mortem assayed using microarrays and expression values reported at gene level. Given the varied collection and processing times, this dataset showed a lot of underlying confounders that could be identified and corrected using CVA.

1623W**Prediction of pathogenic non-coding variants associated to monogenic Mendelian diseases through an integrative supervised learning approach mining signals of recent and ongoing purifying selection in human.**

A. Rausell^{1,2}, B. Caron², Y. Luo². 1) INSERM UMR 1163, Institut Imagine, 24 boulevard du Montparnasse, 75015 Paris, France; 2) Paris Descartes – Sorbonne Paris Cité University, Institut Imagine, 24 boulevard du Montparnasse, 75015 Paris, France.

The study of rare Mendelian diseases through exome sequencing typically yields incomplete diagnostic rates (~8-70%), depending on the disease type. Whole genome sequencing of the unresolved cases allows addressing the hypothesis that causal variants could lay in non-coding regions with damaging regulatory consequences. Computational methods are then needed to score the pathogenicity potential of the large amount of rare and singleton variants found in each individual genome. However, state-of-the-art methods to prioritize non-coding variants have been characterized on sets largely composed of trait-associated polymorphisms and associated to common diseases. In this work we first curated a collection of $n=737$ high-confidence pathogenic non-coding single-nucleotide genetic variants in proximal cis-regulatory regions associated to monogenic Mendelian diseases. We then systematically evaluated the ability to predict causal variants of an exhaustive set of genomic features extracted at three levels: the affected position, the flanking region and the affected gene. In addition to epigenetic features and inter-species conservation scores, a complete set of ongoing purifying selection signals in human was explored. This represents a main novelty allowing to exploit sequence constraints potentially associated to recently acquired regulatory elements in human. A supervised learning approach using gradient tree boosting on the previously described sets of features reached a predictive performance characterized by an area under the ROC curve = 0.84 and an area under the Precision-Recall curve = 0.53. These figures represent, respectively, a relative improvement of >12% and >51% upon the performance of current state-of-the-art methods for prioritizing non-coding disease variants. The approach consistently outperformed reference methods under multiple configurations of the sets of variants used for learning and independent testing. A detailed comparative benchmark is presented and results discussed in terms of the type of the targeted regulatory region.

1624T**Scripting support in StrandNGS.** M.W. Saifi, S. Ramachandrala, S.C. Cherukuri. Strand Life Sciences Pvt. Ltd., Bangalore, Karnataka, India.

Analyzing Next Generation Sequencing (NGS) data involves multiple teams from varied backgrounds. Also, for most comprehensive analysis, generally multiple approaches are used to compare and contrast the results to derive valid conclusions. However, given that bioinformaticians and biologists use different approaches, the collation of their respective analyses becomes cumbersome. StrandNGS is an NGS data analysis software designed for the biologists. It provides this flexibility in the form of scripting support within the framework of the tool. This ensures that the same inputs/data could be used to perform various analyses using different techniques in the software as well as store and compare results on the same. StrandNGS provides support for Jython & R scripting. We intend to showcase this functionality using example study workflows on how one could use scripting to: 1. Integrate HISAT, an open-source software that helps in transcriptome alignment. This flexibility ensures that researchers can continue to use their lab-specific practices while taking advantage of the data management and visualization capabilities of StrandNGS. Also, the aligned files could then be used for all downstream analyses available in StrandNGS. 2. Use pre-packaged R-scripts for multiple differential expression analyses (using external R packages) in the same experiment. This eliminates the need for preparing the data set multiple times to suit the input needs as required by various packages. 3. Write custom scripts to perform vital steps in the workflow. Given that sequencing data is primarily used for answering a myriad of biologically relevant research questions, no software could singularly provide all options readily. Therefore, this option helps researchers to dynamically customize the software for their specific use. Moreover, one can also integrate these scripts in a custom pipeline to either perform additional analysis or replace existing techniques with the pertinent latest external techniques to obtain faster and accurate results or interface with third-party tools. We aim to showcase the technical strengths and robustness of StrandNGS which enables effortless and smooth integration of customized scripts in to the tool. This all-encompassing ability of StrandNGS ensures that multidisciplinary research teams can make the best use of their space, time, and resources to answer diverse biological research problems through a collaborative utilization of this integrated platform.

1625F

Comparison of African genome resource and 1000-Genomes imputation. *M. Schmidt¹, B. Kunkle¹, K.L. Hamilton-Nelson¹, M. Jean-Francois¹, F. Rajabli¹, A.J. Griswold¹, G. Beecham¹, G.D. Schellenberg^{2,4}, C. Reitz², M.A. Pericak-Vance¹, E.R. Martin^{1,3}.* 1) University of Miami, Leonard M. Miller School of Medicine Dr. John T. Macdonald Foundation Department of Human Genetics Miami, FL; 2) Columbia University; 3) Alzheimer's Disease Genetics Consortium (ADGC); 4) University of Pennsylvania, Perelman School of Medicine.

The African Genome Resource (AGR, S. McCarthy et al. 2016) recently released a data from a large sample of African individuals to help improve imputation in individuals with African ancestry. The goal of our investigation is to determine if AGR works better for imputing African genomes than the widely used 1000 Genome Project (1KGP, The 1000 Genome Project Consortium, 2015). In addition we also investigate the impact of SNV confidence thresholds, "gen" probability cutoffs and minor allele frequency (MAF) on the accuracy of the imputed data. We selected a group of individuals with African heritage from the Alzheimer's Disease Genetics Consortium (ADGC) that were genotyped with genotype-wide arrays (GWAS), whole genome sequencing (WGS) and whole exome sequencing (WES). Using genotype array data as a backbone, genotypes were imputed twice, once using the 1KGP reference panel and the AGR. In each imputed set, we separated positions into two disjoint groups, GWAS-only and purely imputed positions. Then we intersected the positions of all three sets, AGR, 1KGP with either WGS or WES to determine the positions common to all. For both groups, imputed and GWAS, we conducted a pairwise comparisons between all three sets to determine concordance. Our results show the concordance of the GWAS data and of the purely imputed data versus the sequencing data under various thresholds and MAFs. Initial results on chromosome 22 show that AGR yields better quality results for subjects with African ancestry.

1626W

Quantifying the unknown unknowns in molecular genetic assays. *D. Siegel¹, O. Le Tonqueze¹, A. Biton^{1,2}, D. Erle¹, N. Zaitlen¹.* 1) University of California, San Francisco, San Francisco, CA; 2) Pasteur Institute, Paris, France.

Transformative advances in molecular technologies, such as massively parallel reporter assays (MPRAs) and CRISPR screens, can efficiently characterize the effects of genetic and genomic variation on cellular phenotypes such as gene expression, growth, and chemical resistance. These technologies complement genetic association studies by providing direct identification of causal factors, an important step towards mechanistic understanding. To date, analyses have proceeded in a bottom-up fashion, identifying individual genomic regions or genetic variants that perturb phenotypes of interest. In this work, we seek to determine the overall contribution of sequence variation to phenotypic variation in massively parallel experiments and to demonstrate how this can lead to an improved understanding of mechanistic biology. To achieve these goals we developed a method to determine the fraction of variance in molecular assays explained by the included sequences, as opposed to other sources of variability such as technical factors and statistical noise. We can apply this method to the entire assay, akin to estimating the "heritability" of the experiment; or to subgroups of sequences, akin to enrichment analyses for regions of interest. We have applied this approach to several simulated and real data sets. In our MPRA for 3'UTR function (Fast-UTR), we show that 63% of the variance in phenotype (reporter RNA level) is explained by sequence variability; while at most 10% can be explained by any single known 3'UTR category. That is, the function of the 3'UTR is mostly unknown. Then we show that the flanking sequence of a known 3'UTR functional element (an AU-rich element (ARE)) accounts for 16% to 56% of the variance of the effect of a killer mutation, depending on the length of the ARE. This illustrates the critical importance of genetic context to the effect size of a mutation. Finally we will discuss the applicability of this method to related assays like CRISPR screens. Our method is implemented as a free and open source software package and can be used in any barcoded molecular assay, as it is based on holding out or resampling over barcoded clones within a dataset, and it does not require parametric model assumptions. Going forward, this method will set an upper limit to the expected performance of prediction methods, and provide a better estimate of the relative importance of various classes of motif- and sequence-based factors in massively parallel experiments.

1627T

A fast and accurate open-source variant caller. *M.A. Simpson¹, S. Thiem¹, A. Dorr¹, E. Sanders¹, A. Rimmer¹, J. Penn², X. Bai², J.G. Ried², S. Teasdale¹, A. Heger¹, G. Lunter¹.* 1) Genomics plc, Oxford, United Kingdom; 2) Regeneron Genetics Center, Tarrytown, NY, USA.

The use of short-read sequencing for whole exome and whole genome analysis is now routine for genetic research and clinical applications. The accurate identification and genotyping of positions at which an individual's genome varies from the reference genome is a critical step in drawing inference from such data. In addition to accuracy, as sample sizes grow to hundreds of thousands of individuals there is a requirement for computationally efficient approaches to variant calling. Whilst many solutions for variant calling have been developed in academia and industry, they are often either limited by accuracy or by computational performance, inhibiting deployment across many thousands of genomes. *weCall* is an open-source variant caller developed by Genomics plc. *weCall* capitalises on highly performant algorithm to enable variant detection and genotyping of aligned short read sequencing data with low CPU, memory and I/O usage while achieving high accuracy. The software, implemented in C++11, undertakes local re-alignment of sequencing reads, identifies and handles duplicate reads, calls and determines the locally phased genotypes of variant positions, and performs quality filtering, all in a single pass of an input BAM file. The performance of *weCall* has been evaluated internally by Genomics plc and externally as part of a beta testing program. It consistently outperforms other variant calling approaches in terms of sensitivity, specificity of detection and genotyping SNPs, MNPs and notably short (<80 bp) indels. In addition, *weCall* is able to call larger indels up to ~100 kb by analysing read pairs, which further helps to improve the specificity of SNP calling. *weCall* will be the primary variant calling tool deployed in the Regeneron Genetics Center-led consortium to exome sequence the UK Biobank participants.

1628F

Sharing aggregate genomic data on the web: re-engineering and extending the gnomAD browser. *M. Solomonson^{1,2}, N. Watts^{1,2}, T. Singh^{1,2}, H. Huang^{1,2}, F.K. Satterstrom^{1,2}, B. Weisburd^{1,2}, K.J. Karczewski^{1,2}, M.J. Daly^{1,2}, D.G. MacArthur^{1,2}.* 1) Analytic and Translational Genetics Unit (ATGU), Massachusetts General Hospital, Boston, MA, USA; 2) Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA.

Increasing the accessibility of aggregate genomic data can have an immediate impact on human health. For example, online access to variant frequency resources such as ExAC and gnomAD has proven critical for gene discovery and diagnosis in rare disease. As genomic resources continue to grow in scale and complexity, it is becoming increasingly challenging to engineer and maintain the technical infrastructure necessary to share them with the broader community. A centralized, community-driven, open-source suite of developer tools is needed for the construction of portals that share and display aggregate data of diverse types. Such tools should include a library of robust, modular web components geared for data visualization which serve as elemental building blocks in the creation of complex web applications. This toolkit must support extremely large genomic datasets stored in the cloud and should integrate with scalable genomic analysis software. Here, we describe a general framework for sharing aggregate genomic data on the web. We showcase the framework by demonstrating a suite of portals, beginning with an updated version of the gnomAD browser. We have re-engineered the gnomAD browser, future-proofing it for an eventual dataset of more than 1 million exomes and genomes. We also present the SCHEMA consortium exome results browser, which displays summary statistics from a case-control sequencing study of over 25,000 schizophrenia exomes and 50,000 controls. The exome results browser is designed to be extendable to other large-scale genotyping and sequencing studies of other disorders. Finally, we present a prototype for dbLoF, a catalogue of loss-of-function variants and associated phenotypic data from large human populations, for leveraging gene-inactivating mutations to explore the function of uncharacterized human genes.

1629W

Systematic discovery of conservation states for single-nucleotide annotation of the human genome and variant interpretation. A. Sperlea, J. Ernst. University of California, Los Angeles, Los Angeles, CA.

Effectively annotating the non-coding human genome is a key challenge to being able to interpret and prioritize potential disease associated genetic variants. Comparative genomics based annotations have been an important source of information for whole genome annotation, but have focused primarily on scores or binary elements of evolutionary constraint. In this work we present a novel computational method, ConsHMM, for systematically annotating entire genomes into one of a number of different 'conservation states'. ConsHMM learns 'conservation states' based on the combinatorial and spatial patterns of which species align to and match a reference genome in a multiple-species DNA-sequence alignment using a multivariate hidden Markov model. We applied this approach to a number of different whole genome multiple species alignments, including a 100-way vertebrate alignment annotate the human genome at single nucleotide resolution into one of 100 conservation states. ConsHMM states show diverse enrichments for orthogonal genomic annotations. Both locations called as constrained and those called as non-constrained by existing constraint measures are heterogeneous in their assigned conservation state. Conservation states also had greater or complementary information than constraint scores or elements for predicting external annotations. Specific conservation states exhibited enrichments for common human variants, while a different set of states exhibited enrichments for variants in the Genome-Wide Association Studies catalog relative to common variation. We demonstrate that bases within constrained elements can have substantially different enrichments for phenotype-associated heritability depending on which set of conservation states they overlap. We also show how the conservation state annotations help clarify the relationship between inter-species constraint and the recently introduced context-dependent tolerance score for identifying intra-species constraint, which were previously reported to have limited overlap. We identify conservation states that strongly enrich for bases prioritized by both approaches as well as uniquely by each approach. In addition to annotating the reference genome, we developed an approach to accurately estimate the conservation state assignment for every possible nucleotide at every base in the genome. These variant specific conservation states give additional information to identify and interpret potential disease-associated variants.

1630T

Quality assessment and comparison of exome data from neonatal dried blood spot DNA with whole blood data. U. Sunderam¹, K. Kundu¹, A.N. Adhikari², Y. Wang², D. Vaka³, R.J. Currier³, J.M. Puck³, P-Y. Kwok³, S.E. Brenner², R. Srinivasan¹. 1) Tata Consultancy Services, Hyderabad, TS, India; 2) University of California, Berkeley, CA, USA; 3) University of California, San Francisco, CA, USA.

Several countries and American states routinely collect and store neonatal dried blood spots (DBS) through their newborn screening (NBS) programs. These archived collections of DBS provide a potential DNA and analyte source for a large cohort of healthy and affected individuals for scientific inquiry. In order to use DBS samples for DNA analysis, including variant discovery, the quality of the DNA sequence obtained from these should be comparable to those from whole blood. In our study, we compared the DNA sequence quality of exomes of a large cohort of 180 DBS archived between 2006 and 2011 with that of 35 whole blood samples sequenced between 2013 and 2014 at the same facility. The de-identified DBS were obtained from the California Department of Public Health under an approved IRB protocol. In the absence of matched samples, we performed an extensive analysis to assess the quality of DNA extracted from DBS. We compared 16 metrics that can be categorized broadly into 3 categories - evaluation of how effectively the reads mapped to the reference genome, the degree of DNA damage, and assessment of the variants that were ultimately called for the samples. Additionally, we evaluated these parameters for a smaller set of DBS samples that were archived between 1980 and 1982, and had been stored for over 30 years. Our assessment showed that DNA sequenced from DBS mapped with similar efficiency as reads from whole blood, while considering parameters such as number of unmapped reads, reads mapped to different chromosomes, uniformity of coverage, and gene level coverage. The overall DNA damage rates, measured by the mismatch rate with the reference genome, were also comparable across the 3 sets of samples. While examining single nucleotide changes that may have arisen from DNA damage related nucleotide mis-incorporation, the A>T and T>A changes were higher in the DBS samples. However, this does not appear to have impacted the accuracy of variant calling. DBS samples yielded a comparable number of high quality SNVs and comparable numbers of indels were observed in samples sequenced with at least 50x coverage. The Transition-Transversion ratios were found to be comparable and in the expected range indicating adequate quality of the variants. Overall, we conclude that dried blood spots, stored desiccated and at -20 degrees C, provided exomes of sufficient quality for clinical interpretation.

1631F

Construction and benchmarking of a multi-ethnic reference panel for the imputation of the HLA. *M. Wendorff, F. Degenhardt, M. Wittig, A. Franke.* Institution of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Schleswig-Holstein, Germany.

Genotype imputation of the human leukocyte antigen region (HLA) is a cost-effective means to infer classical HLA alleles from inexpensive and dense SNP array data. In the research setting, imputation helps to avoid costs for wet lab-based HLA typing and thus renders association analyses of the HLA in large cohorts feasible. Yet, most HLA imputation reference panels are targeting Caucasian ethnicities and multi-ethnic panels are scarce. Here, we present a high-quality multi-ethnic reference dataset based on genotypes measured with Illumina's Immunochip genotyping array including samples from Europe, China, India, Iran, Japan, Korea, Malta and an admixed African American dataset. We HLA-typed more than 1,300 samples from the afore-mentioned eight different ethnic groups using a high-resolution next generation sequencing approach. Manually curated 4-digit level alleles for the classical HLA class I and II loci were included in the reference panel. Using extensive cross validation and different combinations of reference datasets, we benchmarked the imputation using the HLA imputation tool HIBAG across our multi-ethnic reference and an independent, previously published, dataset for the subpopulations of the 1000 Genomes project. We combined HLA data from 1000 Genomes with our multi-ethnic reference to create references with a full 4-digit HLA context for our own multi-ethnic samples alone and at G group level for the combined dataset for all classical HLA class I and II genes as well as the *HLA-DRB3/4/5*. We achieve high average imputation accuracies across *HLA-A*, *-B*, *-C*, *-DQA1*, *-DQB1*, *-DPA1*, *-DPB1*, *-DRB1* and the *-DRB3/4/5* within a 5x cross-validation framework with our own multi-ethnic data and the 1000 Genomes data. We also present an overview of the current literature on *HLA-DRB3/4/5* and show that the imputation of these loci should be considered with care. In conclusion, our new multi-ethnic reference dataset allows for high resolution HLA imputation for all of the classical HLA class I and class II genes as well as the *HLA-DRB3/4/5* for different ethnicities.

1632W

Introducing Aneris: A hadoop cluster for bioinformatics at scale. *N.R.*

Wheeler¹, M. Warfe², W. Bush¹. 1) Institute for Computational Biology, Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH; 2) Institute for Computational Biology, Division of University Technology, Case Western Reserve University, Cleveland, OH.

Human genetics datasets are expanding in size, variety, and number at an accelerating rate, and the integration of multiple data types is becoming a more critical component of genetics research. For example, the Alzheimer's Disease Sequencing Project is generating whole-genome sequencing data on thousands of individuals across multiple studies, many of which have orthogonal high-dimensional data of their own like gene expression and neuroimaging values. While the accumulation of biological data of this magnitude presents lucrative opportunities for data analysis, it also creates new burdens for data storage, manipulation, and processing. Just as novel analytical approaches help to address the new statistical issues inherent to these datasets, new data processing and analytical systems are needed to handle this incoming deluge of data, all while maintaining the reproducibility of scientific studies at scale. Here we demonstrate a scalable analytics platform we call ANERIS based on the Hadoop software ecosystem, which addresses these challenges through the application of feature rich and widely used open source technologies. By leveraging distributed storage and processing across a cluster of powerful computational nodes, bioinformatics questions can be asked interactively, providing useful insight from large-scale data at unprecedented speeds. ANERIS provides researchers the ability to work collaboratively across numerous and massive biological datasets in ways that would not be possible with conventional computational methods. Building upon the open-source HAIL software, currently supported applications include the storage and retrieval of large datasets, effortless GWAS style statistical modeling, and fast SQL table operations for variant annotation processing. Extended operations with Spark enable annotation-based analyses to run orders of magnitude faster, with sequences of operations that take several days using traditional SQL processing being completed in only minutes on ANERIS. Presently, over 90TB of data has been uploaded and formatted for easy access by researchers for downstream processing and use in reproducible genomic studies. Taken together, the shift from serial analysis operations to distributed parallel processing provides more stable, rapid, and reproducible workflows for genomic and bioinformatics data processing.

1633T

A computational approach to detect site specific change in read patterns from high resolution sequencing assays. *N. Yamada, S. Mahony.* Department of Biochemistry & Molecular Biology, Pennsylvania State University, University Park, PA.

Recent developments in high throughput sequencing assays enable us to profile biological activities genome-wide. Genome-wide sequencing assays such as ChIP-exo and ATAC-seq utilize nuclease digestion with or without chromatin immunoprecipitation. Dense mapping of nuclease cleavage sites identifies small regions protected from digestion by the presence of a bound transcription factor. We seek to take advantage of DNA footprint shapes and perform analysis beyond finding peak locations and signal strength in high resolution sequencing assay technology data. Specifically, we develop a computational method to detect site specific change in DNA footprinting shape and read enrichment levels in different biological conditions such as development stages and disease and healthy conditions. Such analysis will define a set of factors that are responsible for changing biological activities and provide us per site quantification of the molecule activity differences. Our computational analysis method may provide novel insights into precise regulatory elements interactions exist in the genome.

1634F

KMgene: A unified R package for gene-based association analysis for complex traits. *Q. Yan¹, Z. Fang², W. Chen^{1,2}.* 1) Division of Pulmonary Medicine, Department of Pediatrics, Children's Hospital of Pittsburgh of UPMC, University of Pittsburgh, Pittsburgh, PA; 2) Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA.

The gene-based tests are becoming an attractive complement to the single variant tests used in GWAS. Compared to single variant tests, gene-based tests are able to identify weak individual signals by combining the effects of variants in the same gene, and greatly reduce the number of multiple testing. One widely used gene-based test is the sequence kernel association test (SKAT) based on a Kernel Machine (KM) regression framework. Although SKAT was first developed for testing rare variants, it could be easily applied to common variants. After SKAT was introduced for testing independent samples with continuous and binary traits, a number of methods and corresponding tools have been developed to extend the approach to complex traits, such as familial, multivariate and longitudinal traits. These methods are based on a GLMM framework. Since SKAT has imbalanced power performance when the single variant effects are in the same direction (i.e., all rare alleles are risk or protective) or in different directions, the optimal version, SKAT-O, was developed to balance these two scenarios. However, most of the extended SKAT methods do not consider the optimal tests balancing genetic effects. In addition, those programs have different input and output formats, making it difficult to use in practice for bioinformaticians with limited genetics knowledge. Therefore, in this report, we introduce KMgene, a one-stop solution that combines SKAT-type methods for complex traits and extends them to include their corresponding optimal tests. KMgene can perform association tests between a set of genetic variants and familial, multivariate, longitudinal or survival traits.

1635W

Homozygosity mapping provides evidence of pathogenicity in recessive Mendelian disease. *M.N. Wakeling¹, T.W. Laver¹, C.F. Wright¹, E. De Franco¹, K.L. Stals², A. Patch³, A.T. Hattersley¹, S.E. Flanagan¹, S. Ellard^{1,2}, DDD Study.* 1) Institute of Biomedical and Clinical Science, University of Exeter, Exeter, England, United Kingdom; 2) Department of Molecular Genetics, Royal Devon & Exeter NHS Foundation Trust, Exeter, UK; 3) QIMR Berghofer, Herston, Queensland, Australia.

One of the greatest challenges currently facing those studying Mendelian disease is identifying the pathogenic variant from the long list produced by a next generation sequencing test. The ACMG guidelines provide a framework for variant classification using all available evidence. Homozygosity mapping has been previously used to guide the identification of pathogenic variants in recessive disease in consanguineous populations. Using 179 homozygous pathogenic variants from three independent cohorts, we investigated the predictive ability of homozygosity mapping for identifying the regions likely to contain the causative variant. We demonstrate that pathogenic variants are not randomly distributed among the regions of homozygosity, but are disproportionately likely to be found within the largest regions of homozygosity. The size and rank of the region of homozygosity can thus be used as supporting evidence of pathogenicity. The maximal predictive power is achieved in patients with <8% homozygosity and variants >3Mb from a telomere, this gives an AUC of 0.735 and results in 92% of the causative homozygous variants being in one of the top 10 largest regions of homozygosity in a sample. Only 62% of benign rare homozygous variants fulfil the same criterion. This predictive power is comparable to SIFT or PolyPhen. The number and size of homozygous regions within an individual's genome is influenced both by ancestral population effects and by recent consanguineous events. It is important to differentiate the two cases as disease causing variants are likely to be in regions of recent homozygosity – variants in ancestral regions of homozygosity have been exposed to selection in a homozygous state for sufficient time for selection to act on them. Ancestral regions of homozygosity are likely to be smaller than homozygous regions which are the result of recent consanguinity. This can explain why variants which cause recessive Mendelian disease are contained in the largest regions of homozygosity. This method can be used to prioritize the list of candidate variants in gene discovery studies. When classifying a homozygous variant using the ACMG guidelines, the size and rank of the region of homozygosity in which the candidate variant is located can be considered as supporting evidence for pathogenicity.

1636T

Limit of detection study of targeted sequencing for germline copy number variation. *Q. Cai, T. Kuang, X. Meng.* WuXi NextCODE, Shanghai, Shanghai, China.

The whole exome or panel based NGS method is promising to detect the copy number variation (CNV) of exons due to its low cost comparing to whole genome sequencing. However, methodological difficulties still exist considering biases during library construction and sequencing process. Barely a study worked carefully to assess the factors that influence the detection limit of the method. When it comes to clinical diagnostic applications, the detection limit of the CNV types across the whole panel and its validation is demanding for genetic counselors and doctors to make accurate diagnosis. We developed a CNV detection software iConer: it integrated multiple depth-of-coverage methods to achieve a high sensitivity; it also incorporated B-allele frequency information to filter out false positives. With further clinical phenotype information, 100% precision could be achieved. Five main factors were included to investigate the CNV calling accuracy: CNV length, ploidy, sequencing depth, batch effect, and pseudogene. In total, 10 samples with spike-in CNVs, 25 CNV positive cell line samples, and 67 clinical blood/saliva samples were used for validation in this study. According to the results: 1) CNV length: for the 10 Mb panel with average read depth 150x, sensitivity increased with the CNV length; > 90% sensitivity could be achieved for CNV length > 5 kb, and ~100% sensitivity could be achieved for CNV length > 100 kb; 2) Ploidy: deletion was easier to be called than duplications; homologous/heterozygous deletion was detected as small as 500 bp/5 kb with ~100% sensitivity, respectively; the detection limit for duplications was about 100 kb with ~100% sensitivity. 3) Sequencing depth: at least 100x average coverage was required for effective CNV calling at single exon level; 4) Batch effect: the results varied a lot due to different evenness among batches; thus, quality controls were introduced before the pipeline; 5) Pseudogene: The mapping problem of pseudogenes could be mitigated by using more specific genomic markers on or near the target regions. In conclusion, our study provided an accurate CNV calling pipeline for targeted sequencing data and valuable assessment of the factors that influence the limit of detection of the method.

1637F

Exome CNV calling improves diagnostic yield in rare disease families.

K.R. Chao^{1,2,4}, J.K. Goodrich^{1,2,4}, A.S. Lee^{6,7}, M. Babadi⁸, S.K. Lee⁸, A. Smirnov⁸, M. Lek^{1,3}, H.L. Rehm^{1,2,4}, D. MacArthur^{1,2,4}, A.H. O'Donnell-Luria^{1,2,4,5}. 1) Broad Institute Center for Mendelian Genomics, Broad Institute of MIT and Harvard, Cambridge, MA; 2) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA; 3) 16 Department of Genetics, Yale School of Medicine, New Haven, Connecticut, USA; 4) Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA; 5) Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA, USA; 6) Department of Neurology, Children's Hospital Boston, Boston MA; 7) Department of Neurology, Harvard Medical School, Boston, MA; 8) Methods Group, Data Sciences Platform, Broad Institute of MIT and Harvard, Cambridge, USA.

The Broad Institute Center for Mendelian Genomics (CMG) has sequenced over 2,600 families with rare genetic diseases since 2016. A diagnosis has been identified for 27% of these families through exome analysis of rare single nucleotide variants (SNVs) and small indels. Additional copy number variant (CNV) analysis may contribute to diagnosis of families that remain unsolved after initial exome analysis. Traditionally, CNVs that are too large to be identified through standard exome analysis have been detected using array comparative genomic hybridization (CGH) or whole genome sequencing (WGS). However, the combination of small indel detection during standard exome analysis and large CNV detection via array testing fails to capture a class of mid-sized CNVs. A number of tools (xHMM, CoNIFER, CNVnator, etc.) have been developed in the past few years to detect medium sized CNVs in whole exome sequencing (WES) data. GATK's new CNV tool, the Germline CNV (gCNV) caller, incorporates and improves upon many features of these existing tools and uses a probabilistic framework to infer both rare and common CNVs from read depth data in the presence of systematic bias. The Broad CMG ran gCNV on over 5,000 samples from 20 different rare disease cohorts. After applying modest CNV quality score filtration, gCNV detected an average of 25 CNVs per exome. gCNV diagnosed 22 previously unsolved cases and found an additional 23 candidates across the CMG cohorts by detecting two duplications and 45 deletions, including several that involved only a single exon. The CNVs ranged from 122 base pairs to 356,701 base pairs in size and involved a median of two exons. Due to their relatively small size, the majority of the pathogenic CNVs called by gCNV would not have been found by standard clinical CGH arrays. In addition, WGS validated several CNVs, and defined breakpoints, but was not required to detect any of the CNVs in these samples. Our results with gCNV demonstrate that exome CNV callers can increase diagnostic yield for existing WES data for individuals that remain undiagnosed after standard analysis approaches, providing a more comprehensive approach, with higher sensitivity, than ordering a separate array, and a more cost-effective option than moving directly to whole genome sequencing.

1638W

A public benchmarking resource of 194,731 CNV calls generated from a combination of short reads, linked long reads and true long read technologies.

X. Chen¹, S. Chen¹, A. Gross¹, F.J. Sedlazeck², M. Bekritsky², D.R. Bentley¹, M.A. Eberle¹. 1) Illumina Inc., San Diego, CA; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX.

High throughput whole-genome sequencing (WGS) allows for rapid identification and classification of all variant types. While high quality benchmarking datasets exist for small variants such as SNPs and indels, the ability to systematically call structural variants and copy number variants (CNVs) is less well characterized since we lack high quality benchmarking datasets. To address this problem for CNVs, we have developed a targeted software pipeline that genotypes previously characterized common CNVs. This approach consists of several stages. First, we aggregate reads across multiple samples to refine the variant boundaries. We then perform targeted calling using these refined boundaries by normalizing read counts against a baseline of reference samples negative for each variant. We test the variant calls using Mendelian errors and Hardy Weinberg Equilibrium (HWE) across the 17-member Platinum Genomes pedigree, 50 Polaris trios and 957 samples of European ancestry. We show the utility of this method by analyzing 3,818 CNVs called in these 1,124 samples by Canvas/Manta and an additional 46 CNVs independently identified from PacBio data in NA12878. 2,441 of these CNVs were common enough that they could be positively assessed by Mendelian inheritance and HWE tests. Of these, we identified 2,083 deletions (257 longer than 10kb) and 146 duplications (64 longer than 10kb) that pass both validations. To further validate these CNVs, we also sequenced 87 of these samples using linked reads produced by the 10x Chromium platform and use the per-haplotype depth as an additional validation of our calls. For example, a deletion should only have reads from one haplotype in the deleted region. Among the CNV calls in these samples, the median ratio between read counts of the minor and major haplotype is 0.014, 0.858 and 0.548 for heterozygous deletion (CN1), reference (CN2) and duplication (CN3) calls, respectively, consistent with their copy number states (one haplotype deleted, both haplotypes present, and one haplotype duplicated). Of the samples analyzed here, 313 are consented for public release and the raw sequence data (40x depth HiSeqX data for all samples and an additional 30x depth linked read data for 87 samples) and their CNV calls are publicly available as a community resource (<https://github.com/Illumina/Polaris>). This CNV truth set includes 194,731 total CNV calls and is a valuable public resource for algorithm benchmarking and clinical genetics research.

1639T

Genome-wide structural variant identification from PromethION long read sequencing data. *W. De Coster¹, A. De Roeck¹, T. De Pooter², S. D'Hert², P. De Rijk², M. Strazisar², C. Van Broeckhoven¹.* 1) Neurodegenerative brain diseases group, Center for Molecular Neurology, VIB & University of Antwerp; 2) Neuromics Support Facility, Center for Molecular Neurology, VIB & University of Antwerp.

The Oxford Nanopore PromethION is an emerging platform for affordable long read whole genome sequencing, including human genomes. Long read sequencing enables structural variation detection with unprecedented resolution, especially in repetitive regions complicating their discovery using currently dominant short read sequencing technologies. For benchmarking of the device and evaluation and development of bioinformatic tools we have sequenced the Yoruban reference genome NA19240 to 80x coverage. We discuss the characteristics and accuracy of the obtained data and the results of a de novo assembly. Making use of the plethora of sequencing and variant data available for this individual we evaluate alignment and structural variant calling software and describe their performance. Additionally, we provide a flexible and efficient computational workflow for cohort-based genome-wide structural variant detection and annotation from long read datasets integrating multiple state-of-the-art tools and provide recommendations for its application in identification of pathogenic structural variation.

1640F

Comprehensive benchmarking of structural variant callers. *R. Littman¹, S. Mangul¹, V. Sarwal¹, M. Distler², E. Eskin¹, J. Flint².* 1) Computer Science Dept, UCLA, Los Angeles, CA; 2) Semel Institute for Neuroscience and Human Behavior, UCLA, Los Angeles, CA.

Structural variance (SV) in genomics including insertions, deletions, copy number variation, translocations, and inversions, have varying pathogenicity of disease. The detection of these variants is necessary for both the detection and treatment of disease. Currently there are no benchmarking studies available, making it difficult to choose proper SV discovery software. The wide spectrum of effects of SV in the genome provides a necessity to discover the most accurate SV detection software. We compared over 70 SV detection tools' abilities to detect deletions, to determine which tools have the best sensitivity and precision. We detected the performance of these tools on 8 different strains of inbred mice in which all SV on chromosome 19 have been detected and confirmed by PCR. Since we have high coverage on these mouse strains we have been able to subsample our data with varying coverage (54x - 1.5x), to determine the effects of coverage on each tool's accuracy. We randomly subsampled ten times to increase the validity of our results. We determined the resolution of each tool by varying the resolution threshold by 100, 1000, and 10,000 base pairs of our gold standard. The resolution is determined as the distance between the gold standard deletion coordinate and that determined by the computation method. As we increase the required distance from the gold standard, generally precision decreased, and sensitivity increased. Generally as coverage increased, precision decreased, and sensitivity increased. The effect of coverage was much more significant between 1.5x and 22x, rather than between 22x and 54x. Many methods reach sensitivity of 100% with precision of 0%, these tools have both a high true positive and false positive rates. Clever is the only sv tool that has a good balance between sensitivity and PPV. Many other tools have high sensitivity, and low PPV, or low sensitivity and high PPV. Our recommendations can help researchers choose the best SV detection software, as well as inform the developer community of the challenges of SV detection, especially with regard to low coverage WGS data.

1641W

FusorSV: An algorithm for optimally combining data from multiple structural variation detection methods. A. Malhotra, L. Lozhovskiy, C. Lee. Jackson Laboratory for Genomic Medicine, Farmington, CT.

Despite the large scope and potential for critical impact on human biology Structural Variations (SVs) remain poorly characterized primarily due to the lack of comprehensive methods for SV detection. To the best of our knowledge, no algorithm comes close to identifying all types of SVs and across all size ranges in human genomes. For example, DELLY and LUMPY appear similar in methodology and produce very similar deletion calls but diverge significantly for inversion calls. To overcome this issue, more recently, certain genome consortia studies have combined multiple SV-calling algorithms into a single pipeline to generate a unified SV callset comprising of primarily overlapping calls. However, the optimized way to combine SV calls from multiple algorithms still remains unclear. As a solution we present Structural Variation Engine (SVE), a computational infrastructure that currently includes an ensemble of eight popular SV calling methods and a novel algorithm – *FusorSV* to merge the SV calls from the ensemble. *FusorSV* employs a data-mining approach to characterize the performance of a group of SV callers against a given truth set. The aim is to select all possible combinations of callers that satisfy a performance threshold. By utilizing 27 deep-coverage human genomes from the 1000 Genomes Project we developed a data fusion model for merging SVs. For these 27 genomes, we identified 843 novel SV calls that were not reported by the 1000 Genomes Project. Experimental validation of a subset of these calls yields a validation rate of 86.7%. These results highlight the power of the *FusorSV* approach and we believe that it has the potential to become a new gold standard for SV calling from short read whole genome sequencing datasets. Currently we are planning to add features to *FusorSV* that will make it easy for users to use any reference genome. *FusorSV* will then be useful for analyzing newer hg38-based datasets as well as model organisms such as the mouse. Furthermore, we wish to train a new *FusorSV* model using newer truth datasets, with a particular focus on using more comprehensive set of SV calls predicted using long read sequencing data. We will deeply investigate the integration of new SV calling methods and the interplay amongst the method to develop a more lightweight version *FusorSV* that can be applied at scale to thousands of genomes. Finally, we will explore the use of machine learning in *FusorSV*'s SV scoring and merging methods to improve performance.

1642T

SV-INFERNO: A Spark based pipeline for INFERRing the molecular mechanisms of Noncoding structural variants. E.E. Mlynarski^{1,2,3}, A. Amalie-Wolf^{1,2,3,4}, P.P. Kuksa^{1,2,3}, O. Valladares^{1,2,3}, G.D. Schellenberg^{1,3}, L.-S. Wang^{1,2,3}. 1) Penn Neurodegeneration Genomics Center, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA., Select a Country; 2) Institute for Biomedical Informatics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA; 3) Department of Pathology and Laboratory Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA; 4) Genomics & Computational Biology Graduate Group, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA.

The GTEx consortium recently demonstrated that noncoding structural variants (SVs) can substantially alter gene expression and are enriched at regulatory elements, highlighting the importance of noncoding SVs in disease pathology. However, such noncoding SVs are poorly understood in part because the few tools available for SVs primarily focus on protein coding genes or previously reported variants, and largely ignore potentially causal noncoding structural variants. Here we introduce SV-INFERNO, the first Spark-based annotation and analysis pipeline for structural variants. SV-INFERNO expands upon our INFERNO method (<http://inferno.lisanwanglab.org/>), which provides a principled approach to harmonize information from diverse data sources across tissues types using a tissue categorization system to prioritize causal noncoding single nucleotide variants and identify the affected regulatory mechanisms, tissue contexts, and target genes in a hypothesis-free manner. Spark, coupled with the INFERNO tissue categorization schema, enables us to efficiently query millions of SVs in a distributed manner against >300 functional genomic datasets in parallel. The pipeline currently includes the following data sources and we will continue to integrate new functional datasets as they become available: enhancers from the Roadmap Epigenomics and FANTOM5 consortia; GTEx eQTLs; TFBS from HOMER and Factorbook; ncRNAs from DASHR (Database of small RNAs, <http://dashr2.lisanwanglab.org/>); and predicted miRNA binding sites from TargetScan. SV-INFERNO converts the SV-annotation queries to Spark RDDs and uses distributed Spark join operations to efficiently identify loci with multiple types of overlapping functional annotations in a tissue context specific manner. The Spark RDDs corresponding to each annotation type/source are then joined to form the SV annotation feature matrix for subsequent SV classification and prioritization using the TensorFlow deep learning framework. SV-INFERNO also includes traditional GSEA, pathway and gene network analyses using overlapping protein coding genes +/- the predicted target genes from applicable functional annotation types. SV-INFERNO is implemented in Python/pyspark and will be available as a webserver and as a Docker container that can be run on any Spark cluster (<http://lisanwanglab.org/SV-INFERNO>).

1643F

A novel approach for structural variant calling: Combining data from whole genome next-generation sequencing and optical mapping. Y. Porat¹, O. Lev¹, M. Einhorn¹, O. Shani¹, E. Vilain², H. Barseghyan², S. Bhat-tacharya², N. Paz-Yaacov¹. 1) Genoox, Palo Alto, CA 94306, USA; 2) Center for Genetic Medicine Research, Children's National Health System, Children's Research Institute, Washington, DC 20010, USA.

Structural variant (SV) detection via short-read sequencing is typically characterized by high false-positive rates and considerable uncertainty regarding variant breakpoint position, owing to the size of alterations typically far exceeding the read length. Thus, alignment patterns between short reads and the reference genome often can provide only limited evidence for the existence of a variation. In addition, variant calling and genotyping employ probability inference surrounding combined alignment signals, which carries inherent uncertainty surrounding exact breakpoint positions and confidence. Optical mapping technology for SV detection results in high sensitivity and specificity but often carries limited certainty of breakpoint positions. We validated a novel joint pipeline for SV detection that integrates next-generation sequencing (NGS) raw data with optical mapping-based SV calls using an advanced reference graph structure. By applying the reference graph structure throughout the analytical pipeline, evidences from both technologies can be considered simultaneously to provide increased breakpoint accuracy and confidence. The novel SV-calling pipeline was validated for deletions and insertions larger than 500bp, using sample data and high-confidence SV truth sets available for NA12878 in Genome in a Bottle (GIAB; <http://jimb.stanford.edu/giab>), by comparing NGS-only, optical-mapping-only, and combined technologies calling. Per optical-mapping-only data, 817 deletions and 1,670 insertions had high breakpoint uncertainty (tens of thousands bp). With application of the combined technologies, the uncertainty of accurate breakpoint positions for the 0-75th percentile of the 756 (92.5%) deletions and 1,248 (75%) insertions called was reduced to only 4bp and 140bp, respectively. Further, results of a specificity evaluation using *in silico* simulations on regions that did not overlap the GIAB truth set indicated 99.3% (971/978) specificity for deletions and 94.3% (910/965) for deletions. As such, the respective false-positive rates were 0.7% (7/978) and 5.7% (55/965). A novel joint pipeline integrating NGS raw data with optical mapping provided SV calls characterized by precise breakpoint positions and low false-positive rates, suggesting the advanced graph structure could have considerable utility in clinical practice.

1644W

Precise common and rare germline CNV calling with GATK. A.N. Smirnov¹, M. Babadi¹, S.K. Lee¹, L.D. Gauthier¹, M.E. Talkowski², H. Brand², X. Zhao². 1) Broad Institute of MIT and Harvard, Cambridge, MA; 2) Massachusetts General Hospital, Boston, MA.

We propose, implement, and evaluate a novel method (GATK gCNV) for the accurate discovery of rare and common copy-number variations (CNVs) from read-depth data derived from either whole exome sequencing (WES) or whole genome sequencing (WGS) data. GATK gCNV particularly outperforms other read-depth methods on whole exome sequencing (WES) data due to a sophisticated Bayesian approach that learns bias factors arising from sequencing and library preparation. GATK gCNV consists of two models - an initial model that infers the ploidy of sex chromosomes and the presence of autosomal aneuploidies, and a primary model that infers bias factors and CNVs, treating GC bias probabilistically and automatically determining the necessary level of model complexity in a data-driven manner. In analyses performed to date, GATK gCNV maintains a high level of sensitivity in common CNV regions due to a hierarchical hidden Markov model used for accurate genotyping of multi-allelic loci. Furthermore, GATK gCNV performs bias modeling and CNV discovery simultaneously and self-consistently, resulting in significantly improved sensitivity and precision. Our implementation of GATK gCNV utilizes the PyMC3/Theano framework for performing automatic differentiation variational inference (ADVI). In addition, GATK gCNV automatically scatters large tasks across multiple machines using the Cromwell/WDL framework, enabling the scalable processing of large cohorts. We have evaluated the ability of GATK gCNV to detect both rare and common coding CNVs in neurodevelopmental disorder cohorts and are continuing the process of benchmarking its sensitivity, precision, and reproducibility for clinical diagnostic applications by direct comparisons to short-read and long-read WGS datasets. GATK gCNV is available in the GATK framework and provides a research tool that could have significant implications for clinical diagnostic genetic testing.

1645T

A deep learning method to increase the accuracy of copy number variation detection from whole-genome sequencing data. B. Trost¹, L. Lee¹, S. Walker¹, E.J. Higginbotham¹, C.M. Nguyen¹, S.W. Scherer^{1,2,3}. 1) The Centre for Applied Genomics, Hospital for Sick Children, Toronto, Ontario, Canada; 2) Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada; 3) McLaughlin Centre, University of Toronto, Toronto, Ontario, Canada.

Copy number variations (CNVs) are major sources of genetic diversity and have been associated with many disorders. Several algorithms for detecting CNVs from whole-genome sequencing (WGS) data are available; however, they have high false discovery rates (FDRs), making it difficult to separate true CNVs from false positives. Previously, we identified ERDS and CNVnator as the most accurate read depth-based CNV-detection algorithms and presented a workflow that reduced the FDR substantially with only a modest reduction in sensitivity. Our workflow partially relied on various simple filters for eliminating false positives, such as using only CNVs identified by both ERDS and CNVnator and discarding CNVs rich in repetitive genetic elements. In this study, we hypothesized that a deep learning model could be used to achieve more accurate prioritization of true CNVs. To test this hypothesis, we developed a method called DNN-CNV that uses a deep neural network to separate real CNVs from false positives based on several features: CNV size and type (deletion or duplication), whether a CNV was detected by ERDS and/or CNVnator, read-depth data, and the repeat elements in and around the predicted CNV. Six genomes with corresponding CNV benchmarks were used to train and test our model. DNN-CNV improved upon the accuracy of previous filters for most combinations of CNV size and type and was particularly effective for duplications, which have the highest false positive rate. For instance, DNN-CNV detected 22% more true and 58% fewer false duplications between 10 and 100 kb compared to the best simple filter. For 5-10 kb deletions, DNN-CNV had the same sensitivity as the best simple filter but a lower FDR (13% versus 17%). Besides increasing accuracy, DNN-CNV offers several other benefits: it obviates the time-consuming process of determining appropriate simple filters; it simplifies the filtering process itself by eliminating the need to use different criteria depending on CNV size and type; and it facilitates modifying the sensitivity/FDR balance depending on application-specific requirements. DNN-CNV should thus prove useful for disease-association studies, population-genetics studies, and other work that depends on reliable CNV detection. DNN-CNV is freely available either as a web server or downloadable source code at <http://dnn-cnv.tcag.ca>.

1646F

DeepCNV: A deep learning approach for detecting copy number variants. Z. Wei¹, X. Hou¹, J. Glessner², M. Khan², P. Sleiman², H. Hakonarson^{2,3}. 1) Department of Computer Science, New Jersey Institute of Technology, Newark, NJ 07102; 2) The Center for Applied Genomics, Abramson Research Center, The Children's Hospital of Philadelphia, Philadelphia, PA; 3) Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, PA 19102.

Copy number variants (CNVs) represent the largest component of structural genomic variations. The recent high-throughput technologies, such as high-density SNP genotyping and DNA sequencing, have made it possible to compressively identify and catalog CNVs. Detection of CNVs has become routine in many genetic studies of disease susceptibility. As a result, many computational methods have been proposed for detecting CNVs. However, it has been shown that the current CNV detection tools still have limited performance, due to the complexity of biological samples, and high level of noise and biases in the high-throughput data. We applied a deep learning approach to overcome these challenges and improving CNV detection performance over existing approaches. To this end, we have compiled a large sample size of manually annotated and validated 4839 CNVs and 9062 controls. We divided these samples into 70:15:15 for training, validation, and testing, respectively. We developed a novel deep neural network named DeepCNV for recognizing CNVs from blood or saliva-derived DNA samples. It is capable of automating feature engineering and eliminates the need of human feature engineering. We obtained an overall AUC of 0.942, MCC of 0.873, and accuracy of 0.943 on the testing dataset. Specifically, for the 788 testing CNVs, DeepCNV has a precision of 0.947, recall of 0.885 and F1 score of 0.915, and a precision of 0.941, recall of 0.974 and F1 score of 0.957 for predicting the 1361 samples. In comparison, SVM using carefully handcrafted features has only an accuracy of 0.905.

1647W

Evaluation of the performance of copy number variation prediction tools for the detection of deletions from whole genome sequencing data. *W. Whitford^{1,2}, K. Lehnert^{1,2}, R.G. Snell^{1,2}, J.C. Jacobsen^{1,2}.* 1) School of Biological Sciences, The University of Auckland, New Zealand; 2) Centre for Brain Research, The University of Auckland, New Zealand.

Whole genome sequencing (WGS) has seen both an increase in popularity and concurrent decrease in cost over the past decade, rendering this approach a viable and sensitive method for variant detection. In addition to its utility for single nucleotide variant detection, WGS data has the potential to detect Copy Number Variants (CNV) to fine resolution. Many CNV detection software packages have been developed which exploit four main types of data: read pair, split read, read depth, and assembly based methods. The aim of this investigation was to evaluate the efficiency of each of these main approaches, including a commonly used combinatorial approach. WGS and CNV location data for the individual NA12878 from the Genome in a Bottle consortium was used as the benchmark dataset. The performance of Breakdancer (read pair), CNVnator (read depth), Delly (read pair and split read), FermiKit (assembly), and Pindel (split read) was assessed by comparing the accuracy and sensitivity of deletion calls $\geq 1\text{kb}$ made by each software package. There was wide variability in the outputs of the different WGS CNV detection programs. The best performance was seen from Breakdancer and Delly, with 92.6% and 96.7% sensitivity, respectively and 34.5% and 68.5% false discovery rate (FDR), respectively. In comparison, CNVnator, FermiKit, and Pindel were less effective with sensitivities of 66.0%, 15.8%, and 69.1%, respectively and FDR of 69.0%, 31.7%, and 91.3%, respectively. Concordance across software packages was poor, with only 27 of the total 612 benchmark CNVs identified by all five methodologies. The WGS based CNV detection tools evaluated show disparate performance in identifying deletions $\geq 1\text{kb}$, particularly those utilising different data input characteristics. Software that exploits read pair based data showed the highest sensitivity, namely Breakdancer and Delly. Breakdancer also displayed the second lowest false discovery rate. Therefore, in this analysis read pair methods (Breakdancer in particular) were the optimal approach for the identification of deletions $\geq 1\text{kb}$, balancing accuracy and sensitivity. There is clear potential for improvement in the detection algorithms, particularly in reducing FDR. This analysis has validated the utility of WGS based CNV detection software to reliably identify deletions, and these findings will be of use when choosing appropriate software for CNV detection, for use in both research and diagnostic medicine.

1648T

How well can you detect structural variants: Towards a standard framework to benchmark human structural variation. *J.M. Zook¹, L. Chapman¹, N.F. Hansen², F.J. Sedlazeck³, A. Wenger⁴, A. English⁵, C. Xiao⁶, J. Oliver⁷, A. Hastie⁸, I.T. Fiddes⁹, A.M. Barrio⁹, T. Marschall¹⁰, M. Chaisson¹¹, J.J. Farrell¹², A. Carroll¹³, M. Salit¹⁴, Genome in a Bottle Consortium.* 1) National Institute of Standards and Technology, Gaithersburg, MD; 2) Cancer Genetics and Comparative Genomics Branch, National Human Genome Research Institute, Rockville, MD; 3) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 4) Pacific Biosciences, Menlo Park, CA; 5) Spiral Genetics, Seattle, WA; 6) National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD; 7) Nabsys, Providence, RI; 8) BioNano Genomics, San Diego, CA; 9) 10x Genomics, Pleasanton, CA; 10) Max Planck Institute, Munich, Germany; 11) University of Southern California, Los Angeles, CA; 12) Boston University Medical School, Boston, MA; 13) DNAnexus, San Francisco, CA; 14) Joint Initiative for Metrology in Biology, Palo Alto, CA.

The Genome in a Bottle Consortium (GIAB) has characterized an Ashkenazi trio from the Personal Genome Project (NIST Reference Material 8392) with short, long, and linked read sequencing and optical/nanopore mapping methods. In this open science project, a community of experts has contributed public data and analyses of structural variants (SVs). Experts contributed SV calls from over 30 mapping- and assembly-based variant callers from 5 technologies. Sequence-resolved insertions and deletions were integrated using sequence-based, repeat-aware clustering methods in SVanalyzer. We have evaluated support for the merged calls using svviz with Illumina paired-end and mate-pair reads as well as 10x Genomics and haplotype-partitioned PacBio. We used BioNano and Nabsys to confirm size predictions of large insertions and deletions. We launched the web-based SVCurator.com platform to enable a community manual curation of our candidate SVs. The community has evaluated multiple iterations of draft benchmark SV calls for their utility, which has led to our current benchmark callset with two tiers: (1) ~13,000 isolated, non-complex, sequence-resolved SVs, and (2) ~7,000 likely SV regions that are complex with multiple calls or with unresolved size or sequence. In addition, we have used four diploid assemblies to develop the first "high-confidence regions" for SV benchmarking, covering ~2.7 Gbp of the genome in which our characterization of SVs is close to comprehensive, which allow users to assess false positives as well as false negatives when benchmarking. We have developed a tool truvari to enable users to compare any SV callset to our benchmark calls and regions at a user-defined sequence-based, size-based, or type-based stringency. Calls are annotated as tandem repeat expansions/contractions, mobile element insertions, etc. to enable stratification of performance by variant type, size, and genome context. These results represent a significant step in GIAB work towards improved benchmarking of large variants in research and clinical settings. These benchmarks are already enabling technology and bioinformatics development, optimization, and demonstrations. All data, analyses, and benchmark calls are publicly available and will continue to be improved based on user feedback and as sequencing data and analysis methods improve.

1649F

Polaris: Extensive resources from multiple technologies for next generation sequencing analysis. M.A. Bekritsky¹, J. Leng², L. Janin¹, N. Johnson², B. Lajoie², A. Moshrefi², T. Truong², E. Jaeger², F. Chen², G.P. Schroth², D.R. Bentley¹, M.A. Eberle¹. 1) Illumina Cambridge Ltd., Chesterford Research Park, Saffron Walden, Essex, CB10 1XL, United Kingdom; 2) Illumina Inc., 5200 Illumina Way, San Diego, CA 92122, USA; 3) Illumina Inc., 200 Lincoln Centre Drive, Foster City, CA 94404, USA.

As next generation sequencing becomes more commonplace, increasing access to whole genome sequencing data becomes an imperative. As has been previously demonstrated by Illumina's Platinum Genomes and the NIST Genome in a Bottle Consortium, readily available sequencing data drives improvements to downstream analysis as informatics developers are able to train and test their methods on the most current data. The goal of the Polaris project is to provide the research community with access to readily available sequence data. To do this, we sequenced DNA from 270 samples obtained from the Coriell Institute that are consented for full release of personally identifiable genetic information. These samples were originally collected as part of the HapMap, 1000 Genomes, and Personal Genomes projects and have been sequenced to an average of 40x depth using Illumina's HiSeqX platform with TruSeq PCR-Free sample preparation. We have grouped our current samples and publicly released this data in four panels (ENA Study IDs are in parentheses): 1. **Diversity Panel (PRJEB20654)**: 150 unrelated samples from 3 super-populations (European, East Asian, and African)². 2. **Trio Panel (PRJEB25009)**: 50 children whose parents are part of the Diversity Panel, enabling extensive trio analyses³. 3. **PGx Panel (PRJEB19931)**: 70 samples with extensive orthogonal characterizations of pharmacogenetic variants performed as part of the CDC's GeT-RM program⁴. 4. **PGx 10X[®] Panel (PRJEB26950)**: PGx panel samples that have been prepared with 10X's Chromium controller and sequenced on the HiSeq4000, providing linked read information. The sequence data for every sample from these panels is available on ENA as part of the Polaris umbrella project (PRJEB23074) and we will continue to add more sequence data for additional samples and new technologies under this umbrella project. Updates to this project will enable novel analyses and validation methods on new platforms. We will also be making these WGS panels available on BaseSpace along with the variant calls made using the latest bioinformatics pipelines developed by Illumina, including Edico's Dragen platform. As the pace of genomic innovation continues to accelerate, Illumina's Polaris project will enable researchers from diverse backgrounds and with varying access to sequencing resources to explore, understand, and contribute to our shared understanding of the role of genomics in human health.

1650W

Curating clinically relevant variants at scale: ClinGen's Variant Curation Interface and opportunities for data mining the clinical genome. S. Dwight¹, M. Wright¹, C. Thomas¹, B. Wulf¹, J. Zhen¹, S. Prabhu¹, H. Wand¹, D. Azzariti², L. Babb², C. Bizon², M. DiStefano², S. Harrison², T. Nelson², A. Nie¹, R. Patel⁵, P. Pawliczek⁵, E. Rooney Riggs⁴, A. Milosavljevic⁵, S.E. Plon⁵, H.L. Rehm^{3,7,8,9}, J. Zou¹, G. Bejerano¹, C. Bustamante¹, *Clinical Genome Resource (ClinGen)*. 1) Department of Biomedical Data Science, Stanford University, Stanford, CA; 2) Laboratory for Molecular Medicine, Partners Healthcare Personalized Medicine, Cambridge, MA; 3) Renaissance Computing Institute, University of North Carolina, Chapel Hill, NC; 4) Geisinger Health System, Danville, PA; 5) Baylor College of Medicine, Houston, TX; 6) Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA; 7) Brigham and Women's Hospital, Boston, MA; 8) Harvard Medical School, Boston, MA; 9) Broad Institute, Cambridge, MA.

Over the coming decade, clinical genome sequencing will identify millions of genetic variants and query their potential association with disease phenotypes. Therefore, it is imperative for us as a community to create gold-standard curation of individual variants at scale in the context of clinical care, a task made more challenging by "data silos." Central to the mission of the Clinical Genome Resource (ClinGen; clinicalgenome.org) is breaking down these silos and allowing for the accurate and consistent curation of the clinical relevance of genes and variants to inform precision medicine and research. ClinGen's Variant Curation Interface (VCI) is designed to support routine application of the ACMG/AMP Guidelines as well as application of gene and disease specifications of the ACMG/AMP Guidelines by ClinGen's Expert Panels, and thus serves as a platform for gold-standard curation of clinical variants. It achieves this by aggregating evidence in a normalized manner from external resources and by supporting the manual curation of evidence in a structured manner. The VCI is now also publicly accessible to any user wishing to interpret the clinical relevance of genetic variants using, as well as contributing to, a shared evidence pool. We describe here 1) efforts aimed at large-scale identification and "tagging" of relevant literature with standardized terminologies to scale curation efficiency within the VCI, and 2) a ClinVar submission workflow supported by ClinGen's common data model, establishing the VCI as a thoroughfare for curation and submission/resubmission of standardized variant interpretations to ClinVar. The above workflow facilitates the collection of shared, normalized variant evidence from and across a vast array of publications, clinical and research laboratories, clinicians, and genomics databases, providing an ideal forum for discrepancy resolution and a rich dataset for subsequent machine learning that will feed back to further increase curation efficiency. Importantly, the VCI integrates curation and ClinVar submission by programmatically converting the ACMG/AMP criteria evaluations and supporting evidence into a ClinVar submission format. This facilitates submission to ClinVar, a public resource, and therefore public access to standardized interpretations of clinical variants and the supporting evidence behind them.

1651T

Resources for archiving, sharing and analysis of controlled-access human data at DDBJ Center. Y. Kodama¹, O. Ogasawara¹, A. Fukuda¹, J. Mashima¹, M.T. Minowa², M. Kawashima², L. Toyo-oka², N. Mitsuhashi², K. Okubo¹, Y. Nakamura¹, T. Takagi^{1,2}, M. Arita¹. 1) DDBJ Center, National Institute of Genetics, Shizuoka 411-8540, Japan; 2) National Bioscience Database Center, Japan Science and Technology Agency, Tokyo 102-8666, Japan.

The Japanese Genotype-phenotype Archive (JGA, <https://www.ddbj.nig.ac.jp/jga>) provides authorized access to individual-level phenotype and genotype data as a centralized repository in collaboration with the National Bioscience Database Center (NBDC) of the Japan Science and Technology Agency (<https://humandbs.biosciencedbc.jp>). NBDC has established the ethical guidelines and policies, and its Data Access Committee (DAC) grants data submission and data access requests to JGA. JGA accepts metadata, raw and processed data from array-based and next-generation sequencing platforms, clinical images, and phenotype data associated with data samples. For submitted studies and metadata, JGA assigns stable, unique identifiers prefixed with 'JGA'. Summaries of available JGA studies are listed at <https://ddbj.nig.ac.jp/jga/viewer/view/studies>. As of 4 June 2018, total 85 studies are available. The JGA system is connected to our secure supercomputer through a high-speed network so that users can smoothly download and analyze data with pre-installed bioinformatics tools on the supercomputer. The computing system is compliant with the NBDC security guidelines. To enhance researchers' collaboration through sharing and analyzing pre-publication data, we also started a new service "DDBJ Group Cloud (DGC)" since February 2017. The first use is the "AMED Genome Sharing Database (AGD)" for human genomes among researchers funded by AMED (Japan Agency for Medical Research and Development). Because AGD and JGA use the same data model, users can easily transfer the AGD data to JGA upon publication. In the poster, we present recent developments of our systems for archiving, sharing and analysis of controlled-access human data.

1652F

Data integration techniques to elucidate connections between environmental factors, genetic variants, and human diseases. M.B. Kosnik^{1,2,3}, A. Planchart^{1,3,4}, D.M. Reif^{1,2,3,4}, C.J. Mattingly^{1,3,4}. 1) Toxicology Program, North Carolina State University, Raleigh, NC; 2) Bioinformatics Research Center, North Carolina State University, Raleigh, NC; 3) Department of Biological Sciences, North Carolina State University, Raleigh, NC; 4) Center for Human Health and the Environment, North Carolina State University, Raleigh, NC.

Addressing the complex relationship between genetic influences on disease development and environmental factors is a challenge that requires multiple types and sources of data. Chemical exposure data from high-throughput screening (HTS) efforts such as the Tox21/ToxCast program identify chemical hazard using *in vitro* assays to probe toxicity. While most of these assays target specific genes, assessing how these chemicals exert effects through genetic interactions remains challenging. Integration with datasets like the Comparative Toxicogenomics Database (CTD), a publicly available database that builds networks of chemical, gene, and disease information from curated literature sources, may help to resolve these questions by providing broader biological context to HTS data. Further context for the interplay of gene-environment (GxE) interactions can be provided to disease connections formed between HTS data and CTD by joining the resulting chemical-gene-disease associations to resources like ClinVar, an archive of clinically relevant genetic variants implicated in disease conditions. By linking these datasets together, new chemical-disease associations can be found leading to new hypotheses for GxE interactions involved in disease progression. Here, we developed a technique for integrating these three resources together by linking elements common to the databases and used these integrated data to identify new associations between environmental factors, diseases, and genetic variants to promote discovery. By statistically testing disease enrichment using chemical-gene data from HTS and gene-disease data from CTD, we identified new chemical-disease associations between 2,781 chemicals and 776 diseases, with a median of 129 new, statistically significant ($p < 0.01$) disease associations per chemical. For 833 of these chemicals, we also found new, statistically significant ($p < 0.01$) genetic support for existing chemical-disease associations in a median of 44 diseases per chemical. We connected these new chemical-gene-disease associations to ClinVar gene-condition-variant data and identified a median of 5 variants per chemical-gene-disease association across 61 unique gene-disease combinations to form new hypotheses about GxE interactions involved in disease progression. Thus, we demonstrate a novel method to integrate existing biological data and identify new associations between environmental factors, diseases, and genetic variants to promote discovery.

1653W

NCBI services for variation normalization, remapping, and annotation. L.

Phan, E. Moyer, E. Ivanchenko, D. Revoe, H. Zhang, W. Qiang, E. Shekhtman, D. Shao, M. Ward, A. Glodek, J.B. Holmes, B.L. Kattman. National Center for Biotechnology Information, U.S. National Library of Medicine 8600 Rockville Pike, Bethesda MD, 20894 USA.

Variant normalization and annotation are crucial steps to understanding the genetic basis of their effects on disease and biological functions. Normalizing variants and matching them with data in public variation databases, such as dbSNP and ClinVar, are central to the discovery efforts to identify known and novel variants. However, this process can be difficult because identical variants can be represented differently by variation resources and tools leading to duplications and complicating downstream analysis and annotation. Various tools have been proposed to address variant normalization but they are not robust handling variants in different formats and reporting on different reference sequence coordinates. Here we present a suite of NCBI variation tools (<https://www.ncbi.nlm.nih.gov/projects/variation/services/v0/>) that 1) normalizes representation of genetic variants from dbSNP rs IDs, HGVS expressions, and VCF formats for matching existing variants in dbSNP and other public resources and to identify novel variants, 2) remapping of variants between assembly versions and between coordinates on mRNA, protein, and genomic sequences to a normalized sequence coordinate, and 3) provide annotations from dbSNP and ClinVar. The tools employ a new algorithm and data model called Sequence Position Deletion Insertion (SPDI) to describe sequence changes and variant normalization which allow for:

- Producing a contextual allele, a new feature of our SPDI format that corrects left/right-shifting or shuffling for insertions or deletions
- Transforming data into right-shifted HGVS
- Transforming data into left-shifted VCF fields
- Remap (or lift-over) a variant to all available locations based on the alignment dataset used by ClinVar and dbSNP
- Remap (or lift-over) to a canonical representative at a single location that you can use to group identical variants

Acknowledgments Work at NCBI is supported by the NIH Intramural Research Program and the National Library of Medicine.

1654T

MyVariant.info: Community-aggregated variant annotations as a service.

J. Xin¹, S. Lelong¹, C. Afrasiabi¹, M. Cano¹, B. Jeon¹, G. Tsueng¹, S. Mooney², A. Su¹, C. Wu¹. 1) Department of Integrative Structural and Computational Biology, The Scripps Research Institute, LA JOLLA, CA; 2) Department of Biomedical and Health Informatics, University of Washington, Seattle, WA.

The accumulation of genetic variant annotations has been increasing explosively with the recent advance of the high-throughput sequencing technology. These annotations are sprinkled across dozens of data repositories like NCBI, Ensembl, UniProt, and hundreds of other specialized databases. While the volume and breadth of annotations is valuable, their fragmentation across many data silos is often frustrating and inefficient. Bioinformaticians everywhere must continuously and repetitively engage in data wrangling in an effort to comprehensively integrate knowledge from all these resources, and these uncoordinated efforts represent an enormous duplication of work. We created a platform, called MyVariant.info, to aggregate variant-specific annotations from community resources and provide high-performance programmatic access. Annotations from each resource are first converted into JSON-based objects with their id fields as the canonical names following HGVS nomenclature (genomic DNA based). This scheme allows merging of all annotations relevant to a unique variant into a single annotation object. A high-performance and scalable query engine was built to index the merged annotation objects and provides programmatic access to the developers. MyVariant.info is an open source project and we are actively expanding the coverage by engaging community efforts. Integrating a new variant annotation resource into MyVariant.info is extremely simple, which is to build a parser to convert from the original data source file into a JSON document, with the primary key represented by HGVS ID. And the parser could be contributed not only by the data provider but basically anyone who is interested. Once a parser is contributed, the sophisticated query-engine we built can save users effort to build their own infrastructure, which provides the incentive for them to contribute. As of today, we are able to aggregate annotations for around 900 million unique variants from 19 diversified resources, including well-known resources like dbSNP, ClinVar, as well as specialized ones like CIVIC and GRASP. In addition, we have built a scheduling system that automates the updates for each data source according to its own schedule. Queries to MyVariant.info can be performed through our web-based API at <http://myvariant.info> or by using data access libraries for Python, R or JavaScript. MyVariant.info currently handles more than 1.3 million monthly user queries from thousands of unique IPs.

1655F

Genetic variation influences ground state pluripotency in mouse embryonic stem cells through a hierarchy of molecular phenotypes. D.A. Skelly¹, C. Byers¹, A. Czechanski¹, C. Spruce¹, K. Choi¹, D.M. Gatti¹, S. Aydin¹, A. Stanton¹, D. Porter², T. Choi², G.A. Churchill¹, S.C. Munger¹, C.L. Baker¹, L.G. Reinholdt¹. 1) The Jackson Laboratory, Bar Harbor, ME 04609; 2) Predictive Biology, Inc., Carlsbad, CA 92010.

The influence of genetic variation on differentiation propensity of human pluripotent stem cell lines has important consequences for regenerative medicine and the study of early development. Mouse embryonic stem cells (mESCs) grown under controlled conditions occupy a ground state where pluripotency-associated transcriptional and epigenetic circuitry are highly active. We observed variability in expression of core pluripotency genes in genetically diverse mESCs grown in media lacking inhibitors that reinforce the ground state. To dissect the genetic basis of this variation in stability of ground state pluripotency, we conducted genetic analysis of gene expression and chromatin accessibility in mESCs grown under sensitized conditions (LIF + GSK3 inhibitor). We profiled gene expression and chromatin accessibility in 185 lines derived from genetically heterogeneous Diversity Outbred mice. We mapped thousands of loci that affect chromatin accessibility (caQTL) and/or transcript abundance (eQTL). We found eleven instances where distant QTL co-localized in clusters, suggesting a common regulator. One locus on chromosome 15 controlled the expression of 208 genes including many with known roles in maintenance of pluripotency. We applied causal mediation analysis and identified *Lifr* (leukemia inhibitory factor receptor) transcript abundance as the causal intermediate for these eQTL. Moreover, a joint mediation analysis of gene expression with chromatin accessibility yielded a single peak upstream of *Lifr* containing one SNP with a strain distribution pattern matching founder haplotype effects on *Lifr* downstream targets. Functional assays confirmed that *Lifr* genotype influences self-renewal capacity in an orthogonal set of mESCs derived from recombinant inbred lines. Furthermore, allele swap experiments in mESC lines with opposing alleles upstream of *Lifr* verified that the SNP we identified strongly influences mESC pluripotency. Thus, we detected a causal chain of molecular events: a single SNP modulates regulatory element accessibility, which affects *Lifr* expression, leading to large-scale transcriptional shifts including known and novel pluripotency-associated genes. These results reveal mechanistic details underpinning variability in stability of ground state pluripotency in mESCs, highlight the power of diverse mouse populations for dissecting molecular regulatory networks, and represent a resource for study of molecular events driving early development.

1656W

Identification of novel epigenetically regulated genes and pathways involved in high myopia. C. Uppala¹, K.I. Lakhtaria¹, S.S. Chettiar², S.

Visweswaraiah³, U. Radhakrishna³. 1) Computer sciences, Gujarat University, Ahmedabad, Gujarat, India; 2) Department of zoology, School of sciences, Gujarat University, Ahmedabad-9, India; 3) Department of Obstetrics and Gynecology, Oakland University William Beaumont School of Medicine, Royal Oak, MI, USA.

Myopia or nearsightedness is the most prevalent form of refractive error. In myopic subjects, the image of distant objects falls in front of the retina and either the eye is too long (axial myopia), the cornea is too convex, or the index of refraction of the lens is too high (refractive myopia). The risk of threats to vision can also occur in patients with moderate and/or high-grade myopia. In United States, about 24% of the >40 years aged population is affected by myopia and the estimation predicts that, myopia will affect about 45 million US people by 2050 and about 5 billion people would be myopic globally. However, the prevalence of myopia varies with age, sex, race, occupation, ethnicity, and environmental effects. Its highest prevalence rates are in East Asians, in whom between 80 to 90% of the young adult populations are myopic confirming that both genetic and environmental factors influence the incidence of myopia. Until now there are no genome-wide epigenetic association studies related to high myopia subjects were reported. To elucidate the molecular mechanism(s) underlying the pathophysiology of high-grade myopia, we have conducted a genome-wide methylation analysis on 18 high myopia subjects and an equal number of controls using Illumina epic array. We identified 1544 CpG sites in 1782 unique genes that were differentially hypermethylated (at least 2.0-fold or higher) in high myopia subjects compared to healthy controls. Among the 1544 unique targets, the top 10 hypermethylated targets were cg26526312 (*LINGO1*), cg27541540 (*PTPN11*), cg00609363 (*ZNRD1*), cg24877391 (*PEX-13;PUS10*), cg21790796 (*KIF20A;BRD8*), cg14282407 (*TRAPPC1;CNTROB*), cg18191664 (*ERLIN2*), cg08048517 (*KIAA0528*), cg10646633 (*ZNF224*) and cg14694176 (*TCEA1*). This study will shed light on the etiology of myopia and eye growth patterns and thus has considerable public health application, as the high prevalence of myopia in the world is a formidable health care challenge. Furthermore, the identification of myopia causing genes will be important for improving myopia diagnosis and counseling, promoting the development of therapeutic strategies and increasing our understanding of the normal and abnormal development of the affected tissues. .

1657T

NGS studies on non-model organisms using StrandNGS. V. Chakraborty, S. Ramachandrala, D. Chohan, A. Adewar, S. Cherukuri. Strand Life Sciences Private Limited, Bengaluru, Karnataka, India.

Next Generation Sequencing (NGS) is constantly growing and becoming a popular tool for quantitative studies. Model organisms are widely used in this research methodology as they are well studied which helps researchers answer various biological questions. As the cost of sequencing techniques continue to plummet, more researchers are leaning towards sequencing of non-model organisms to identify interesting genetic traits. Sequencing analysis of non-model organisms addresses questions related to genetic evolution, comparative genetics, genetic breeding, ecology, and so on. However, as these species have less gene annotations and functional knowledge, the analysis is a much more technically involved process. We elaborate an intuitive and user friendly way to perform sequencing analysis with non-model organisms using the StrandNGS bioinformatics software. In StrandNGS, the Annotation Manager is a central interface for updating, deleting, or adding custom annotations and it also contains a repository of annotations for various organisms. The software is flexible and supports creation of custom builds using text, GenBank and fasta files coming from various sources namely NCBI, Ensembl, UCSC or any other source. For this case study, we analyzed two types of datasets from *Callithrix jacchus* (Marmoset): (A) 12 RNA-Seq raw samples generated using Illumina platform (B) already analyzed data present in text format. These datasets were a part of a bigger study from the publicly available dataset GSE50747. The reference build was created using the fasta and the annotations were updated using gtf files from Ensembl. An end to end analysis including alignment, quantification, filtering, and differential analysis was performed. A multi-omics approach was then applied to understand the correlation between the results obtained from the two types of datasets. The significant genes identified by our analysis matched with some genes highlighted by the research group and reported to be conserved among other mammalian species. To further perceive the conservation of these genes, pathway analysis was done by mapping these genes to pathways of model mammalian organisms. This was accomplished in StrandNGS by enhancing the HomoloGene annotations which primarily identifies homologs of a gene in diverse organisms. As the study of non-model organisms is trending, our intention was to showcase StrandNGS as a tool to effortlessly perform a comprehensive analysis on these organisms.

1658F

Optimal correction for hidden covariates improves eQTL-based characterization of disease risk variants. M.J. Gloudemans¹, N.R. Gay², B. Liu³, A.S. Rao⁴, S.B. Montgomery^{2,5}. 1) Biomedical Informatics Training Program, Stanford University, Stanford, CA; 2) Department of Genetics, Stanford University, Stanford, CA; 3) Department of Biology, Stanford University, Stanford, CA; 4) Department of Bioengineering, Stanford University, Stanford, CA; 5) Department of Pathology, Stanford University, Stanford, CA.

Recent dissemination of publicly accessible results from genome-wide association studies (GWAS) has enabled functional investigation of putative disease risk variants. For example, a GWAS risk variant that also affects gene expression as an expression quantitative trait locus (eQTL) may affect disease risk through a mechanism involving the regulated gene. The quality of eQTL calling is often hindered by both biological and technical batch effects including known and unknown covariates, such as the subject's age, or tissue sampling heterogeneity. Using the method PEER to identify and correct for such covariates increases the number and significance of identified eQTLs (Stegle et al. 2012); therefore, application of PEER or related methods has become standard practice for eQTL analysis. However, it remains unclear to what extent the inclusion of PEER covariates in an eQTL model affects the overlap of eQTLs with disease-relevant GWAS variants, as well as exactly how many PEER factors should be included to achieve optimal interpretation. To answer these questions, we evaluated the overlap between eQTLs and disease risk variants for 15 recent GWAS with high SNP density, large study sample size, and many genome-wide significant hits. We called eQTLs in 922 human whole blood RNA-seq samples from the Depression Genes and Networks (DGN) study and in the GTEx project, including every number of PEER factors from 0 to 60. We then ran several commonly used GWAS-eQTL colocalization tools (eCAVIAR, coloc, SMR) at every level of correction (Hormozdiari et al. 2016, Giambartolomei et al. 2014, Zhu et al. 2016). Strength of colocalization increased at some loci and diminished at others with respect to the number of modeled PEER factors. Furthermore, we observed considerable differences between the behavior of the three colocalization methods, suggesting that the choice of colocalization method is a critical factor for interpreting GWAS variants. Our results illustrate that eQTL pipelines should carefully consider the extent of hidden covariate removal, in order to optimize not only the total number of eQTLs identified, but also the biological relevance of these eQTLs in relation to known disease risk variants. To address this, we propose a new method to optimize colocalization discoveries by identifying and removing hidden covariates significantly associated with sample genotypes, and we demonstrate its utility by discovering robust eQTL-GWAS colocalizations.

1659W

An ultra-fast approach to identify eQTLs with many tissues leveraging a linear time mixed-model algorithm. B. Jew¹, B. Han², S. Sankararaman^{3,4}, J.H. Sul⁵. 1) Bioinformatics Interdepartmental Program, University of California, Los Angeles, CA, USA; 2) Department of Medicine, Seoul National University College of Medicine, Seoul, Korea; 3) Department of Computer Science, University of California, Los Angeles, CA, USA; 4) Department of Human Genetics, University of California, Los Angeles, CA, USA; 5) Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles, CA, USA.

Large scale generation of genotyping and gene expression data allows researchers to better understand the genetic basis for traits and disease by identifying genetic variants linked to gene expression levels, or expression quantitative trait loci (eQTLs), that can reveal key insight into the regulatory architecture that drives complex phenotypes. Regulation of gene expression varies by tissue and cell type and to study this variation, the Genotype-Tissue Expression (GTEx) project is generating genotype and tissue-specific gene expression measurements. This data has allowed researchers to significantly expand our understanding of local (cis-eQTLs) and distal (trans-eQTLs) effects on traits, and evidence suggests that sample size is a limiting factor in identifying these regulatory elements. One approach to increase power to detect eQTLs is to combine the tissue-specific measurements from each individual across multiple tissues to simulate larger sample sizes. Meta-Tissue (Sul et al. PLOS Genetics, 2013) is a statistical framework that effectively utilizes multiple tissue samples in these datasets for identifying eQTLs through linear mixed model and meta-analytic methods. However, the feasibility of this method diminishes with increasing sample sizes due to computational time that grows cubically with each sample. For example, Meta-Tissue in current implementation will take years of computational time to analyze the latest GTEx dataset. The main computational bottleneck is estimating a variance component of each SNP-gene pair in the linear mixed model (LMM) framework. To overcome this challenge, we propose an efficient algorithm to estimate variance components in LMM. Our method leverages the structured format of data in LMM because gene expression measurements are usually collected from many tissues in each individual. Our algorithm scales linearly with sample size and generates results identical to Meta-Tissue when all tissues are collected from each individual. At current sample sizes, this method is thousands of times faster than the current Meta-Tissue implementation, and hence will enable detection of eQTL from many tissues like those in GTEx. We are currently extending the algorithm to consider a situation when different sets of tissues are collected from individuals. This method will accommodate current and future studies with increasing sample sizes required to identify novel eQTLs from many tissues.

1660T

The eQTLs catalog and LinDA browser: A platform for determining the effects on transcription of GWAS variants. S. Onano^{1,2,3}, F. Cucca^{1,2}, M. Pala². 1) Department of Biomedical Science, Università degli Studi di Sassari, Sassari, Italy; 2) National Research Council (CNR) Institute for Genetic and Biomedical Research (IRGB), Cagliari, Italy; 3) Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor (MI), USA.

The expression Quantitative Traits Loci (eQTLs) are genetic polymorphisms associated with changes in gene expression levels. They have been successfully used to prioritize the target genes of the variants associated with complex traits and diseases (GWAS variants). Up to date a few eQTLs databases exist and they collect only a small portion of the available datasets. We thus planned to build the largest publically available catalog of eQTLs, coupled with a browser, to optimize and simplify their interrogation. We collected and manually curated 51 eQTL public studies ranging from 2007 to date, corresponding to more than 94 sample types and 15 human populations for a total of 285,157 cis-eQTLs and 33,374 genes with at least one cis-eQTL (cis-eGenes). Most of the eQTLs studies were conducted in blood samples from healthy individuals of European ancestry. We found that for 93% of the known protein-coding genes were eGenes, 22% of them intersecting ($r^2 \geq 0.8$) with the NHGRI-EBI GWAS Catalog and 26% of whom considered as druggable. Furthermore, for those GWAS variants for which at least an eGene was known, we found that the NHGRI-EBI GWAS Catalog proposed at least one of the same genes as candidate target only for the 68% of the times. Our eQTL-Catalog can be used as a reference to measure the degree of novelty for future eQTLs studies; it is provided within a platform with a web interface (LinDA) that we plan to implement with other types of quantitative traits (i.e. epigenetic, proteomic, metabolomics and microbiota) to better dissect the pleiotropy of the GWAS variants.

1661F

Design and analysis of FACS-based CRISPR screens. C.G. de Boer, J. Ray, C. Fulco, N. Hacohen, A. Regev. Broad Institute, Cambridge, MA.

CRISPRi (interference) and CRISPRa (activation) can be used to ectopically activate or repress regulatory regions such as enhancers. Here, a cleavage-defective Cas9 protein tethered to a transcriptional activation or repression domain is targeted to a locus with complementary guide RNAs. Combining pooled CRISPR screens with a gene expression output is a powerful approach to identify regulators¹, and regulatory regions^{2,3,4} controlling gene expression, by allowing us to test potential associations for many genes/loci in a single experiment. Here, I will discuss considerations in experimental and library design, and introduce models and tools for the analysis of FACS-based CRISPR screens. 1. Parnas O et al. A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks. *Cell*. 2015 Jul 30;162(3):675-86. doi: 10.1016/j.cell.2015.06.059. Epub 2015 Jul 16. 2. Simeonov DR et al. Discovery of stimulation-responsive immune enhancers with CRISPR activation. *Nature*. 2017 Sep 7;549(7670):111-115. doi: 10.1038/nature23875. Epub 2017 Aug 30. 3. Klann TS et al. CRISPR-Cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome. *Nat Biotechnol*. 2017 Jun;35(6):561-568. doi: 10.1038/nbt.3853. Epub 2017 Apr 3. 4. Canver MC et al. BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature*. 2015 Nov 12;527(7577):192-7. doi: 10.1038/nature15521. Epub 2015 Sep 16.

1662W

Bioinformatics approaches for quantifying endogenous and transgene mRNA expression by RNA sequencing in gene therapy studies. E.K. Tsang, J.T. Gray, C. Chaivorapol. Audentes Therapeutics, South San Francisco, CA.

Gene therapy has demonstrated profound clinical benefit for patients with serious, life-threatening rare genetic diseases, and AAV vectors are the most advanced tool for delivering genetic payloads to terminally differentiated somatic cells in liver and muscle. Clinical development of these products will benefit from a more precise understanding of transcriptional changes induced by the vector in tissues, including characterization of mRNA expression of the endogenous mutated gene and of the transgene. RNA-seq is a cost-effective way to simultaneously measure transgene and endogenous gene expression and the effects of gene therapy at a transcriptome-wide level. A challenge with ascertaining transgene expression simultaneously with that of the endogenous gene is that the two sequences may or may not differ significantly in sequence, and quantification methods must work in the context of wide genetic variation across patients with the same disease, ranging from single nucleotide mutations to deletions and skipped exons. Here, we present bioinformatics approaches for quantifying endogenous and transgene expression from RNA-seq data at varying levels of sequence divergence between the two transcripts. Using a combination of *in silico* simulations and data from animal studies, we assess two methods for distinguishing endogenous from transgene expression: 1) estimating transcript abundances with Salmon, and 2) directly counting reads that align to sequences where the transgene and endogenous transcripts differ. We find that Salmon works well when the two sequences are sufficiently different, but it incorrectly partitions transgene and endogenous expression when the sequences are similar. The more targeted counting approach works even when the sequences differ by a single nucleotide, but only provides coverage at a limited number of sites that may be influenced by sequencing platform and alignment algorithm biases. To guide when each approach should be used, we quantify the level of sequence divergence for Salmon to provide accurate transcript abundance estimates. Our results demonstrate that RNA-seq can be used to reliably assess transgene expression even when the endogenous and transgene sequences have a single nucleotide difference across the entire gene. Our work provides insight into the applications of a common transcript estimation tool and a direct counting method for estimating transcript abundance in gene therapy.

1663T

Miscall estimation for pathogenic variants in genomic direct-to-consumer microarray data. G. Lennon, M. Cariaso. SNPedia, Potomac, MD.

Direct-to-Consumer (DTC) genomic products based on microarray technology are relatively inexpensive and have been used by millions of consumers. Many of these microarrays report on rare variants that have significant clinical associations based on the scientific and medical literature. The false positive rate for reported pathogenic variants in DTC products is important to establish. As a significant provider of independent, third-party literature retrieval reports based on uploaded DTC data, Promethease is in a unique position to compare variant frequencies in such data to expected frequencies observed in large-scale sequencing projects such as gnomAD and TOPMED. Here, we report our findings based on raw genotype data from microarray DTC products generated by companies such as 23andMe, Ancestry, and MyHeritage. In addition, thousands of individuals have data for their own genomes from two or more DTC companies, allowing us to calculate reproducibility between companies. Overall, we find high reproducibility and very good correlation between microarray data frequencies and sequencing frequencies. For genes with high medical interest such as BRCA1 and BRCA2, we are able to estimate the number of variants likely to be miscalled. Based on the ability to estimate miscall likelihood, we are developing a new feature in Promethease reports allowing users to systematically identify and filter specific genotypes likely to be incorrect. This feature is based on statistical analysis of company- and platform-specific frequency data. We believe this feature will empower DTC users to more confidently assess their data and the relevant literature.

1664F

MiniScrub: De novo long read scrubbing using approximate alignment and deep learning. N. LaPierre¹, R. Egan^{2,3}, W. Wang¹, Z. Wang^{2,3,4}. 1) University of California, Los Angeles, Los Angeles, CA; 2) Department of Energy Joint Genome Institute, Walnut Creek, CA; 3) Lawrence Berkeley National Laboratory, Berkeley, CA; 4) University of California, Merced, Merced, CA.

Long read sequencing technologies such as Oxford Nanopore can greatly decrease the complexity of de novo genome assembly and large structural variation identification. Currently Nanopore reads have high error rates, and these errors, if not corrected, may result in poor quality assemblies. Correcting these errors dramatically improves assembly quality, but requires prohibitive experimental or computing cost. Here we developed a novel Convolutional Neural Network (CNN)-based method, called MiniScrub, for de novo identification and subsequent "scrubbing" (removal) of low-quality Nanopore read segments. MiniScrub first generates read-to-read alignments by MiniMap, then encodes the alignments into images, and finally builds CNN models to predict low quality segments that could be scrubbed based on a user defined quality cutoff. Applying MiniScrub to in-house Nanopore control datasets under several different parameters, we show that it robustly improves overall read quality, which in turn leads to dramatic improvements in the speed and/or accuracy of de novo genome assembly. When applying MiniScrub as a preprocessing step before assembly with MECAT, the percentage of genome assembled increases from 87.92% to 99.71%, the NGA50 increases from 315,975 to 438,077, and the number of misassembled contigs decreases from 18 to 14. As a preprocessing step for Canu, MiniScrub results in similar percentage of genome assembled but enables Canu to run 9 times faster by eliminating excess poor-quality data. We propose MiniScrub as an essential tool for preprocessing Nanopore long reads, and its design makes it generalizable across a variety of datasets and similar sequencing technologies.

1665W

A massive collection of Japanese full-length class-I HLA sequences enhanced the accuracy of HLA genotyping from WGS data. *Y. Wang¹, T. Mimori¹, S. Khor², Y. Hitomi², O. Gervais¹, K. Misawa¹, Y. Kawai², K. Tokunaga², M. Nagasaki¹.* 1) Division of Biomedical Information Analysis, Department of Integrative Genomics, Tohoku Medical Megabank Organization, Tohoku University; 2) Department of Human Genetics, Graduate School of Medicine, The University of Tokyo.

The human leukocyte antigen (HLA) system, which is the most variable gene region in the entire genome, encodes the major histocompatibility complex (MHC) proteins in humans on chromosome 6 and has been reported to be associated with numerous immune-mediated diseases. HLA calling from whole genome sequencing data (WGS) remains to be challenging due to the incompleteness of population specific references. We have previously described the construction of Japanese reference sequences of HLA genes (ToMMo HLA sequences). This study aim to evaluate the accuracy of HLA calling from WGS when the in-house ToMMo HLA sequences were used in conjunction with the IPD-IMGT reference sequences to estimate the most likely HLA genotypes. To assess the usefulness of population-specific HLA references in HLA type estimation, we compared the performance of different sets of references (with and without ToMMo HLA sequences) across HLA-A, HLA-B and HLA-C alleles by using HLA-VBSeq, a flexible software which allows the use of arbitrary reference sequences. In addition, we compared HLA-VBSeq with another widely-used HLA typing software, HLA*PRG:LA. We evaluated their performances by using 418 Japanese samples (Tokyo Healthy Controls, THC) as testing data, in which HLA genotypes were determined with Luminex technology. All of the tests were implemented based on 4-digit resolution. As a result, the performance of HLA calling in the Japanese population significantly increased when we included ToMMo HLA sequences as reference sequences. In conclusion, we suggest that including the sequences obtained from regional population samples is an effective way to improve the accuracy of HLA calling from WGS dataset. Given that this study was performed based on 4-digit resolution, in the future we intend to assess the accuracy of HLA-VBSeq with ToMMo HLA sequences at the 6-digit and 8-digit level resolutions.

1666T

Linkage-disequilibrium regularized support vector machines for genome-wide association studies. *M. Sudarshan, L. Subramanian.* Courant Institute of Mathematical Sciences, NYU, New York, NY.

It is of great interest to perform whole genome analyses (WGA) in order to determine the causal single nucleotide polymorphisms (SNPs) for a particular outcome without manually subsampling the large genome. However, in many experiments, researchers focus only on a small portion of the genome [He and Lin, 2011], [Wu et al., 2010], and [Fan and Lv, 2008]. This may be done to mitigate the effect of spurious correlations between certain SNPs and the outcome. We hypothesize that by considering block structures in the Linkage Disequilibrium (LD) [Slatkin, 2008] matrix of the data, we can return a significantly smaller set of significant SNPs without a large drop in prediction accuracy. LD is a property observed in genomics where adjacent SNPs take on highly correlated values. Our main contribution is a two-step procedure in which we first determine blocks in the LD matrix using a constrained spectral method, then regularize an SVM model based on these blocks. We apply a separate L2 norm for each block of SNPs and an overall L1 norm to enforce a small number of SNPs being returned as relevant. We observe that this model selects blocks of SNPs with high pairwise LD that are significantly predictive of the phenotype. Furthermore, the model selects only the SNPs within each block that are the most significant in terms of predicting the outcome. We also observe that despite a slightly larger set of SNPs being returned in the block-regularized case, the percentage of SNPs that are significant are better than or equal to the L1-regularized models. The use of a max-margin classifier over logistic regression improves classification accuracy by around 6% while returning a set of highly significant SNPs.

1667F

Fine mapping chromatin contacts in capture Hi-C data. C. Eijssbouts^{1,2}, O. Burren¹, P. Newcombe², C. Wallace^{1,2}. 1) Department of Medicine, University of Cambridge; 2) MRC Biostatistics Unit, University of Cambridge.

Capture Hi-C (CHI-C) uses high-throughput sequencing to map physical interactions between chromatin regions (HindIII digestion fragments) in cell nuclei. The technique is suitable for the discovery of regulatory interactions at high resolution, which is achieved by targeting only interactions involving defined regions of interest (baits). However, in a typical CHI-C analysis, the evidence for each of a bait's putative interactions is considered independently, while patterns in the evidence across neighboring regions are ignored. This tends to produce long runs of regions which appear to all make contact with the same bait. Based on peaks within the signal across these, we hypothesised that long runs of apparently interacting regions could result from a small subset of direct interactions. We propose a new method, based on a Bayesian sparse variable selection approach, which attempts to fine map these direct interactions. Our model is conceptually novel, exploiting the spatial pattern of read counts in CHI-C data, and in application to real experimental data prioritised interactions with biological properties that would be expected of true contacts. For bait fragments corresponding to gene promoters, we identify contact fragments with active chromatin that overlap those found in previously defined enhancer-target networks; conversely, for intergenic bait fragments, we identify contact fragments corresponding to promoters for genes expressed in that cell type. We show that long runs of apparently co-contacting fragments can typically be explained using a subset of direct interactions consisting of <10% of the number in the full run. Our results appear largely complementary to those from a per-fragment analytical approach, and provide an additional level of interpretation for CHI-C datasets that helps to map interactions at increased resolution.

1668W

Pathoscore: A tool for unbiased evaluation of variant pathogenicity metrics. J.M. Havrilla¹, B.S. Pedersen^{1,2}, R.M. Layer^{1,2}, E.J. Hernandez^{1,2}, M. Yandell^{1,2}, A.R. Quinlan^{1,2,3}. 1) Department of Human Genetics, University of Utah, Salt Lake City, UT; 2) USTAR Center for Genetic Discovery, University of Utah, Salt Lake City, UT; 3) Department of Biomedical Informatics, University of Utah, Salt Lake City, UT.

When sequencing a rare disease cohort, much of the variation will be either extremely sparse in the population, or entirely de novo. Deciding upon the best metric to evaluate the pathogenicity of each newly discovered variant can be onerous. Many variant pathogenicity predictors exist, yet determining which is most appropriate to use for a particular phenotype and its corresponding set of genes is difficult. One major complication in fairly evaluating metrics lies with ClinVar. ClinVar has numerous limitations: it changes version and its variant set constantly, the criteria for establishing pathogenicity is not consistent, benign variants are primarily common variants from population-scale datasets, and many variants are incorrectly annotated. Nearly all variant pathogenicity predictors have compared their metrics to others on ClinVar alone, and each one uses a different version or subset of the variants, with virtually no consistency between independent evaluations based on ClinVar. To address these difficulties and to aid in the evaluation of new and existing methods, we have created pathoscore, which acts both as a tool and a collection of curated truth sets and metrics. The truth sets are filtered to exclude supposed pathogenic variants that are common in the gnomAD population database. The pathoscore tool allows quickly evaluating existing methods via a simple command-line tool that utilizes the truth-sets included in the repository. Pathoscore creates an interactive HTML page for evaluating the performance of variant pathogenicity predictors on different truth sets. The HTML contains ROC curves, value distribution plots for benign and pathogenic variants for each metric, Youden's J statistic curves, precision recall curves, the number of variants each metric was able to score, and a clinical utility score. There is a detailed README and a wiki to facilitate use of this tool and help interpret the output. Our goal is not only to provide bioinformaticians with a fair and consistent way to evaluate their new tool against existing approaches, but also to provide clinicians and diagnosticians with an easy, comprehensive approach to identifying the best metric for their patients' datasets.

1669T

IGV-Web: An application for viewing and sharing genomic datasets.

J.T. Robinson¹, H. Thorvaldsdóttir², D. Turner¹, J.P. Mesirov^{1,2,3}, 1) University of California San Diego, La Jolla, CA; 2) Broad Institute of MIT and Harvard, Cambridge, MA; 3) Moores Cancer Center, University of California San Diego, La Jolla, CA.

Advances in genomic technologies are revolutionizing our knowledge and understanding of human disease. They enable the generation of diverse types of data to detect genetic variation and analyze its impact. While much of the analysis can be automated, human interpretation and judgment supported by rapid and intuitive visualization are important for gaining insight, increasing the confidence in computational results, and characterizing complex events. First released in 2008, the *Integrative Genomics Viewer (IGV)* is a high-performance desktop tool for interactive visual exploration of diverse genomic data. IGV was one of the first tools to provide next-generation sequencing (NGS) data visualization, and it currently provides a rich set of tools for inspection, validation, and interpretation of genomic datasets. The IGV supports real-time interaction at all scales of genome resolution, from whole genome to base pairs, even for very large NGS datasets. In 2016, we introduced *igv.js*, a JavaScript implementation of IGV, designed to be easily embeddable in web portals, cloud applications, and investigators' web pages. It supports many of the key features of the desktop application, including visualization of NGS alignments, germ-line and somatic variants, copy-number data, annotations, and eQTL data. This version has been widely adopted in web portals, including the cBioPortal for Cancer Genomics, Genotype-Tissue Expression (GTEx) Portal, Exome Aggregation Consortium (ExAC) Browser, and others. Here we present *IGV-Web* (<https://www.igv.org/app>), a new web-based resource based on *igv.js*. IGV-Web enables researchers to view genomic datasets using the familiar IGV user interface without the need to install any software. Importantly, it allows users to visualize and share data stored on web servers, as well as popular cloud providers such as Dropbox, Google Drive, Google Cloud Storage, and Amazon Web Services. Data is protected using the cloud provider's authorization scheme. IGV-Web encodes the exact state of the visualization in a shareable URL, making sharing a visualization of a genomic dataset as easy as sharing a YouTube video. The IGV-Web application, IGV desktop application, and *igv.js* are all open source and freely available at www.igv.org and github.com/igvteam.

1670F

What's new with NCBI tools for genome visualization and analysis.

V.A. Schneider, A. Kuznetsov, T.D. Murphy, F. Thibaud-Nissen. Information Engineering Branch, NIH/NCBI, Bethesda, MD.

To keep pace with the still-increasing number of newly sequenced genomes and offer users the sequence-level details that are now an integral part of most annotation data and browser-based analyses, NCBI introduced the Genome Data Viewer (GDV), a new genome browser, in 2016. Since then, we have continued to add functionality to GDV, making possible the retirement of the aging Map Viewer browser in May 2018. We will present recent GDV updates, which include an integration with BLAST that offers tools for graphical analysis of alignments, new track offerings that show details of assembly-assembly alignments and evolutionary conservation, and added support for the display of custom tracks representing externally hosted data, either from user-provided stand-alone remote files or those in the UCSC Track Hub format. GDV's graphical display uses the NCBI Sequence Viewer, an application that can also be integrated into web pages external to NCBI using its embedding API. Sequence Viewer offers highly customizable views of individual sequences that can include both NCBI-provided annotation tracks and custom data, and publication quality PDFs. We will also demonstrate examples of its recently added features and use. We will additionally highlight recent updates to web-based NCBI genome analysis tools, such as the Coordinate Remapping Service, which now offers improved handling for variant data in VCF format and an alpha release of a stand-alone software package, and the Genome Decoration Page (GDP) that can be used to generate publication quality images of annotated ideograms. These, and other tools, can be easily accessed from the NCBI human resources page (<https://www.ncbi.nlm.nih.gov/genome/guide/human/>). Together, these tools provide users with a comprehensive suite of resources for genomic visualization and analysis. This work was carried out by staff of the National Library of Medicine (NLM), National Institutes of Health, with support from NLM.

1671W

Using heterogeneous networks to systematically record, model and query GWAS follow-up experiments. *J. Barrett, J. Floyd, P. Seth, M. Suci.* Genomics plc, Cambridge, United Kingdom.

To date, more than 3300 genome-wide association studies (GWAS) have identified approximately 62,000 unique SNP-trait associations. However understanding the causal variants, genes and mechanisms underlying these associations has proven difficult. Signals of association frequently span large regions encompassing many genes, and statistical fine-mapping approaches and variant annotation often fail to identify any one variant conclusively. There has therefore been a huge amount of functional follow-up experimentation done to try to resolve these questions. These studies carry out a diverse range of experiments and typically report results for individual loci in an unstructured way in the text of publications. Researchers may be familiar with results in their disease of interest, but making systematic comparisons between disease areas is difficult. Here we present a tool to record, model and enable systematic queries of a wide variety of data types used to convincingly identify causal variants, genes and mechanisms. We represent these data as a heterogeneous network, which have been previously used to encode relationships among a wide variety of biological data types. Here we focus on creating much more detailed paths between traits and genes, which previously have been captured by generic "is associated with" edges that can vary in how confident the underlying data are. In addition to trait and gene nodes we have added variant, transcript, region (e.g. enhancer), and cell type. Edges represent experimentally validated links, such as genetic association, eQTL co-localisation, tissue-specific expression, model organism knockouts, or CRISPR perturbations in cell models. We have built upon existing graph frameworks to benefit from methods and software tools to create and query meta-nodes and meta-paths connecting genes and traits. We have implemented a user-friendly tool to enter data (e.g. while reading a paper) to help expand the database. This framework enables comprehensive analysis of a wide variety of commonly employed GWAS follow-up experiments. We will show how paths with the strongest evidence can be used as "gold standard" loci to evaluate automated algorithms for discovering new causal genes, which has historically been challenged by the absence of robust and computationally usable truth sets.

1672T

GlobAI Distribution of GEnetic Traits (GADGET): Exploring polygenic trait scores. *A.T. Chande^{1,2,3}, L. Wang^{1,3}, L. Rishishwar^{1,2,3}, A.B. Conley^{1,3}, E.T. Norris^{1,2,3}, A. Valderrama-Aguirre^{3,4}, I.K. Jordan^{1,2,3}.* 1) School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA, 30332; 2) IHRC-Georgia Tech Applied Bioinformatics Laboratory, Atlanta, GA 30332, USA; 3) PanAmerican Bioinformatics Institute, Cali, Valle del Cauca, Colombia; 4) Biomedical Research Institute, Faculty of Health, Universidad Libre-Seccional Cali. Cali, Valle del Cauca, Colombia.

Genome-wide association studies (GWAS) have the power to reveal the underlying genome-phenotype associations within complex traits. Over the last decade, thousands of such associations have been discovered for hundreds of traits. Exploration and interrogation of these data in global populations has the potential to aid in the understanding of global and regional risk factors. To address this need, we have developed the GlobAI Distribution of GEnetic Traits web server (GADGET <http://gadget.biosci.gatech.edu>). The GADGET web server provides users with a dynamic visual platform for exploring the relationship between worldwide genetic diversity and the genetic architecture underlying numerous human phenotypes. GADGET combines trait-associated single nucleotide polymorphisms discovered by GWAS with genotype data from the 1000 Genomes Projects and ChocoGEN to compute polygenic trait scores (PTS) for 800+ complex trait for 27 global populations. Population-specific PTS are visualized for each of the 27 global population and 5 continental super-populations. GADGET users can also analyze custom trait-associated SNP lists and visualize their distribution. GADGET was used to explore and describe the global distribution of risk for Post-Traumatic Stress Disorder and Type 2 Diabetes. We also used GADGET and its underlying technology to interrogate genetic difference among African and European ancestry individuals in Colombia.

1673F

An exploration of the clinical phenome and the application of natural language processing for rapid and deep phenotyping of acutely ill children with genetic diseases. *M.M. Clark¹, S. Batalov¹, M. Janes¹, T. Defay², S. Gilmer¹, C. Yacoubian³, N. Veeraraghavan¹, S.F. Kingsmore¹.* 1) Rady Children's Institute for Genomic Medicine, San Diego, CA, USA; 2) Alexion Pharmaceuticals, Inc., New Haven, CT, USA; 3) Clinithink Ltd., London, UK.

Intro: Rapid whole-genome sequencing (rWGS) is now gaining traction for the diagnosis of neonates affected with rare genetic disorders in intensive care units. The accuracy and completeness of patient phenotypic information has a direct impact on the ability to make a timely molecular diagnosis. However, there are three issues that remain stumbling blocks: (1) textbook descriptions are incomplete for rare genetic diseases, (2) existing descriptors lack granularity, and (3) phenotyping from electronic health records (EHRs) is time-consuming, requires expert knowledge, and has a risk of bias. In a retrospective analysis, we examine how closely phenotypes derived with natural language processing (NLP) match those selected by expert analysts and physicians and how both compare to the current understanding of the diagnosed disease. **Methods:** For 76 acutely ill children diagnosed with a genetic disease, an expert reviewed the EHR and manually selected HPO phenotypes determined to be representative of the child's condition. NLP software (CLiX ENRICH, Clinithink) was adapted to extract HPO phenotypes from non-structured EHR fields without human involvement. For each patient, an automated workflow was triggered which imported clinical documents from the Epic Chronicles Database to CLiX ENRICH and automatically initiated a job resulting in a set of HPO terms. The OMIM disease descriptions were used to assess the accuracy of the existing disease-phenotype associations. These three sets of phenotypes were evaluated based on precision, recall, count, and information content (IC), a measure of term specificity. **Results:** We find that the application of NLP to ascertain a deep phenotype has 74% precision and 93% recall. In ~20 seconds, the optimized NLP software identified an average of 93 phenotypes per patient, a 20-fold deeper phenome than by expert manual curation. There was surprisingly little overlap between the observed and expected phenomes; only 3.2% of the NLP phenome matched the OMIM descriptions of those genetic diseases. Interestingly, the phenotypic overlap of OMIM and NLP had significantly higher IC than the overlap of OMIM and the manual extraction ($p < .0001$). This exploration provides quantitative evidence that most acutely ill children with genetic diseases have presentations that imprecisely and noisily match textbook descriptions of disease, which has critical implications for the genetic diagnosis of affected children based on clinical phenotypes.

1674W

Utilizing spike-and-slab Lasso models for epistasis analysis. *J. Wen, J. Chen, X. Shi.* University of North Carolina at Charlotte 9201 University City Blvd., Charlotte, NC 28223.

Epistasis among genetic variants or genes is an important genetic mechanism in complex traits including common diseases. Many studies show that epistasis helps elucidate the genetic foundation for modeling human traits and evolution. However, it is statistically and computationally challenging to identify epistasis in large-scale and high-dimensional genomic data. The challenge lies in many intertwined aspects including (1) how to avoid overfitting induced by high-dimensional data where the number of genetic variants is typically much larger than the number of samples under investigation; (2) how to model the high collinearity or correlations in genomic data; and (3) how to reduce the search space for epistasis analysis to ensure that the computing is feasible for large scale genomic data. Recently, shrinkage based variable selection methods have showed desirable property for modeling this type of genomic data. In this study, we apply a new type of shrinkage variable selection models, namely spike-and-slab Lasso general linear models, to identify epistasis among genetic variants. We develop an iterative workflow with an interactive web portal to support data pre-processing, model training and optimization, and result presentation and visualization for epistasis analysis using these new models.

1675T

Predicting genetic communities using frequently shared identity by decent segments. L. Doroud¹, S. Hateley¹, G. Yalamanchili², H. Guturu¹, K. Noto¹, E. Hong¹, K. Chahine², C. Ball¹. 1) AncestryDNA, San Francisco, CA; 2) AncestryDNA, Lehi, UT.

Direct-to-consumer (DTC) genomics tests provide a unique experience for individuals interested in their family history and genetic matches. AncestryDNA estimates customers' distant and recent ancestry. The recent ancestry evaluation is powered by the Genetic Communities™ algorithm, which constructs a large Identity-By-Descent (IBD) network and clusters individuals if they share a significant amount of genome-wide IBD [1][2]. With the remarkable popularity in DTC genomics and continuous expansion of our genotype database, the IBD connections among individuals grow quadratically, leading to a "hair-ball" effect on the IBD network. To address the computational challenges inherent with a genome-wide IBD network, we propose a novel approach that investigates the significance of the IBD segments within the communities. In this study, we evaluate the impact of this method on the feasibility and accuracy of community assignments. [1] <https://www.ncbi.nlm.nih.gov/pubmed/28169989>[2] <https://www.ancestry.com/cs/dna-help/communities/whitepaper>.

1676F

A Hadoop data lake enables genocentric mining of genetic variation associated with Mendelian disease. A.W. Hansen^{1,2}, L. Wang¹, B.G. Palazzo¹, P. Panchel¹, J.R. Easton-Marks¹, Z. Coban-Akdemir², J.A. Rosenfeld², C. Eng^{2,3}, Y. Yang^{2,3}, J. Posey², J.R. Lupski^{1,2,4,5}, M. Murugan¹, R.A. Gibbs^{1,2}. 1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 3) Baylor Genetics Laboratories, Houston, TX; 4) Department of Pediatrics, Baylor College of Medicine, Houston, TX; 5) Texas Children's Hospital, Houston, TX.

Traditional study design for disease-gene association is 'phenocentric' and assumes that enriching a cohort for phenotypic homogeneity will also enrich for genetic homogeneity. Here we have pursued a complementary genocentric paradigm to accelerate gene discovery in Mendelian disease cohorts. We employed a Hadoop open-source platform to construct a data lake, aggregating whole exome sequencing (WES) data from more than 17,500 samples—approximately ¾ of samples are probands, ¼ are unaffected family members or other healthy exomes. The only phenotypic commonality among probands is an indication for WES due to suspected Mendelian disease. Additionally, we have loaded variant-level and gene-level annotations from a variety of sources, enabling rapid, flexible annotation through Hive or Impala joins. We flagged genes harboring recurrent ultra-rare variants (MAF < 1/10k) across at least 5 samples, placing the genes into bins according to the value of one additional filtering criteria (i.e., ExAC missense intolerance z-score between 4.0 and 4.5, etc.), creating a series of gene lists across a range of possible scores for a given parameter. The fraction of Mendelian disease genes (from OMIM) was then compared across the series of gene lists corresponding to these bins. We hypothesized that when creating these binned gene lists based on filter parameter value, outlier lists with lower-than-expected disease gene enrichment would be 'corrected' as more disease genes are reported. Accordingly, we prioritize gene lists for discovery efforts as those with lower-than-expected disease gene enrichment. Comparing gene list OMIM disease ratio in 2016 vs 2018 database versions validates this hypothesis, as gene lists with lower-than-expected OMIM disease gene enrichment in 2016 have the highest increase in reported OMIM disease-gene associations from 2016 to 2018. Our approach yielded more than 100 candidate novel dominant disease genes, which we are prioritizing for both *de novo* mutation screening and functional study. For example, we identified a Sanger-validated *de novo* variant in *POLR2A* and are collaborating to study 10 patient variants in multiple model organisms. Our approach is also valuable to the study of established disease genes, such as *DDX3X*, where we have identified multiple Sanger-validated *de novo* variants in previously undiagnosed patients.

1677W

ClinGen's Gene Curation Interface (GCI) facilitates gold-standard consistent evaluation of the clinical relevance of genes. *M.W. Wright¹, S.S. Dwight¹, J. Zhen¹, C. Thomas¹, B. Wulf¹, S. Prabhu¹, H. Wand¹, D. Azzariti², M. DiStefano³, J. Goldstein³, S. Harrison³, J. McGlaughon³, E. Rooney Riggs⁴, N.T. Strande⁵, C. Thaxton², L. Babb², C. Bizon⁵, S. Goehringer⁴, T. Nelson⁴, J. Zou¹, G. Bejerano¹, J.S. Berg³, C.L. Martin⁴, S.E. Plon⁶, H.L. Rehm^{2,7,8}, C.D. Bustamante¹, Clinical Genome Resource (ClinGen).* 1) Stanford University, School of Medicine, Stanford, CA; 2) Laboratory for Molecular Medicine, Partners Personalized Medicine, Cambridge, MA; 3) University of North Carolina at Chapel Hill, NC; 4) Autism & Developmental Medicine Institute at Geisinger, Danville, PA; 5) Renaissance Computing Institute, Chapel Hill, NC; 6) Baylor College of Medicine, Houston, TX; 7) Brigham and Women's Hospital, Boston, MA; 8) Broad Institute, Cambridge MA.

With a large increase in the number of reported gene-disease relationships in recent years, there is an urgent need to critically evaluate which associations are most clinically valid. Defining the clinical relevance of genes is one of the primary objectives of the Clinical Genome Resource (ClinGen). The ClinGen gene curation process combines an appraisal of genetic and experimental data in the scientific literature with expert review to classify gene-disease pairs into 1 of 6 categories according to ClinGen's Gene-Disease Clinical Validity Classification framework. ClinGen's Gene Curation Interface (GCI) tool (curation.clinicalgenome.org) was specifically designed to facilitate this process, enabling curators to systematically assess the strength of gene-disease associations. The GCI guides curators through the curation process, enabling them to access and evaluate evidence for a gene's role in disease in a consistent and efficient manner. All evidence is accessible to any registered user, allowing all users to view and evaluate the same set of information (Case Level, Case-Control and Experimental evidence). Members of ClinGen expert panels (EPs) can work collectively on a Classification to edit and score/evaluate the same pieces of evidence. An automated table dynamically tallies the score and provides a calculated Classification. A detailed Evidence Summary is also available and is used by ClinGen's EPs as an integral part of their expert review process. Expert Panels can now publish their final approved Classifications to the public ClinGen website (clinicalgenome.org), via ClinGen's Data Exchange (a data repository that harmonizes curated ClinGen data via a common data model). Controlled vocabularies and ontologies are utilized throughout the GCI in order to promote the capture of discrete evidence that will allow downstream connections between information on the ClinGen website and to promote consistency. We also briefly describe our plans to assist curators in their capture of relevant data from publications via large-scale identification and "tagging" of literature with standardized terminologies. **A test site (curation-test.clinicalgenome.org) with a "demo login" facility is available for anyone who wants to explore the functionality of the GCI. We provide extensive help documentation on how to use the GCI, and there is a detailed SOP on the ClinGen website that describes how to utilize their framework.**

1678T

Xrare: A machine learning method jointly modeling phenotypes and ACMG guidelines improves rare disorder diagnosis. *X. Ma^{1,2}, K. Zhao³, Q. Li², C.D. Bustamante^{3,4}, W.H. Wong¹.* 1) Department of Statistics, Stanford University, 390 Serra Mall, Stanford, CA 94305; 2) GenomCan Inc., 119 University Ave, Palo Alto, CA 94301; 3) Department of Biomedical Data Science, Stanford University, 365 Lasuen Street, Littlefield Center, Room 303, Stanford, CA 94305; 4) Department of Genetics, Stanford University, 300 Pasteur Dr., Stanford, CA 94305.

Despite the successful progress of next generation sequencing technologies brought in diagnosing the genetic cause of rare Mendelian diseases, the current diagnostic rate is still far from satisfactory with ranges from ~28% in WES to 57% in the most comprehensive family trio WGS. While standards and guidelines for the interpretation of sequence variants released by the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) have greatly facilitated the variant pathogenicity interpretation process, subjectivity and inconsistency are still widespread across clinical molecular genetic laboratories, with the phenotype to genotype supporting evidence being the most inconsistent among the laboratories. Even though existing computational phenotype-driven methods (e.g. Exomiser, Phen-Gen) have eased the inconsistency and gained wide applications in clinical diagnoses, the heterogeneity, imprecision and noise in disease phenotype descriptions in real clinical settings posed big challenges. To combine the strength from both approaches, we developed a machine learning method for disease causative variant prioritization. We developed a novel phenotype similarity scoring method named Emission-Reception Information Content (ERIC) highly tolerant of noise and imprecision in the clinical phenotypes and incorporated ACMG domain expert knowledge in feature extraction and learning. Extensive simulations and two real clinical datasets demonstrated that Xrare outperforms existing alternative methods (e.g. CADD, M-CAP, REVEL, Exomiser, Phen-Gen etc.) by 10-40% at various genetic diagnosis scenarios.

1679F

CLUMPAK 2.0: Improved alignment of replicate cluster analyses. *N.M. Kopelman¹, M. Jakobsson², N.A. Rosenberg³, I. Mayrose⁴.* 1) Holon Institute of Technology, Givatayim, Israel; 2) Uppsala University, Uppsala, Sweden; 3) Stanford University, Stanford, CA; 4) Tel Aviv University, Tel Aviv, Israel.

Many model-based clustering methods have been developed for the identification of population structure. These methods take as input multi-locus genotypes of individuals, and either assign these individuals into a few presumed source populations, or – more often – estimate coefficients of membership of each individual in the various clusters, summing to 1 across clusters. Typically, model-based algorithms, among which the most popular is STRUCTURE, involve stochastic steps. Assuming N individuals and K ancestral populations, where K is a parameter that is predefined, the output of a single run of an algorithm is a matrix of membership coefficients of size $N \times K$, termed a "Q matrix". Due to the stochastic nature of the algorithms, a typical analysis consists of multiple runs for the same value of K . Moreover, as the "real" number of ancestral populations is generally unknown, a range of K values is investigated. Thus, many Q-matrices are produced, and a comparison of these Q-matrices, as well as a summary of the results, are required. We have previously developed CLUMPP and CLUMPAK in order to aid users in automating the analysis of the results of genotype clustering programs such as STRUCTURE. Both programs incorporate various algorithms for the alignment of multiple replicates, i.e. multiple Q-matrices, produced for the same value of K . In addition, CLUMPAK separates runs belonging to distinct clustering solutions, and aligns results across different values of K , enabling publication-quality visualization of results with different K values. In this work, we have augmented the three algorithms for the alignment of replicate Q-matrices with a new algorithm. In particular, we have added the classical Hungarian algorithm, treating the alignment problem as an assignment problem. Unlike the greedy algorithms currently utilized by CLUMPP and CLUMPAK, the order in which replicates are handled does not affect the Hungarian algorithm, thus saving the need to examine large numbers of different input orders of replicates. We compare the results and running times for the Hungarian and greedy algorithms using various datasets and species. CLUMPAK is freely available for use as a web-server, allowing users to easily analyze tens or hundreds of results produced by STRUCTURE-like programs.

1680W

A novel non-linear dimension reduction approach to infer population structure for low-coverage sequencing data. *J. Zhou¹, M. Zhang¹, H. Zhou², J. Watkins¹.* 1) University of Arizona, Tucson, AZ; 2) University of California, Los Angeles.

With rapid advances in DNA sequencing technologies, different sequencing designs and platforms appear and provide options balancing between accuracy and cost. As a cost-effective choice, low to medium depth sequencing allows researchers afford to sequence more samples. But this sequencing relies on linkage-disequilibrium (LD)-based methods and has lower accuracy for genotype calling, especially rare-variant identification. Downstream association analyses are advised to incorporate sequencing uncertainties in order to increase the statistical power. As a fundamental dimension reduction technique, principal component analysis (PCA) is widely adopted in genetic studies to infer population structure and to adjust the bias caused by population differences in the association testing procedures. However, PCA is not designed to reveal non-linear patterns that may arise in genetics data, for example, in low-depth genomic sequencing data. We propose an advancement of a data-adaptive approach, *MCPCA_PopGen*. Our method uses genotype likelihoods rather than genotypes and, thus, takes into account the uncertainty of raw sequencing reads to discover the non-linear patterns. *MCPCA_PopGen* is defined via an optimization that infers transformations of genotype likelihoods and maximizes Ky Fan norm of the transformed covariance matrix. We propose an efficient block coordinate descent algorithm to determine transformation. Moreover, we propose a distributed algorithm to compute an approximation of Maximal Correlated Principal Components for dense sequencing data by chromosomal partitions. In the simulation, we reveal two major properties of our method for low-coverage sequencing data. For a locus with a low minor allele frequency (MAF), the transformation emphasizes the uncertainty in calling between heterozygous and the major homozygous loci. On the other hand, the transformation is more linear when both variants are common. We apply *MCPCA_PopGen* to a human sequencing sample from two indigenous Siberian populations, the Nganasan (nomadic hunters, NGA) from the Taymyr Peninsula in the Arctic Ocean, and the Yakut (herders, YAK) of North-Central Siberia with 5x coverage on average. Our method can separate two groups with much higher confidence in comparison with many other existing methods, including PCA.

1681T

A data broker tool to coordinate availability queries and meta-analysis in a secure, compliant environment. *E. Sukharevsky, B. Jackson, F. Syed, N. Hurskainen, T. Kanninen, A. Ahonen-Bishop.* BC Platforms, Boston, MA.

It is a contradictory time for the scientific community: while it is acknowledged that there is a need for genomic data sharing to deliver on the promise of precision medicine, optimize outcomes and reduce costs, data security regulations are becoming stricter and can limit progress. Communicating with tens or hundreds of collections individually on a global scale has been stated as one of the biggest bottlenecks in seeking data and samples for pharmaceutical research. Academic research groups and industry alike demand an ecosystem that facilitates the sharing of data and information in a safe and secure environment. This opens the door for the role of a data broker solution that can guarantee a confidential environment for biobanks to share their data with pharmaceutical researchers who seek suitable patient cohorts within validated data clouds on a global scale. Such a global ecosystem, presented here, does exist so that researchers can efficiently search for specific cohorts of patients for population research and rare tumors. With this system, biobanks also benefit by getting access to researchers, sustaining their academic relevance and financial viability. Additionally, the ability to perform controlled meta-analysis on genomic data provides researchers with necessary tools to run feasibility studies, and to investigate the safety of specific pathway targets utilizing naturally occurring mutations in the cohorts. As the platform is highly flexible in its data model and access control, its utility in the most complex data environments is unique. .

1682F

Undiagnosed Diseases Program Database (UDPdb) of genomic data: Development and utility. *D.L. Ross, N.J. Balanda, N. Ten, B.N. Pusey, C. Lau, W.A. Gahl, D.R. Adams.* UDP, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD.

The NIH Undiagnosed Diseases Program (UDP), now part of the Undiagnosed Diseases Network (UDN), is entering its 10th year of seeking diagnoses for individuals with illnesses that remain undiagnosed despite extensive medical evaluation. The UDP's diagnostic rate is approximately 25%. Extensive phenotyping as well as genome sequencing have been critical in making these diagnoses. Bioinformatic methods have drastically advanced over the last 10 years creating a unique opportunity to reanalyze patient's genomic data. Further, the accumulation of genomic data of 500+ families with undiagnosed diseases presents another unique opportunity to discover rare diseases shared by two or more patients within the cohort. The purpose of this work is to revisit past unsolved cases en bloc in a systematic and organized fashion with the conjecture that more patients can be diagnosed by examining the cohort altogether and performing reanalysis using updated methods, tools and annotation databases. The heterogenous nature of the source data for patients creates a challenge for reanalysis and cohort-wide analyses. To address this challenge, the approximately 1700 exomes and 300 genomes were uniformly aligned and joint-genotyped using the DRAGEN V2+ pipeline (Edico Genome, San Diego, CA), thereby creating a cohort-wide variant call format (VCF) file. Annotation of variants was performed using Variant Effect Predictor (VEP) and Vcfanno (Quinlan Lab, University of Utah), incorporating a broad variety of the most current annotation sources. Finally, the variant and annotation data were loaded into the GEMINI (Genome MINing, Quinlan Lab, University of Utah) platform to enable flexible and structured data queries based on family variant segregation and annotation information. The UDPdb will provide a model for studying reanalysis approaches in our extensively-phenotyped undiagnosed-disease cohort. We have built an approach to systematically reanalyze the entire cohort using new and experimental annotation sources. The speed of the DRAGEN platform combined with the analytic flexibility of the GEMINI system allows us to update the dataset on a frequent basis. Creating a standardized and updatable sequence dataset for our cohort of undiagnosed-disease study participants has the potential to allow for rapid adaptation to new analytic approaches, providing a foundation for ongoing studies of data reanalysis.

1683W

Seekmer: A fast and accurate tool for quantitative transcriptomic analysis at low sequencing depth.

H. Zhang¹, Y. Wang², E. Aziz³, S. Wu¹, G. Omenn^{1,2,3}, R. Mills^{1,2}, Y. Guan^{1,3,4}. 1) Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI; 2) Department of Human Genetics, University of Michigan, Ann Arbor, MI; 3) Department of Internal Medicine, University of Michigan, Ann Arbor, MI; 4) Department of Electronic Engineering and Computer Science, University of Michigan, Ann Arbor, MI.

Low sequencing depth presents a long-term challenge for quantitative transcriptomic analysis, especially in novel single-cell RNA sequencing studies. We present Seekmer, a fast and accurate quantification tool that seamlessly combines the speed of alignment-free and accuracy of alignment-based algorithms without their corresponding shortcomings. Seekmer incorporates fast alignment algorithm into alignment-free mapping, resulting in a 95% increase in speed over alignment-based tools, and up to 15% improvement in accuracy over commonly used tools. We applied Seekmer on single-cell mRNA sequencing data of K562 erythroleukemic cell line, and identified heterogeneous expression of marker isoforms across different cell types. The Python-powered package enables interactive transcriptomic analysis using IPython notebook and Python-based data analysis packages.

1684T

Quality control of highly structured RNA-seq data from the Multimuther Study.

M.W. Stevens^{1,2}, M. Abdalla^{4,5,6}, Y. Jiao^{4,5,6}, G. Leday⁷, J.S. El-Sayed Moustafa⁸, C. Menni⁸, G. Nicholson⁸, C. Holmes⁸, T.D. Spector⁸, M.I. McCarthy^{4,5,8}, S. Richardson⁷, K.S. Small⁸, E.T. Dermizakis^{1,2}. 1) Institute of Genetics and Genomics of Geneva, Geneva, Switzerland; 2) Department of Genetic Medicine and Development, Faculty of Medicine, University of Geneva, Geneva, Switzerland; 3) Department of Twin Research and Genetic Epidemiology, King's College London, London, UK; 4) Wellcome Centre Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, OX3 7BN, UK; 5) Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of Medicine, University of Oxford, Oxford, OX3 7LE, UK; 6) Department of Statistics, University of Oxford, Oxford, OX1 3TG, UK; 7) Medical Research Council Biostatistics Unit, Cambridge, United Kingdom; 8) Oxford NIHR Biomedical Research Centre, Oxford University Hospitals Trust, Oxford, OX3 7LE, UK.

We describe the steps taken in our QC'ing of the highly structured RNA-seq data set generated by the Multimuther study. This study is designed to investigate the use and determine the usefulness of longitudinal multi-omics in identifying patterns associated with clinically relevant phenotypes using the extensively phenotyped TwinsUK cohort. As part of this project, 1014 RNA-seq libraries of twin pairs at three time-points over 7 years were generated. While many large data sets are presumed to be randomly sampled from a population sample, this set is highly structured containing both family structure as well as multiple, longitudinal samples from the same individual providing additional challenges as well as opportunities. Firstly, to take into account both twin structure as well as all time points to minimize bias we processed all RNA samples together and randomized them in the process of library preparation and sequencing. Important early QC steps were taken including checking for sample mislabeling, cross-sample contamination, and PCR amplification biases, which can be estimated by comparing RNA-seq reads to genotypes at both hetero- and homozygous SNPs, as in QTLtools, MBV mode (Fort et al., *Bioinformatics*, 33(12), 2017). However, 7 subjects did not have a genome available. By using the longitudinal structure, we were nonetheless able to check for coherence between time points by extending the MBV methodology to look at similarly hetero- and homozygous variant frequencies but between two RNA-seq libraries from the same person. We characterize the effect of contamination and amplification bias on this statistic using simulations. We provide this QC step as a tool, R2R, <https://qtltools.github.io/r2r/>, for use in longitudinal studies or any study with repeated measurements on the same subject or twin pair. The highly structured nature of the data, particularly the longitudinal, allowed us to residualize the effect of the genomic variation on expression to look for association of expression levels with potential sources of bias, such as RNA quality, QC content, percent exonic reads as well as batch effects due to time of collection. We compare these associations with and without residualization.

1685F

Deep learning in accurate cell type classification and robust surface protein abundance prediction for droplet-based single cell transcriptomic data. H. Xin¹, J. Ma², Y. Jiang^{1,4}, T. Chen³, S. Fong², T. Ideker², W. Chen¹. 1) Department of Pediatrics, University of Pittsburgh, Pittsburgh, PA; 2) Department of Medicine, University of California San Diego, La Jolla, CA; 3) Department of Pathology, University of Pittsburgh, Pittsburgh, PA; 4) School of Medicine, Tsinghua University, Beijing, China.

Recent developments in droplet-based single cell gene expression profiling technologies, such as the Chromium system from 10X genomics, allow biomedical researcher to measure transcript abundance for thousands of cells simultaneously. While cells from the same cell types can be clustered together based on similarities in expression profiles using unsupervised clustering methods, we observe that such approaches often lead to limited accuracy in cell type classification, due to imperfect signature gene markers, heterogeneities across cell lineages, and inevitable dropout events. In this work, we present CytoNet, a novel supervised cell classification method that predicts the abundance of surface protein markers based on the gene expression profile of each cell; and classifies the type of each cell individually thereafter. Specifically, we employed an emerging droplet-based gene expression profiling technology, CITE-seq with cell hashing, which measures the abundance of surface proteins and RNA transcripts together for each cell. We assembled a total of 100k PBMC cells from 9 samples with both RNA expression and surface marker profile available. Then we applied deep learning methods to train a neural network that predicts the surface marker profile given expression profile and classifies each cell into 8 cell types, including CD4+/CD45RA+ T cells, CD4+ T cells, CD8+ T cells, CD8+/CD45RA+ T cells, B cells, monocytes, NK cells and hematopoietic cells. Furthermore, to mitigate input noise introduced by the low sampling rate of droplet-based gene expression profiling technologies, we applied a random-walk-based data normalization method, which averages expression levels of functionally closely-related genes. We evaluated CytoNet with 10-fold cross validation and observed that CytoNet achieves over 90% accuracy in predicting cell types. Finally, we applied CytoNet to over 200k human PBMC cells from both internal and external resources. We compared cell compositions predicted by CytoNet against flow cytometry and observed that CytoNet produces similar results to flow cytometry. We implemented CytoNet using Pytorch and built an interactive webserver at <http://www.pitt.edu/~wec47/CytoNet>, as a one-stop solution of cell type classification for single-cell research community.

1686W

Identification and prediction of cell cycle signatures in induced pluripotent stem cells using single-cell gene expression data. C.J. Hsiao¹, P. Tung², J.D. Blischak¹, J. Burnett², K.K. Dey², K. Barr², Y. Gilad^{1,2}, M. Stephens^{1,3}. 1) Department of Human Genetics, University of Chicago, Chicago, IL; 2) Department of Medicine, University of Chicago, Chicago, IL; 3) Department of Statistics, University of Chicago, Chicago, IL.

The cell cycle is known to control cell fate decisions in cell state transitions, during which the cell cycle is tightly regulated by a family of proteins conserved across species and cell types. However, at the transcriptional level, the effects of the cell cycle on cell fate decisions can vary between different cell types, even for cells at the same cell cycle stage. To date, the cell cycle is receiving a renewed importance due to the rapid advancement of single-cell genomics technology. Especially in the analysis of single-cell gene expression, the cell cycle plays a key role in expression variation across cell states and cell types. The current study was designed to develop effective tools for assessing and predicting the cell cycle in single-cell gene expression data analysis. We collected both Fluorescent Ubiquitination-based Cell Cycle Indicator (FUCCI) – for determining the cell cycle – and single-cell RNA-seq (scRNA-seq) data from each individual using STRT-seq on the Fluidigm C1 platform. The cells were collected from iPSCs derived from 6 genotyped Yoruba cell lines. The experimental design controlled for C1 processing batch effects, as well as individual and gender effects. Using these data, we are developing a supervised approach for predicting the cyclical ordering of single cells in the cell cycle using single-cell gene expression data. We use the FUCCI fluorescent intensities to determine a cyclical ordering of the individual cells, and to assign cell time labels for individual cells representing each cell's position in one complete cell cycle. We estimate the cyclical trend of expression levels for each gene based on the FUCCI-derived cell times, and identify candidate sets of cyclical genes for model training. We train our model in 5-fold cross-validation and evaluate the trained model in held-out validation samples and external datasets. We will compare the prediction results with existing approaches for estimating the cell cycle using single-cell gene expression data: including unsupervised approaches to construct the cyclical ordering of single cells and approaches to categorically assign the cell cycle to individual cells. These results will provide a benchmark for assessing the cell cycle in scRNA-seq data analysis. We also expect that our results will give insights into the effects of the cell cycle on gene expression variation in stem cells.

1687T**Imputation of single-cell gene expression with an autoencoder neural network.**

R. Li¹, B. Liu², M.B. Badsha¹, Y.I. Li³, N.E. Banovich⁴, A.Q. Fu¹. 1) University of Idaho, Moscow, ID; 2) Stanford University, Stanford, CA; 3) University of Chicago, Chicago, IL; 4) Translational Genomics Research Institute, Phoenix, AZ.

Due to dropout and other technical limitations in single-cell sequencing technologies, single-cell RNA-seq (scRNA-seq) data typically contain many zero expression values, resulting in a nonzero rate that are often <20% and can be as low as just a few percent. This low nonzero rate leads to problems in downstream analyses (e.g., obscuring gene-gene relationships, and blurring differences in cell subpopulations). It is of great interest to impute the zeros (i.e., missing values) in this type of data. The idea is analogous to genotype imputation methods that have been used extensively. Whereas those methods exploit the dependence structure across the DNA sequence, we will exploit dependence in gene expression. We have developed imputation algorithms based on autoencoder, a deep learning architecture, for scRNA-seq data. Our architectures make use of the dependence structure across genes and across cells. Our LATE (Learning with AuToEncoder) algorithm trains an autoencoder of 1, 3 or 5 hidden layers de novo on the data. Our TRANSLATE (TRANsfer learning with LATE) algorithm further allows for the use of a reference gene expression panel (e.g., bulk gene expression, a larger scRNA-seq data set, or scRNA-seq data of similar cells types collected elsewhere), and employs the transfer learning technique in deep learning to provide LATE with an initial set of parameter estimates. We assess the performance of LATE and TRANSLATE on simulated and real data, and compare them to existing methods (MAGIC, SAVER, scImpute, DCA, and scVI). In terms of mean squared error, LATE and TRANSLATE perform similarly, and both outperform other methods, across multiple data sets from human and mouse with different numbers of genes and cells and different nonzero rates. Our methods recapitulate nonlinear relationships between gene pairs well; when the cells are collected during a differentiation process, these nonlinear relationships reflect the differentiation trajectory. When the cells are a mixture of cell types, LATE and TRANSLATE both capture the cell types, and TRANSLATE demonstrates better capability in capturing less frequent cell types. Additionally, LATE is highly scalable: it easily runs on a large mouse brain data set (28K genes, 1.3M cells, and a nonzero rate of 7%; from 10x Genomics), and finishes the job within 17 hours on a GPU, whereas no other method can handle this data set.

1688F**BAMM-SC: A Bayesian mixture model for clustering droplet-based single cell transcriptomic data from population studies.**

Z. Sun¹, L. Chen², Q. Huang³, A. Cillo⁴, T. Tabib⁴, Y. Ding¹, J. Kolls⁵, R. Lafyatis⁴, D. Vignali^{4,6}, K. Chen⁴, M. Hu⁷, W. Chen^{1,8}. 1) Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA; 2) Harrison School of Pharmacy, Auburn University, Auburn, AL; 3) The Dietrich School of Arts & Sciences, University of Pittsburgh, Pittsburgh, PA; 4) School of Medicine, University of Pittsburgh, Pittsburgh, PA; 5) School of Medicine, Tulane University, New Orleans, LA; 6) Tumor Microenvironment Center, UPMC Hillman Cancer Center, Pittsburgh, PA; 7) Lerner Research Institute, Cleveland Clinic Foundation, Cleveland, OH; 8) Department of Pediatrics, Children's Hospital of Pittsburgh of UPMC, University of Pittsburgh, Pittsburgh, PA.

Single cell RNA sequencing (scRNA-seq) technologies have been widely used to measure gene expression for each individual cell, facilitating deep understanding of cell heterogeneity. The recently developed droplet-based technology has quickly gained popularity because of its high-throughput and low cost. It is now feasible to conduct population scale single cell transcriptomic profiling studies, where several to tens or even hundreds of individuals are sequenced. A major task of analyzing droplet-based scRNA-seq data is the identification of clusters of single cells with similar transcriptomic profile. To achieve this goal, numerous clustering methods have been developed for unbiased inference of cell types. It is unclear how these methods can be scaled up to study population scale data with tens of thousands of cells collected from many individuals and conditions. In addition, no existing clustering method is tailored for modeling multiple layers of heterogeneity imbedded in population studies. To fill in this gap, we propose a **BA**yesian **M**ixture **M**odel for **S**ingle **C**ell sequencing (BAMM-SC) to simultaneously cluster droplet-based scRNA-seq data from multiple individuals. Specifically, BAMM-SC adopts a Bayesian hierarchical Dirichlet multinomial mixture model, which explicitly characterizes three levels of variabilities (i.e., genes, cell types and individuals). BAMM-SC has the following three key advantages: (1) BAMM-SC is able to account for potential batch effect among multiple individuals, such as unbalanced sequencing depths, variable read lengths and hidden technical biases. (2) BAMM-SC borrows information across different individuals, leading to improved power for detecting individual-shared cell types. (3) BAMM-SC can account for data heterogeneity among multiple individuals, and improves the reproducibility and stability of clustering results. We have conducted comprehensive simulation studies and applied our method to three in-house scRNA-seq datasets collected from mouse lung tissue, human peripheral blood mononuclear cells (PMBCs), and human skin tissue, respectively. We further applied BAMM-SC to a dataset from CITE-seq which is a cutting-edge method that combines multiplexed protein marker detection with transcriptome profiling for single cells. We demonstrated that BAMM-SC outperforms existing clustering methods with improved clustering accuracy and reduced impact from batch effects.

1689W

A generative model for manifold learning for single cell RNA sequencing experiments. *A. Verma, B. Engelhardt.* Princeton University, Princeton, NJ.

Effective analysis of single cell RNA sequencing (scRNA-seq) experiments often requires dimension reduction for visualization, regularization, and efficiency. This practice is based on the manifold assumption, that expression patterns lie on a low dimensional surface that captures the space of possible high dimension gene measurements. A robust manifold learning model should account for the nonlinear and noisy nature of gene expression from single cells. We present a generative, nonlinear latent variable model with robust, heavy-tail error modeling and adaptive kernel learning to properly capture low dimensional nonlinear structure in scRNA-seq data. Each cell's expression is modeled as a noisy draw from a Gaussian process in high dimensions with latent variable weights, or the Gaussian Process Latent Variable Model. We model noise with a heavy tailed Student t-distribution to account for observed technical and biological error including dropouts. An adaptive kernel is used to automatically estimate the smoothness of the manifold and number of latent dimensions. Black Box Variational Inference is used to overcome the intractability of likelihood function for inference of the latent manifold. Batch learning enables scalable fitting for large data sets as experiments grow to hundreds of thousands and millions of cells. We demonstrate our model's ability to accurately capture clustering and developmental behavior from previously published experimental data. Preliminary results show superior separation of labeled cerebral cortex cells compared to current dimension reduction techniques commonly used for analyzing single cell experiments. The generative model also allows quantification of uncertainty in the latent positions and identification of poorly sampled regions of the manifold to inform decisions about further experiments.

1690T

Greatly reducing the noise of single-cell RNA sequencing with transfer learning. *J. Wang, M. Huang, G. Hu, D. Agarwah, Z. Zhou, M. Li, N. Zhang.* University of Pennsylvania, Philadelphia, PA.

Single-cell RNA sequencing (scRNA-seq) has rapidly gained popularity for profiling transcriptomes of hundreds to thousands of cells. This technology has led to the discovery of novel cell types and revealed insights into complex tissues. However, the data generation process is very noisy. There always need to be a trade-off between the number of cells and the sequencing depth. Even when the reads are deeply sequenced, efficiency loss during reverse transcription is still inevitable. We develop a fast computational method, SAVER X, which denoises and greatly improves the quality of scRNA-seq data without an extra budget to increase the number of cells or the sequencing depth. SAVER X is based on the transfer learning technique in deep learning and borrows information from public single-cell RNA sequencing data across protocols, cell types, tissues, and even species. It uses a hierarchical transfer learning framework which benefits both its wide applicability and specificity. Particularly, SAVER X builds up a complex autoencoder network that can learn both common gene networks that are widely preserved and tissue/cell type-specific gene features. The outputs are finally combined with Bayesian shrinkage and cross-validation filtering so that over-fitting is effectively reduced. As a first example, we show using down-sampling experiments of several datasets [Zeisel et.al (2015), Shekhar et.al.(2016)] that SAVER X acts as increasing sequencing depth. By borrowing information from cells in the same tissue and species but are measured with different protocols and laboratories, the down-sampled data can have comparable quality with the original data. Though neither the raw nor the imputed down-sampled data without using our pre-trained model are too poor to identify cell sub-types, these cell sub-types can clearly be detected with the SAVER X denoised data. Immune cells widely exist in many tissues. However, the cell types are very hard to identify using scRNA-seq both due to their similarity and the low total number of immune cells in a specific dataset. As a second example, we show that by pre-training SAVER X using both the half a million immune cells from Human Cell Atlas and PBMC cells from Zheng et. al. (2017), we can more accurately identify the cell types of immune cells from other tissues. Finally, using the data from Manno et. al. (2016) where the brain cells are measured in both human and mouse, we show that information across species is as well helpful. .

1691F

1,000x faster than PLINK: Genome-wide epistasis detection with logistic regression using combined FPGA and GPU accelerators. *D. Ellinghaus, L. Wienbrandt, J.C. Kässens, M. Hübenthal.* Institute of Clinical Molecular Biology, Christian-Albrechts-University Kiel, Kiel, Schleswig Holstein, Germany.

Logistic regression as implemented in PLINK is a powerful and commonly used framework for assessing gene-gene (GxG) interactions. However, fitting regression models for each pair of markers in a genome-wide dataset is a computationally intensive task. Performing billions of tests with PLINK takes days if not weeks, for which reason pre-filtering techniques and fast epistasis screenings are applied to reduce the computational burden. Here, we demonstrate that employing a combination of a Xilinx UltraScale KU115 FPGA and an Nvidia Tesla P100 GPU leads to runtimes of only minutes for logistic regression GxG tests on a genome-wide level. In particular, a dataset with 53,000 samples genotyped at 130,000 SNPs was analyzed in 8 minutes, resulting in a speedup of more than 1,000 when compared to PLINK v1.9 using 32 threads on a server-grade computing platform. Furthermore, on-the-fly calculation of test statistics, p-values and LD-scores in double-precision make commonly used pre-filtering strategies obsolete.

1692W

Improving NCBI's dbSNPdata quality and annotation for variant interpretation. *M.T. Ward, L. Phan, Q. Wang, H. Zhang, G. Shekhtman, T. Murphy, C. Farrell, E. Moses, A. Astashyn, J.B. Holmes, B.L. Kattman.* NCBI/NLM/NIH, Bethesda, VA.

For nearly two decades, dbSNP - NCBI's Short Genetic Variation database - has been serving as the premier reference variation database for many disciplines of genetic research including personal genomics and medical genetics. dbSNP aggregates and annotates genetic variation data from multiple submitters, and assigns stable Reference SNP (rs) identifiers that can be used as citation in publications and for annotating and integrating with other data sources. The database houses variation and frequency data from thousands of submitters, ranging from large-scale projects including HapMap, 1000Genomes, GO-ESP, ExAC, GnomAD, TOPMED and HLI to more focused studies such as locus-specific databases (LSDB) and clinical variants in ClinVar. It has over 650 million human RS identifiers, averaging one variant every five nucleotides across the genome. In this presentation, we describe several approaches that integrate dbSNP with other NCBI resources (Genome, RefSeq, etc.) to improve the data quality and annotation of this unprecedented large variation dataset and to aid variant interpretation and prioritization. - Identify suspect Paralogous Sequence Variants (PSVs) due to sequence differences in paralogs. - Identify potential misreporting of SNVs in repeat regions by analyzing mapping results along with RepeatMasker. - Remap variants from one assembly to another in aligned regions with gaps. - Identify possible false variant calls and other data issues including differences in alleles due to assembly-to-assembly differences. - Add functional and regulatory elements including enhancers, promoters and CpG islands to variant annotation. - Compare data from multiple submitters to check for concordance and for methodological biases for both variant and allele frequencies. This work is supported (in part) by the intramural research program of the National Library of Medicine, National Institutes of Health.

1693T

Integrate genome, phenome and transcriptome data to detect novel disease-gene association using GWAS summary data. *B. Wu¹, B. Guo¹, N. Liu².* 1) University of Minnesota, Minneapolis, MN; 2) Indiana University, Bloomington, IN.

Genetics have the great potential to personalize and improve the disease treatment by tailoring to patients' genetic profiles. In the past decade many large-scale genome-wide association studies (GWAS) have been conducted and successfully identified tens of thousands of genetic variants associated with various diseases and traits. However in total these identified genetic variants explained only a small proportion of the heritability for most traits, suggesting that there remain many more genetic variants to be discovered. This is (i) mainly due to the small effect sizes of most genetic variants, which requires very large sample sizes to achieve decent detection power; (ii) partially due to the current dominating practice of single trait and single variant association test in the field of GWAS analysis, despite the fact that most traits are polygenic in nature; and (iii) the extreme multiplicity associated with testing individual variants genome-wide. We discuss several of our recently developed methods to integrate phenome (multiple traits), genome (multiple variants), and transcriptome data to detect novel disease-gene association that need only GWAS summary data. We build statistical models to develop sound association test methods, and develop efficient and accurate numerical algorithms to quickly compute the analytical p-values for all our proposed methods without the need of resampling. Our methods nicely unify (i) pleiotropy of variants, (ii) polygenicity of traits, and (iii) transcript mediated association, into a coherent modeling framework. We demonstrate the utility of our proposed methods through rigorous numerical studies and analysis of GWAS summary data for multiple lipids traits, glycemic traits, and psychiatric disorders. Our approach identified many novel loci that were not detected by the existing methods. .

1694F

Amplimap: An open-source analysis pipeline for targeted high-throughput sequencing data. *N. Koelling, M. Bernkopf, E. Calpena, G. Maher, K. Miller, H. Ralph, A. Goriely, A.O.M. Wilkie.* MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, United Kingdom.

Next-generation targeted sequencing, for example with PCR-based amplicons or single-molecule molecular inversion probes (smMIPs), is widely used to resequence candidate genes in large panels or to identify low-frequency mutations (<5% allele frequency) using high-coverage sequencing. This enables cost-effective screening of patient cohorts for novel pathogenic mutations and detection of somatic mutations from cancer samples or mosaic cases. While various tools exist to analyse such data, they often either require specialized bioinformatics knowledge or only support a small set of analyses. Pitfalls specific to this type of experiment, for example off-target capture events introducing false positive variant calls, are also common yet not always considered during primer design and analysis. We introduce amplimap, a free open-source pipeline to automate the processing and analysis of data from the most commonly used protocols, including standard PCR-based amplicons and probes with/without Unique Molecular Identifier (UMI) sequences. Taking raw sequencing data (FASTQ) as input, amplimap generates a variety of output files with a single command, including read alignments, per-basepair nucleotide counts (pileups) and annotated germline variant calls. By leveraging knowledge of primer sequences and their corresponding genomic regions, amplimap improves accuracy and identifies off-target capture events. It can also extract UMI sequences from reads to generate consensus pileups: particularly important in ultra-high-coverage applications, where true low-frequency variants must be distinguished from errors introduced during sequencing or PCR amplification. As a Python 3 package, amplimap can easily be installed on Linux and Mac systems. Built on top of Snakemake, it can intelligently determine which processes need to be run and take advantage of computing clusters such as LSF or SGE. Under the hood, it interfaces with commonly used tools such as BWA, GATK, Annovar and SAMtools to follow best practices. All source code is freely available from <https://github.com/koelling/amplimap> amplimap was developed in close collaboration with wet-lab scientists to ensure maximum user-friendliness and wide applicability. Here, this will be illustrated through three case studies: germline variant screening in a patient cohort, quantification of transcript abundance to detect nonsense-mediated decay and validation of low-frequency somatic mutations in a survey of testis samples.

1695W**A workflow to improve variant calling accuracy in molecular barcoded sequencing reads.**

M. Tai¹, C. Yin², G. Smith¹, W. Xu¹. 1) Department of Molecular Genetics, True Health Diagnostics, 6170 Research Rd. Frisco, TX; 2) Department of Mathematics, Statistics, and Computer Science, University of Illinois Chicago, Chicago, IL.

Molecular barcoded multiplexed amplicon sequencing can be used to identify and correct sequencing errors and amplification biases introduced during library preparation in low-frequency variant calling. Barcode aware variant calling algorithms can achieve higher sensitivity detecting variants at very low allelic fractions in somatic tissue. Additionally, a linear relationship between the read depth and the copy number can be more consistently established by reducing PCR duplicates to the unique templates captured during the initial enrichment stages. We propose a generic workflow to improve variant calling accuracy taking advantage of molecular barcoded sequencing reads by applying a base score correction method to duplicate or overlapping read pairs across the targeted panel. The workflow introduces a few additional steps to the typical processing pipeline. First, we strip the molecular barcode from the sequencing read and associate it with the read name for downstream processing. After alignment, we perform steps to mask the primer sequences used for enrichment and apply base score correction on discordant bases called within any overlapping or duplicate read pairs sharing the same barcode for the same primer. After the correction is applied, a deduplication step is performed to select the read pairs with the highest overall base call qualities between duplicates. We demonstrated the ability of the workflow to reduce the false positive rate of variant calls in homopolymer and repetitive regions where the sequencer commonly encounters phasing errors interfering with base calling. The analysis was focused on three specific regions within a custom QIAseq targeted DNA (Qiagen, USA) of 220 genes with high false positive rate in the clinical validation study. Uncorrected and corrected datasets were compared at targeted regions to reference calls from NIST and exome data from a reference laboratory (Macrogen, USA). The base correction method was able to remove all false positives previously identified without the correction method while retaining the true positive call in the dataset. We have shown that our workflow incorporating base correction of molecular barcoded sequencing data can be applied to germline sequencing to improve variant calling accuracy in genomic regions with certain repetitive sequence motifs.

1696T**DeFine: Deep convolutional neural networks accurately quantify intensities of transcription factor-DNA binding and facilitate evaluation of functional non-coding variants.**

M. Wang. Center for Bioinformatics, School of Life Sciences, Peking University, Beijing, 100871, P. R. China.

The complex system of gene expression is regulated by the cell type specific binding of transcription factors to regulatory elements. Identifying variants that disrupt the normal regulation and give rise to human diseases remains a great challenge. To address this, we built and trained sequence-based deep learning regression models that accurately predicted the cell type specific transcription factors ChIP-seq signal intensity, which measured the *in vivo* DNA sequence binding affinity of transcription factors. Unlike existing methods that just predicted the binary outcome binding/unbinding, our model was capable of predicting the continuous TF binding intensity to any sequence. We showed that the deep learning models captured known motifs of different TFs. The capability of accurately predicting the sequence binding affinity enabled us to develop the tool DeFine (Deep learning based Functional impact of noncoding variants evaluator, <http://define.cbi.pku.edu.cn>) with improved performance for assessing the functional impact of noncoding variants including SNPs and indels. Furthermore, DeFine integrates three-dimensional genome contact maps so that it is able to indicate the potential affected genes for each regulatory variant. DeFine is an effective and easy-to-use tool that facilitates systematic evaluation of functional noncoding variants.

1697F

To bind or not to bind: A deep learning approach for explaining TF-DNA recognition. A. Zheng¹, M. Lamkin², H. Su¹, M. Gymrek^{1,3}. 1) Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA; 2) Department of Bioengineering, University of California San Diego, La Jolla, CA; 3) Department of Medicine, University of California San Diego, La Jolla, CA.

Binding of transcription factors (TFs) is one of the major transcriptional regulation mechanisms. Most TFs have intrinsic binding preferences to specific motifs. However, these motifs cannot completely explain TF binding affinity: a TF motif can occur as many as one million times across the human genome, whereas only 1% or fewer of these occurrences may actually be bound *in vivo*. To comprehensively understand TF-DNA recognition, we developed "Agent Bind", a deep learning approach, which enables both prediction and interpretation of TF binding affinity. Previously, several deep learning approaches, such as DeepSEA and DeepBind, proposed to tackle similar problems, including predicting noncoding variant effects and chromatin accessibility. However, these methods restrict analysis to regions with known binding activities and do not include regions without regulatory functions. Thus, they have limited power to identify key elements that contribute to TF-DNA recognition. Importantly, our approach incorporates regions both with and without binding activities, and pay special attention on identifying signals in the motif context that differ between these two categories. The prediction module in Agent Bind consists of a five-layer convolutional neural network (CNN). For each TF, we identified all genomic regions containing the target motif. Regions were labelled as positive or negative based on their overlap with ENCODE ChIP-seq peaks and used as input to our CNN model training. Our overall classification performance reached AUC=0.92 on average, demonstrating that sequence context contains critical additional signals beyond the core motifs for determining TF binding. For the interpretation module, we implemented deep Taylor decomposition, a state-of-the-art model interpretation method, to explain prediction decisions. This method annotates positions in every input sequence with quantitative scores according to their contributions to the prediction results. We found that 78% of the most influential bases overlap known motifs and are highly enriched for motifs recognized by pioneer factors. The additional 22% of scores are concentrated in proximal flanking regions and contain sequence features that are yet to be characterized. Overall, our results show that Agent Bind can shed light on the critical challenge of determining whether a given motif instance will be bound or unbound by its target TF and help in explaining the underlying determinants of TF-DNA recognition.

1698W

Consistent variant annotation using the Clinical Annotation Reference Templates (CART). S. Yost, M. Munz, E. Ruark, S. Mahamdallie, A. Renwick, N. Rahman, *The Transforming Genetic Medicine Initiative (TGMII)*. Division of Genetics, Institute of Cancer Research, London, United Kingdom.

NGS gene testing involves variant annotation against the reference genome. However, in clinical reports nucleotide and protein annotation against a transcript is given. This has the advantage of not changing between genome builds, but the disadvantage that no 'reference transcripts' equivalent to the reference genome is available. This leads to inconsistency because different labs use different transcripts. Moreover, most clinical labs use RefSeq NM transcripts which may not match the reference genome, whereas databases, such as ExAC use Ensembl ENST transcript releases, which match the reference genome, but change frequently. To address these problems we have integrated clinical and scientific data to define a set of Clinical Annotation Reference Templates (CARTs), which can be used to give consistent gene variant annotation and frequency data, similarly to how the reference genome is used to provide consistent genomic annotations. The CART defines the five properties required to annotate gene variation for clinical reports: 1) reference genome build, 2) strand, 3) total number of exons, 4) exon boundary positions, 5) translation start and stop positions. We also provide the ENST and related NM and if/how the NM differs from the reference genome. For each gene the CART matches an ENST, which by definition matches the reference genome and a related RefSeq NM, for which we have generated the matching reference genome sequence. For 17,957 CARTs the coding sequence is identical between build 37 and 38, and 14,350 of these have identical sequence end-to-end. For most genes, using the CART would not alter the clinical annotation, and other annotations will not be required. However, the CART annotation can and should be used in addition to other transcript annotations, where appropriate. For example annotation against tissue-specific transcripts may be appropriate in some clinical or scientific scenarios. The CART annotations aim to deliver a single consistent annotation of a variant, to facilitate accurate interchange and integration of variant calls. Having a stable, versioned, set of templates for annotation will make clinical reporting and data sharing of WES and WGS gene testing simpler and more robust. Our work has also highlighted important discrepancies that need to be addressed, for example when comparing RefSeq and Ensembl transcripts, there are 53 transcripts that match against different genes on build 37 and a further 14 mismatches on build 38.

1699T

ERVcaller: Identifying and genotyping non-reference unfixed endogenous retroviruses (ERVs) and other transposable elements (TEs) using next-generation sequencing data. X. Chen¹, D. Lj^{1,2,3,4}. 1) Department of Microbiology and Molecular Genetics, University of Vermont, Burlington, Vermont 05405, USA; 2) Department of Computer Science, University of Vermont, Burlington, Vermont 05405, USA; 3) Neuroscience, Behavior, and Health Initiative, University of Vermont, Burlington, Vermont 05405, USA; 4) University of Vermont Cancer Center, University of Vermont, Burlington, Vermont 05405, USA.

More than 8% of the human genome is derived from endogenous retroviruses (ERVs). In recent years more and more polymorphic (unfixed) ERVs were found to be associated with human diseases. However, it is still challenging to accurately detect polymorphic ERVs using next-generation sequencing (NGS) data. As such, development of new tools is needed to rectify this shortcoming. We designed a new tool, ERVcaller (<http://www.uvm.edu/genomics/software/ERVcaller.html>), to accurately detect and genotype non-reference unfixed transposable elements (TEs), including ERVs, in the human genome. We evaluated the tool using both simulated and real benchmark whole-genome sequencing datasets. The detection power and precision were ~98% and ~99%, respectively, and the genotyping accuracy was ~97% under 30X sequencing depth. We compared ERVcaller with four existing tools. ERVcaller consistently showed the highest sensitivity and precision for detecting and genotyping non-reference unfixed TEs, particularly ERVs, as well as produce the most accurate breakpoints at single nucleotide resolution, especially under low sequencing depths. By applying ERVcaller to the 1000 Genomes Project samples, we detected 100% of the ERVs and 95% of the TEs that were experimentally validated. In conclusion, ERVcaller is capable of accurately identifying and genotyping non-reference unfixed ERVs and other TEs using NGS data. Additionally, this tool is applicable to analyze NGS data from animals and plants.

1700F

reGenotypeTE: A tool to improve genotype prediction for mobile element insertions. J. Thomas¹, C. Goubert², L.M. Payer³, J.M. Kidd⁴, J.E. Feusier¹, W.S. Watkins¹, K.H. Burns³, C. Feschotte², L.B. Jorde¹. 1) Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, UT; 2) Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY; 3) Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, MD; 4) Department of Human Genetics, Department of Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, MI.

Non-LTR retrotransposons constitute more than one third of the human genome, stemming from their proliferative activity over the last 100 million years. Their continued activity in the human lineage has generated polymorphic insertions, where individuals or populations share the presence or absence of a mobile element insertion (MEI). Such polymorphic MEIs are a source of genetic variation and can lead to regulatory polymorphisms that alter gene expression. The polymorphic MEIs are also used as markers of ancestry in population genetics to understand the structure and evolution of populations. These studies depend on accurate genotyping of MEIs, thereby enabling correct estimations of their allele frequencies. However, accurate MEI genotyping using short-read alignments is inherently challenging due their repetitiveness, and dedicated tools are scarce. In addition, we found that a large proportion of MEIs genotyped in the 1000 Genomes Project (1KGP) show excess heterozygosity compared to SNPs, which contradicts expectations for such largely neutral markers. We thus have developed a pipeline called *reGenotypeTE* that aims to genotype both reference and non-reference MEIs from whole genome paired-end resequencing data aligned to the reference assembly. The pipeline analyzes the MEI breakpoints detected by other tools, such as *MELT*, for non-reference insertions. First, all mapped reads including discordant reads and their mates are extracted for each locus. The consensus sequence of MEI at each locus is either generated through a local *de novo* assembly or extracted from the Repbase database. The pipeline then identifies the hallmarks of non-LTR retrotransposition (i.e. a poly-A tail and target site duplications) and the strand orientation of the MEI. The extracted reads are then mapped to two alleles (with and without TE insertion) separately, and the genotype likelihoods are calculated using established methods. To confirm the accuracy of prediction, we compared *MELT* and *reGenotypeTE* predicted genotypes to PCR-based genotypes generated for 37 reference and 102 non-reference MEIs in 45 CEU individuals from the 1KGP. On average, our genotype predictions were 91% concordant with the PCR genotypes whereas those predicted by *MELT* v2.1.4 were 72% concordant. These data indicate that *reGenotypeTE* improves the accuracy of MEI genotype prediction and will be valuable in identifying candidate regulatory variants and markers of ancestry in population genetics.

1701W

Refining the ability to detect human-specific LINE-1 insertions using long-read sequencing technology. *W. Zhou¹, S. Emery², D. Flasch², Y. Wang², K. Kwan^{2,3}, J. Kidd^{1,2}, J. Moran^{2,4}, R. Mills^{1,2}.* 1) Department of Computational Medicine and Bioinformatics, University of Michigan Medical School, 100 Washtenaw Avenue, Ann Arbor, MI 48109, USA; 2) Department of Human Genetics, University of Michigan Medical School, 1241 East Catherine Street, Ann Arbor, MI 48109, USA; 3) Molecular and Behavioral Neuroscience Institute, University of Michigan Medical School, 109 Zina Pitcher Place, Ann Arbor, MI 48109, USA; 4) Department of Internal Medicine, University of Michigan, 1500 East Medical Center Drive, Ann Arbor, MI 48109, USA.

Long Interspersed Element-1 derived sequences (LINE-1s or L1s) comprise ~17% of the human genome. The mobilization of active L1s contributes to inter- and intra-individual genetic variation, and L1 insertions into genes occasionally can lead to human genetic disorders. Various strategies have been developed to identify human-specific LINE-1 (L1Hs) insertions from short-read whole genome sequencing (WGS) data; however, these approaches are limited in their ability to identify L1Hs insertions that reside within genomic regions containing pre-existing LINE-1 sequences. The advent of single molecule, third generation Pacific Biosciences (PacBio) DNA sequencing has allowed the direct sequencing of long stretches of contiguous DNA, permitting an additional approach to identify L1Hs insertions. Here, we describe the development of a computational pipeline (PALMER) to identify non-reference L1Hs insertions in PacBio DNA sequencing data. PALMER identified 199 non-reference L1Hs insertions in the well-characterized NA12878 sample, including 92 that were not detected using standard short-read WGS data, representing a ~2-fold enrichment of L1Hs insertions in NA12878. Using this L1Hs dataset as a gold standard, we then assessed the ability of WGS to identify germline L1Hs insertions in single-cell sequencing data obtained from four individual NA12878 cells. Approximately 66% of L1Hs insertions were not detected in our single-cell experiments and the majority of these L1Hs insertions were nested inside of reference LINE-1 sequences. We observed a wide variation in L1Hs insertion detection accuracy between individual single-cell WGS experiments. Our preliminary data further suggest that a 3' targeted L1 capture assay had a higher sensitivity and lower false discovery rate at detecting L1Hs insertions when compared to single-cell WGS. Together these results indicate that ongoing studies of mobile element insertions in bulk and single cell experiments using standard WGS approaches may underestimate the numbers of L1Hs insertions near or nested within reference LINE-1s.

1702T

Manta: Improving the accuracy of structural variant calling. *X. Chen, N. Narai, A. Gross, S. Ji, C. Saunders.* Illumina, Inc., 5200 Illumina Way, San Diego, CA, USA.

Manta is an open-source method designed to discover structural variants (SVs) from next generation sequencing data. Manta is optimized for rapid germline and somatic analysis, calling SVs, medium-sized indels and large insertions on standard compute hardware. Since its initial publication in 2015, ongoing development of Manta has led to improved call accuracy and expanded scope of applications such as multi-sample germline calling and tumor-only calling. A TIGRA-like, iterative assembler has been recently introduced in Manta to improve the quality of alternative contigs assembled from evidence reads. The new iterative assembler outperformed a few state-of-the-art local assemblers in our comparative analysis, and has been shown to improve the call sensitivity across SV types and sizes, and to be able to accurately assemble insertions of size larger than 500 bp with high-depth, short-read data. Furthermore, the contig aligners are tuned to improve the germline call precision in noisy and/or low complexity regions, particularly for indels of size between 51 and 200 bp and large SVs of size more than 10 kbp. This study reports improved Manta performance for both germline and somatic calling, as determined by comparison of germline calls to GIAB germline calls and comparison of somatic calls to COSMIC database variants. This study also highlights directions of ongoing Manta improvements including utilization of depth information to improve the precision of deletion/duplication calls, and recovery of unmapped reads to improve the accuracy of large insertion calls. Such improvements reflect the development work of Manta towards an accurate, fully featured structural variant caller.

1703F

Highly-accurate detection of SNVs from long read sequencing with high error rates. *P. Edge¹, V. Bansal².* 1) Department of Computer Science Engineering, University of California, San Diego, La Jolla, CA; 2) Department of Pediatrics, University of California, San Diego, La Jolla, CA.

Single molecule sequencing (SMS) technologies such as Pacific Biosciences (PacBio) Single Molecule Real-Time (SMRT) sequencing have seen increased adoption in recent years. Compared to short reads, SMS long reads have better capability for genome assembly, haplotype assembly, and resolution of large structural variations. Further, long reads are capable of uniquely mapping to more of the genome than short reads, including coding regions coinciding with gene duplications. However, the high indel error rate of SMS reads makes the detection of Single Nucleotide Variants (SNVs) more challenging than with short Illumina reads, in particular for heterozygous variants. For this reason, there are limited methods for calling SNVs in diploid organisms using SMS reads. We introduce REAPER, a realignment-based diploid SNV caller that calls haplotype-resolved SNVs from long reads. REAPER uses a realignment-based approach to allelotype and estimate quality values for potential variant sites, and iteratively refines genotypes and haplotypes using the HapCUT2 algorithm. We use REAPER to call SNVs with high precision and recall for four human individuals, and validate the calls against high-confidence variant calls from the Genome in a Bottle (GIAB) consortium. We find that REAPER calls variants with excellent precision and recall (see table), achieving 99.6% precision and 97.0% recall from 44x SMS reads for the NA12878 genome. We analyze the locations of false positive SNVs and find that 80.7% occur at true indel variant sites, indicating that classification of these sites can improve the precision even further. Further, we use simulated SMS reads to show that REAPER is capable of calling variants in highly similar (99.9% similarity) segmental duplications. REAPER called variants in these regions with a recall of 93.2% from 40x coverage simulated SMS reads, compared to 33.9% recall using 40x coverage simulated short reads. The ability to call variants in low-mappability regions allows the detection of variants in highly similar gene paralogs and more complete profiling of human genetic variation.

Table: Summary of variants called on chromosome 1 with REAPER using SMS reads for 4 GIAB individuals.						
Genome	SMS read coverage	SNVs called	Precision	Recall	Outside GIAB high-confidence	Runtime (hours)
NA12878	~44x	281,525	0.996	0.97	40,673	13:39
NA24385	~69x	277,921	0.99	0.975	42,235	20:39
NA24149	~32x	225,890	0.997	0.837	30,097	5:51
NA24143	~30x	212,458	0.997	0.767	26,415	5:31

1704W

Comprehensive characterization across the spectrum of tandem repeat variants. *H. Kang, A. Tan.* Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI.

Accurately detecting polymorphic alleles from a population of genome sequences is a very important step to comprehensively understand the genetic architecture of disease-related traits. One of the known challenges in variant calling is that Indels calls are much less reproducible than SNPs and the discordance rate of indels is particularly high in the repetitive (low-complexity) regions of genome. This is because tandem repeat variants (TRVs), which includes STRs (Short Tandem Repeats or microsatellites) and VNTRs (Variable Nucleotide Tandem Repeats or minisatellites), encompass a large proportion of indels, especially in repeat-rich regions, and they are more prone to alignment errors and misleading representations. As a result, existing callers tend to undercall TRVs or only calls a specific class of TRVs, making a comprehensive understanding of TRVs very challenging. We present a principled way to characterize TRVs across the entire spectrum of tandem repeats variants, and present a robust algorithm for comprehensively detecting and characterizing TRVs among indels. Unlike previous methods that rely only on reference genomes to identify TRVs, our algorithm integrates reference sequence and allelic variation together to infer "allelic window", which allows us to determine repeat tracts and units. Starting from an arbitrary variant call set in VCF (Variant Call Format), our method groups specific indels together based allelic window inferred jointly from reference and non-reference alleles. Using the allele windows, we were able to quantify relative abundance of observed TRVs across all allele lengths and repeat context. We found that the TRV is 10 times more enriched than average when allele window is 5, and more than 100 times enriched when allele window is greater than 10, across all allele lengths. Our results can be used to model relative mutation rates of TRVs, or to be used as priors for calling Indels and TRVs. Furthermore, our method allows us to properly integrate multiple TRVs in the presence of alignment and sequencing errors, accounting for repeat-rich sequence context, especially multi-allelic indels. Instead of dichotomously classifying indels and TRVs separately, we demonstrate that dual representation of Indels as Indel or TRV allows us to determine more accurate representation of the variant at the time of genotyping. Our algorithms are publicly available at <http://github.com/atks/vt>.

1705T

A comparative evaluation of accelerated variant calling pipelines for human germline whole-genome sequencing data. *S. Maciuga, T. Hickman, S. Chacko, A. Daly, K. Walter, J. Randall.* Wellcome Sanger Institute, Cambridge, United Kingdom.

The analysis of next-generation sequencing data typically starts with mapping reads to the reference genome, followed by calling small variants observed in the read pileup, in the form of SNPs and INDELs. The accuracy of this process is critical to downstream analyses of variant interpretation and disease association. A number of tools have been published, widely used and typically benchmarked against high-confidence variant calls from a small number of well-characterized human genomes, developed by projects such as *Genome in a Bottle* and *Platinum Genomes*. The limitation of these gold standard call sets is that they were built by integrating variant callers based on short reads, which biases testing towards genomic regions that are accessible to short-read technologies. More recently, a new “truth set” was built from a synthetic diploid mixture of two haploid PacBio assemblies of hydatidiform mole cell lines, which are homozygous across the whole genome. This truth set provides a less biased benchmarking framework because it was derived from a method orthogonal to the ones being tested, based on long-read sequencing data. With growing initiatives for population-scale genome sequencing (*100,000 Genomes Project*, *UK Biobank*), there is a need to evaluate variant calling pipelines in the context of computational performance. In the past year, several hardware-accelerated implementations have been released, such as **GATK4** with Intel FPGA (field-programmable gate array) and avx512 support and **DeepVariant**, with GPU (graphics processing unit) support. We evaluate these solutions against the synthetic diploid truth set, along with standard **GATK4** deployed on a public and private cloud, in order to assess their scalability to very large sample sizes, on the order of 10s to 100s of thousands of genomes. While offering a robust genome sequence to compare against, the synthetic diploid benchmark has the limitation of consisting of a single sample. For multi-sample calling, we perform testing on 3 additional data sets and compare with results obtained from independent sources. We assess concordance with genotyping chip data, Mendelian inconsistencies for a set of trios and heterozygous calls in regions of homozygosity for a set of consanguineous samples.

1706F

Joint calling and PacBio SMRT Sequencing for indel and structural variant detection in populations. *A.M. Wenger, A. Toepfer, Y. Li, L. Hickey, J. Korlach.* PacBio, Menlo Park, CA.

Fast and effective variant calling algorithms have been crucial to the successful application of DNA sequencing in human genetics. In particular, joint calling – in which reads from multiple individuals are pooled to increase power for shared variants – is an important tool for population surveys of variation. Joint calling was applied by the 1000 Genomes Project to identify variants across many individuals each sequenced to low coverage (about 5-fold). This approach successfully found common small variants, but broadly missed structural variants and large indels for which short-read sequencing has limited sensitivity. To support use of large variants in rare disease and common trait association studies, it is necessary to perform population-scale surveys with a technology effective at detecting indels and structural variants, such as PacBio SMRT Sequencing. For these studies, it is important to have a joint calling workflow that works with PacBio reads. We have developed pbsv, an indel and structural variant caller for PacBio reads, that provides a two-step joint calling workflow similar to that used to build the ExAC database. The first stage, discovery, is performed separately for each sample and consolidates whole genome alignments into a sparse representation of potentially variant loci. The second stage, calling, is performed on all samples together and considers only the signatures identified in the discovery stage. We applied the pbsv joint calling workflow to PacBio reads from twenty human genomes, with coverage ranging from 5-fold to 80-fold per sample for a total of 460-fold. The analysis required only 102 CPU hours, and identified over 800,000 indels and structural variants, including hundreds of inversions and translocations, many times more than discovered with short-read sequencing. The workflow is scalable to thousands of samples. The ongoing application of this workflow to thousands of samples will provide insight into the evolution and functional importance of large variants in human evolution and disease.

1707W

Pluto: Ultra-fast haplotype phasing and genotype refinement tool for ultra-low coverage DNA sequence reads. *F. Zhang, H.M. Kang.* University of Michigan, Ann Arbor, MI.

As the number of deeply sequenced genomes increases, genomes sequenced in ultra-low coverage is continuously growing in various contexts, including non-invasive prenatal testing (NIPT), targeted sequencing with off-target coverages, and ultra-low coverage genome sequencing as alternatives of SNP arrays. Insufficient coverage inflates the error rate in genotype calling, and haplotype-based genotype refinement is a critical step to obtain high-quality genotypes from ultra-low coverage sequence reads. Recently, statistical methods to leverage a large and diverse reference haplotype panel, such as Haplotype Reference Consortium (HRC) are demonstrated to be effective in haplotype phasing or imputation. However, for ultra-low coverage sequencing, existing tools are inappropriate due to either high computational cost or the lack of ability to account for genotype uncertainty. We propose an efficient statistical method *Pluto* that enables haplotype phasing and genotype refinement from individual genomes sequenced in ultra-low coverage with orders of magnitude smaller computational cost compared to alternative methods such as Beagle. Our method integrates the Positional Burrows-Wheeler Transform (PBWT) and Variable Length Markov Chains (VLMC) to build a haplotype graph from reference haplotypes to account for genotype uncertainty. We leverage statistical methods, such as Kolmogorov-Smirnov (KS) test, to accurately and efficiently build haplotype graphs for VLMC from conditional distributions rapidly obtained from PBWT. Our experiments show that our method can achieve comparable genotype accuracy with Beagle under various sequencing depth between 0.1x or 4x when using 1000 Genomes as reference haplotypes. Specifically, our initial results show that *Pluto* provides relatively lower homozygous error rate but higher heterozygous error rate than Beagle. The computational speed of *Pluto* is 8.5x faster than Beagle when evaluated in 4x whole genome sequence reads.

1708T

Profound perturbation of the human metabolome by obesity. *E.T. Cirulli¹, L. Guo², C. Leon Swisher³, N. Shah¹, L. Huang¹, L.A. Napier⁴, E.F. Kirkness¹, T.D. Spector⁵, C.T. Caskey⁶, B. Thorens⁵, J.C. Venter⁶, A. Telenti⁷.* 1) Human Longevity, Inc., San Diego, CA; 2) Metabolon, Inc., Durham, NC; 3) Department of Twin Research and Genetic Epidemiology, King's College London, London, U.K; 4) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 5) Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland; 6) J. Craig Venter Institute, La Jolla, CA; 7) The Scripps Research Institute, La Jolla, CA.

Obesity is a heterogeneous phenotype that is crudely measured by body mass index (BMI). More precise phenotyping and categorization of risk in large numbers of people with obesity is needed to advance clinical care and drug development. Here, we used non-targeted metabolome analysis and whole genome sequencing to identify metabolic and genetic signatures of obesity. We collected anthropomorphic and metabolic measurements at three timepoints over a median of 13 years in 1,969 adult twins of European ancestry and at a single timepoint in 427 unrelated volunteers. We observe that obesity results in a profound perturbation of the metabolome; nearly a third of the assayed metabolites are associated with changes in BMI. A metabolome signature identifies the healthy obese and also identifies lean individuals with abnormal metabolomes – these groups differ in health outcomes and underlying genetic risk. Because metabolome profiling identifies clinically meaningful heterogeneity in obesity, this approach could help select patients for clinical trials.

1709F

Single-nucleus RNA-sequencing of human subcutaneous adipose tissue reveals a heterogeneous population of 12 cell types. *M. Alvarez¹, K. Garske², J.N. Benhammou^{1,2}, J.R. Pisegna^{1,2}, K.L. Mohlke³, M. Laakso⁴, K.H. Pietiläinen⁵, P. Pajukanta^{1,6,7}.* 1) Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA; 2) Vache and Tamar Manoukian Division of Digestive Diseases, UCLA, Los Angeles, CA, USA; 3) Department of Genetics, University of North Carolina, Chapel Hill, NC, USA; 4) Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland and Kuopio University Hospital, 70210 Kuopio, Finland; 5) Obesity Research Unit, Research Programs Unit, Diabetes and Obesity, University of Helsinki, Biomedicum Helsinki, Helsinki, Finland; 6) Bioinformatics Interdepartmental Program, UCLA, Los Angeles, USA; 7) Molecular Biology Institute at UCLA, Los Angeles, CA, USA.

Adipose tissue functions as both a metabolic and an endocrine organ that plays a key role in obesity and cardiovascular disorders. Furthermore, several cell types, including adipocytes, immune cells, and mesenchymal cells, interact to affect adipose function at the tissue level. However, the detailed cellular landscape of human adipose tissue has not been precisely determined. We used single-nucleus RNA-sequencing (snRNA-seq) to identify cell types in adipose tissue based on nuclear gene expression. We isolated nuclei from human subcutaneous adipose tissue and prepared single-nuclei RNA Illumina libraries using the 10X Chromium platform. Raw reads were aligned to the hg38 human genome with the 10X CellRanger software. Utilizing Seurat to cluster nuclei based on expression, we identified 12 cell-types in the human adipose tissue, suggesting that this tissue consists of more cell types than previously reported using FACS sorting. Differential expression analysis identified signature genes specific to each cluster and pinpointed adipocyte, endothelial cell, macrophage, dendritic, T-cell, and 7 potentially novel cell-types. To further explore these 7 cell types, we searched for their pathway enrichments using DAVID. Three of the 7 clusters were enriched for extracellular matrix processes, suggesting that they are preadipocyte- and fibroblast-related cell-types. The other 4 clusters showed enrichments for muscle/contractile and mitochondrial-related processes. Lastly, we explored if these cluster-specific marker genes exhibited associations between whole adipose tissue gene expression and metabolic traits in bulk adipose RNA-seq data from the METSIM cohort (n=335). Six of the 12 cell-type clusters were enriched for genes whose expression level is correlated with ≥ 1 of 47 metabolic traits ($p < 3.14 \times 10^{-3}$). For example, adipocyte signature genes were enriched for genes whose expression is correlated with serum triglycerides (TGs) (OR=4.82; $p = 8.18 \times 10^{-19}$), and 92.9% of these genes had a negative correlation, suggesting that TGs correlate negatively with the proportion of adipocytes in METSIM adipose transcriptomes. Taken together, snRNA-seq from adipose tissue revealed a heterogeneous population of 12 cell-types, each expressing a unique set of markers. These results, leveraging bulk RNA-seq and snRNA-seq data from human adipose tissue, provide novel insight into adipose function and can be used to decompose cell-type proportions contributing to disease phenotypes.

1710W

Metabolite Quantitative Trait Loci (mQTLs) provide insights into genetically driven pathobiology in asthma. R.S. Kelly¹, B.A. Raby¹, B.D. Levy², S.T. Weiss¹, J.A. Lasky-Su¹. 1) Channing Division of Network Medicine, Brigham and Womens Hospital, Boston, MA; 2) Department of Medicine, Brigham and Womens Hospital, Boston, MA.

Inflammatory mediators play a key, but not fully understood, role in the pathogenesis of asthma; with metabolites derived from omega-6 and omega-3 fatty acids of particular importance. *FADS2* (fatty acid desaturase 2) encodes a crucial rate limiting enzyme within these unsaturated fatty acid pathways, and has been linked to asthmatic phenotypes. Metabolomics, which captures both genetic and environmental influences and current phenotype, is ideally suited to explore the downstream functional implications of genetic variants in *FADS2* on asthmatic phenotypes. The Childhood Asthma Management Program is a randomized clinical trial investigating the long-term effects of inhaled treatments for asthma in children. Blood samples from 375 participants with asthma were submitted for mass-spectrometry based metabolomic, transcriptomic and genome-wide profiling. A metabolite Quantitative Trait Loci analysis (mQTL) was performed using the R package *matrIXeQTL*, to identify metabolites whose abundance was associated with a functional variant in *FADS2*; rs968567. Mediation analysis was conducted to determine whether the genetic burden of disease severity in asthma was mediated through alterations in the levels of these metabolites. Rs968567 [C>T] was shown to be associated with increased expression of *FADS2*; ($p=6.6 \times 10^{-14}$); with multiple lung function metrics, including bronchodilator response ($p=0.013$), and with 67 metabolites; the majority of which were omega-6 or omega-3 derived fatty acids, and which were themselves associated with asthmatic lung function. Mediation analysis revealed that arachidonic acid ($p<2 \times 10^{-16}$); eicosapentaenoic acid ($p<2 \times 10^{-16}$); docosapentaenoic acid ($p=0.06$) and docosahexanoic acid ($p=0.07$) were mediating the relationship between the *FADS2* variant and bronchodilator response, although only a proportion of the change attributable to rs968567 was explained by these metabolites. This study leverages population-wide integrative-omic data to demonstrate *FADS2* variants may act as mQTLs driving differential abundance of key inflammatory mediating metabolites that influence asthma severity. The balance of omega-3 versus omega-6 fatty acid conversion regulated by *FADS2* in the alpha-linolenic acid pathway is crucial for the resolution of inflammation and dampening of airway hyperresponsiveness. However, mediation analysis suggests that *FADS2* may also exert its genetic influence on asthma through the regulation of additional metabolic pathways.

1711T

Gut microbiota composition, DNA methylome and susceptibility of developing allergic diseases in children living in rural and urban. Z. Yang^{1,2}, Z. Chen², Z. Zheng¹, L. Zhu¹, X. Chen², C. Wang^{1,2}, J. Li¹. 1) Department of Allergy and Clinical Immunology, the National Center for Respiratory Diseases, State Key Laboratory of Respiratory Disease, Guangzhou Institute of Respiratory Health, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China; 2) Center for Genomics and Department of Basic Sciences, School of Medicine, Loma Linda University, CA 92350, USA.

The interplay between the diet and reduced microbiome diversity has been linked to allergy diseases. Our previous study has shown a marked difference in dietary patterns between urban and rural children exposed in different environments. However, it is unknown if this different environmental exposure and diet patterns between urban and rural children will give rise to different susceptibility of developing allergic diseases. Our study is to investigate whether different environmental exposure-modulated gut microbiota and DNA methylome may be related to food allergy (FA) between urban and rural children in Guangdong, China. A questionnaire survey was performed in 11,473 children aged 7–12 years old from urban Guangzhou and rural Shaoguan, China. A nested age-matched case-control consisting of 402 children from Guangzhou city and 349 children from Shaoguan rural area was studied. Skin-prick test and serum-specific IgE to allergens were determined. We conducted 16S rRNA gene sequencing focused on V3-V4 region consisting of 226 stool samples, i.e., 92 urban FA, 53 urban control, 38 rural FA and, 37 rural controls. We also carried out metagenomic shotgun sequencing on 96 stool DNA samples from children living in urban and rural (24 samples in each group). In addition, we obtained peripheral blood DNA methylomes on 36 subjects (both FA and controls) using Illumina Human Methylation 450k beadchips. Our study showed although the self-reported food allergy prevalence (4.0% vs. 3.5%) was not significantly different between urban Guangzhou and rural Shaoguan, there was a distinct difference in dominant genus or endotype with *Prevotella* more prevalent in rural children vs. *Bacteroides* more prevalent in urban children. In addition, there was a significant different DNA methylation profiles between food allergy subjects and healthy controls. We identified ninety-one candidate differential CpGs in urban and rural samples. Our preliminary data suggest urbanization or long-term lifestyle may have a significant impact on gut microbiota and there is a distinct different dominant genus or endotype between food allergy and control. In addition, there is a marked different blood DNA methylation profiles between food allergy and control. Overall, different environment-modulated gut microbiota and DNA methylation may be one of the main determining factors for different susceptibility of developing allergic diseases between urban and rural children in China.

1712F

Transcriptomics and metabolomics show the *ORMDL3* asthma gene regulates pleiotropic effects on epithelial inflammation. Y. Zhang¹, S. Willis-Owen¹, S. Spiegel², C. Lloyd¹, M. Moffatt¹, W. Cookson¹. 1) National Heart and Lung Institute, Imperial College London, London, London, United Kingdom; 2) Department of Biochemistry and Molecular Biology and the Massey Cancer Center, Virginia Commonwealth University School of Medicine, Richmond,.

Human chromosome 17q21 locus confers the major genetic susceptibility to childhood asthma, within which asthma-associated SNPs are strongly correlated with the transcript abundance of *ORMDL3*. The locus influences disease severity and the frequency of human rhinovirus (HRV) initiated exacerbations. The mechanisms how *ORMDL3* modulates airway inflammation and infection remain unclear. In order to investigate the role of *ORMDL3* in cellular inflammation, we performed time-series of IL1B-induced inflammation in A549 cells, using cytokine production as outputs and testing effects of *ORMDL3* siRNA knockdown, *ORMDL3* overexpression, and the SPT antagonist myriocin. We replicated selected findings in primary human bronchial epithelial (NHBE) cells. We measured global gene expression and biochemical profiles in quadruplicate. Silencing *ORMDL3* led to steroid-independent reduction of IL6 and IL8 release and reduced ER stress after IL1B stimulation. Overexpression of *ORMDL3* and myriocin increased IL6 and IL8 release. Silencing *ORMDL3* reduced expression of genes regulating host-pathogen interactions, stress responses and ubiquitination, in particular strongly reduced expression of the HRV receptor *ICAM1*. Silencing led to changes in levels of transcripts and metabolites integral to glycolysis. Increased concentrations of ceramides and the immune mediator sphingosine-1-P (S1P) were also observed. The results show *ORMDL3* has pleiotropic effects during cellular inflammation, consistent with its substantial genetic influence on childhood asthma and provide potential new therapeutic targets for the disease.

1713W

Pan-Immunome Project: Building an integrated T cell receptor repertoire profile for immune-related diseases. X. Liu^{1,2}, W. Zhang^{1,2}, J. Wu^{1,2}. 1) BGI, shenzhen, China1BGI-Shenzhen, Shenzhen, 518083, China; 2) China National GeneBank-Shenzhen, BGI-Shenzhen, Shenzhen 518083, China.

The human adaptive immune system provides protection against internal and external environmental pressures, such as genome somatic mutation and infected pathogens. The protection function is directly triggered by the receptors on the surfaces of T cells (TCRs) and B cells (BCRs) to specifically recognize antigens. However, up to now, it is frustrating to exactly detect some immune related diseases such as autoimmune diseases, which hinder the disease prevention and treatment. Here, we proposed a pan-immunome project, building an integrated T and B cell receptor repertoire map for immune related disease, and aimed to explore and mine the disease's immune characteristics. As the early stage of this project, we have established a pan-immune-repertoire database (PIPD, <https://db.cngb.org/pird/>) in China National GeneBank to store TCR and BCR repertoire data of this project. We have collected more than 10,000 patients comprising sorts of autoimmune diseases and thousands of healthy people. All samples have been amplified TCR b repertoire and sequenced using high throughput sequencer. We developed a bioinformatic pipeline to cluster the large number of discrepant TCR b clones that are associated with the same antigen. The method has been evaluated using previously reported clones associated with certain antigen. As for each disease, the disease specific clustered clones were identified from TCR b repertoire of thousands of patients using machine learning. A statistical model was generated to diagnose diseases using those disease-associated TCRs, and performed well with high specificity and sensitivity. Thus, the immune repertoire data could reflect the individual immune status and serve as a strong biomarker for molecular classification of immune related diseases, especially for autoimmune diseases that are fairly complicated for pathogenesis and have not exactly diagnostic method for early stage patients. It is worthwhile to build an enormous immune database with kinds of diseases and healthy samples, for the potential applications, such as assisting diagnosis, prognosis evaluation, early screening and risk evaluation of disease.

1714T

Fine-mapping autoimmune disease variants in cytokine induced cell states. B. Soskic¹, E.C. Gamez¹, D. Plowman¹, N. Nakic², D. Tough², W. Rowan², C. Larmine², T. Carlson⁴, P.G. Bronson³, K. Estrada³, G. Trynka¹.

1) Wellcome Sanger Institute, Cambridge, United Kingdom; 2) GSK, Stevenage, United Kingdom; 3) Statistical Genetics & Genetic Epidemiology, Biogen, Cambridge, MA, USA; 4) Tempero Pharmaceuticals, GlaxoSmithKline Co., Cambridge, MA, USA.

Autoimmune disease variants are enriched in active chromatin regions of stimulated T cells and macrophages. The functional response of both cell types differs depending on the environmental cues. Here, we sought to investigate whether disease variants play a role in specific cytokine induced conditions. Therefore, we stimulated T cells and macrophages in the presence of different autoimmune disease relevant cytokines in two time points and profiled active enhancers with H3K27ac ChIP-seq and regions of open chromatin with ATAC-seq, generating a total of 54 T cell and macrophage cell states. We observed that the majority of peaks were shared across conditions, however the levels of acetylation of H3K27 and chromatin accessibility differed. Using the available SNP enrichment methods, we were unable to fine-map disease relevant cell states, as these methods are based on the binary overlap between peaks and disease variants. We therefore developed a more sensitive method that uses the changes in histone modifications and chromatin accessibility to prioritise relevant cell states. We observed that rheumatoid arthritis (RA), multiple sclerosis (MS), type 1 diabetes (T1D), Crohn's disease (CD), ulcerative colitis (UC) and inflammatory bowel disease (IBD) variants were enriched in peaks that were upregulated during early activation of naïve and memory T cells irrespective of the cytokine induced environment. At the late time point, IBD, CD and UC were specifically enriched in regulatory regions active in Th1 and Th17 cells, as well as cells stimulated in the presence of IFN- γ . Other immune traits were not enriched in any of the conditions at the late time point. Furthermore, CD and IBD variants were also enriched in regulatory regions specific to TNF α and IL-23 polarized macrophages. Our method allows for discrimination between cell subsets with subtle changes in chromatin activity and predominantly points towards the effects of autoimmune variants in early T cell activation irrespective of cytokine polarization. In addition, we have demonstrated that IBD, CD and UC variants are enriched in regulatory regions active in Th1 and Th17 cells. Disclaimer: The human biological samples were sourced ethically and their research use was in accord with the terms of the informed consents under an IRB/EC approved protocol.

1715F

Development and validation of next-generation sequencing assay for carrier screening of alpha-thalassemia. C. Ye, Q. Cai, T. Kuang, X. Meng. WuXi NextCODE, Shanghai, China.

Alpha-thalassemia ([MIM 604131]) is one of the most common genetic diseases worldwide, especially in southern China. It usually results from single or double deletions of alpha globin genes within the alpha globin gene cluster that spans about 30 kb on the genome. Conventionally, Gap-PCR or MLPA is used for alpha-thalassemia detection. However, these PCR-based methods are costly, low throughput, and difficult to identify novel variant. Recently, NGS is promising as a replacement for patient-based and population-based carrier testing in Mendelian recessive diseases. However, NGS-based carrier screening for alpha-thalassemia is challenging due to the identical coding sequences of *HBA1* and *HBA2* and its complicated genotypes. In our developed NGS assay, besides exonic regions, capture-based probes were also designed for *HBA1/2* intronic regions and UTRs. Then, the normalized read depths of the unique regions were used for CNV calling. The called CNV was further genotyped by the maximum likelihood estimation algorithm. In validation, both NGS and MLPA assays were conducted for 986 samples. All 28 carriers (heterozygous deletion) and normal samples identified by MLPA were all successfully called by NGS assay. In addition, a novel short deletion (~100bp) on exon 1 of *HBA2* was only found through NGS assay. The frequencies of alpha-thalassemia subtypes ($\alpha\alpha$ - α^{Δ} , $\alpha\alpha$ - $\alpha^{3.7}$, and $\alpha\alpha$ - $\alpha^{4.2}$) in this cohort were 0.6%, 1.92% and 0.2%, respectively. Another 600 samples were enrolled in a prospective NGS carrier screening program with optional MLPA confirmation. According to NGS results: 1) 7 samples were carriers; 2) 32 samples were low confident results; 3) the rest 561 samples were high confident negatives. According to confirmations by MLPA, 1) the sensitivity is 77.8% (2 low confident results were later found as positives); 2) the PPV is 100%; 3) the estimated specificity is 100%. The performance could be improved by balancing the low sensitivity and high PPV/specificity in future. Particularly, these two false negative samples resulted from bad coverage evenness due to batch effects. The total alpha-thalassemia carrier frequency in this cohort (eastern Chinese) as a result is 1.5% (9/600). In conclusion, we have developed and validated a NGS assay for carrier screening of alpha-thalassemia. Our results also indicated that clinical application should take more factors, such as batch effects and quality problems, into consideration in addition to analytical validations.

1716W

Functional dissection of common genetic variation in the autoimmune-associated *TNFAIP3* locus. J.P. Ray¹, C.G. de Boer¹, C. Fulco^{1,2}, R. Tewhey^{1,2}, C. Lareau^{1,2}, J.C. Ulirsch^{1,2}, L. Ludwig^{1,2}, S.K. Reilly^{1,2}, J. Engreitz¹, A. Regev^{1,4,5}, N. Hacohen^{1,3}. 1) Broad Institute, Cambridge, MA; 2) Harvard University, Cambridge, MA; 3) Massachusetts General Hospital, Boston, MA; 4) Massachusetts Institute of Technology, Cambridge, MA; 5) Howard Hughes Medical Institute, Cambridge, MA.

Common genetic variation has been shown to explain a significant component of the heritability for complex, non-Mendelian inheritance of many diseases, including autoimmune diseases. However, moving from disease association to causal variant is extremely difficult, in part due to linkage disequilibrium, and in most cases neither the causal variant, the cell types in which it functions, nor its mechanism of action are known. *TNFAIP3* encodes for the well-known A20 protein, which functions to attenuate NF- κ B signaling in a multitude of immune cell populations and contexts. Here, we study 2,781 genetic variants in the *TNFAIP3* locus, which contains multiple independent haplotypes associated with at least 8 autoimmune diseases. In order to (1) nominate causal variants that could underlie these diverse autoimmune disorders and to (2) gain a more comprehensive understanding of the regulatory architecture of this important locus, we dissect regulatory variant functionality using massively parallel reporter assays, create maps of the regulatory landscape with ATAC-seq and Hi-C, and probe the functional relevance of identified regulatory elements and variants using complementary CRISPRi and CRISPRa high throughput screens across a diverse spectrum of immune cell lines under relevant stimulation conditions. Combining all evidence, we nominate several putatively functional regulatory variants that are within linkage disequilibrium blocks for genetic association with inflammatory bowel disease, systemic lupus erythematosus, and psoriasis. Several of these variants are occupied by and predicted to disrupt the binding of key immune transcription factors. Overall, our comprehensive experimental approach is applicable to the study of disease associated common genetic variants and may be used to refine the search for causal variants in a scalable manner.

1717T

Predicting and characterizing kidney transplant rejection risk by B cell immune repertoire sequencing. S. Pineda^{1,2}, T. Sigdel¹, J. Liberto², M. Sarwal¹, M. Sirota¹. 1) Institute for Computational Health Science, UCSF, San Francisco, CA, USA; 2) Department of Surgery, UCSF, San Francisco, CA, USA.

The diversity of the B cell immune repertoire (IR) may drive humoral injury and rejection in organ transplantation (tx). We studied functionally correlated variations in the VDJ-CDR3 regions by B cell sequencing, examining 124 DNA and RNA samples from the same tx patient before tx, and at 6 and 24 mo post-tx. Three patient groups sorted into non-progressors (NP), progressors/ no rejection (PNR) and progressors/ with rejection (PR) were scored by progressive changes in the chronic allograft damage index (CADI). Pre-tx IR diversity was significantly greater in patients that went on to reject their tx (PR vs NP, $p = 0.008$). Post-tx, there was a strong interaction between clinical outcome (CADI score, AR) and time, with a significant reduction in clonal diversity ($p = 0.01$) in PR, and an increase in NP. We defined a network for each sample considering the clone definition (same V and J segments, same CDR3 length and 90% nucleotide identity between CDR3s). We used Gini index for quantification and found relevant changes at 24 mo post tx (PR vs. NP, $p=0.003$ and $p=0.001$). On a further evaluation of the data, we also found specific clones and an over-expression of several IGHV genes in the rejection group, which may suggest a biological driven mechanism for rejection. The diversity of recipient IR plays a major role in driving AR and progressive chronic tissue injury, and this effect can be observed even prior to engraftment of the tx organ. Selected dominant clones expand in patients who reject allografts, and the preservation of these dominant clones results in a reduction of their diversity. Pre-tx prediction of recipient risk of rejection, provides a powerful approach for precision medicine in tx.

1718F

Genetic and environmental factors associated with the diversity and biases of immune system. *W. Zhang^{1,2}, X. Liu^{1,2}, Z. Wan^{1,2}.* 1) BGI-Shenzhen, Shenzhen, 518083, China; 2) China National GeneBank-Shenzhen, BGI-Shenzhen, Shenzhen 518083, China.

The extremely large diversity of T-cell receptor (TCR) repertoire make up of individual immune system, acting against the external Pathogenic infection and enabling individual to keep healthy. The process of germline VDJ rearrangement, that comprise of VDJ gene recombination and random nucleotides deleted or inserted during the junction of V-D and D-J regions, contribute the huge TCR diversity and is regulated by both genetic factors and environment stimulation. The genetic influence are involved during the period of negative and positive selection of T cells in thymus and external antigen presentation. Additionally, auto or exogenous antigen stimulation causes some TCRs expansion. To explore how genetic factors and environment stimulation regulate the diversity of TCR repertoire, we collected more than 1000 volunteers and all were amplified TCR β repertoire and sequenced using high throughput sequencer. Simultaneously, multi-dimensional data were generated, including whole genome sequences (WGS), clinical tests, metabolomes, microbiomes, tumor screen, psychological test and so on. Using all of these data, we utilized bioinformatic method to identify genetic variations from WGS data associated with TCR β V genes and overall clone diversity, and found many associated variations in major histocompatibility complex (MHC) region. According to comparison of TCR repertoire and internal/external environmental data, we can find which factors are strong correlated with the diversity of TCRs. Our results clearly demonstrate the genetic environmental factors regulate the diversity of TCR repertoires, which provide clues for avoiding decreasing the capacity of immune system, and prevention and therapy of immune related disease.

1719W

Single-cell RNA-sequencing of a patient with a mosaic RUNX3 deletion reveals an aberrant T-cell repertoire in peripheral blood mononuclear cells. *J.H. Tsai¹, V. Montel¹, L.C. Watson¹, M. Kellogg¹, J. McEachern¹, B. Lajoie¹, M. Bainbridge¹, J. Friedman^{2,3,4}, R.J. Taft¹.* 1) Illumina, Inc, San Diego, CA; 2) UCSD Department of Neurosciences and Pediatrics, San Diego, CA; 3) Rady Children's Hospital Division of Neurology, San Diego, CA; 4) Rady Children's Institute for Genomic Medicine, San Diego, CA.

The application of whole exome and whole genome sequencing to patients with rare, undiagnosed and genetic diseases (RUGD) has led to substantial increases in diagnostic yield, but even amongst cohorts with refined selection criteria the number of patients resolved rarely exceeds 60%. Recent advances in single-cell RNA-sequencing (scRNA-seq) enable massively parallel analysis of transcriptomes across diverse cell types, which may help deconvolute previously complex genetics and improve disease understanding. Here, we applied scRNA-seq to a patient carrying a mosaic 68kb deletion in RUNX3 with progressive neurodegeneration and an unresolved clinical diagnosis. We performed scRNA-seq on isolated peripheral blood mononuclear cells (PBMCs) from the proband and both biological parents, followed by unbiased dimensionality reduction-based clustering analysis (tSNE) on at least 5000 cells to profile the PBMC subpopulation. Using canonical immune cell gene markers to annotate the subpopulations, the tSNE plot revealed a loss of a defined CD8 T-cell cluster in the proband. Comparative analysis of the proband with the biological mother showed that the subpopulation of cells with low CD8 gene expression also had low RUNX3 gene expression, consistent with the previous studies showing that RUNX3 is linked to CD8a expression. Using flow cytometry (FACS) to profile the peripheral immune cell surface protein expression, we found a higher CD4- CD8- population in the proband compared to the biological mother (35.6% versus 19.7%, respectively), but similar levels of CD8+ cells (24.4% versus 28.3%). Combined with the scRNA-seq data, we believe the CD4/CD8 canonical protein surface markers alone may not be sufficient to detect the subtle gene expression differences observed in the T-cell populations that are identified by scRNA-seq. To our knowledge, this is the first study to show that scRNA-seq may provide insight into the biological impact of a mosaic CNV. Further studies are required to understand the link between RUNX3, the immune cell phenotype and the proband's neurodegeneration.

1720T

Droplet assisted RNA targeting by single cell sequencing (DART-seq) enables simultaneous multiplexed amplicon sequencing and transcriptome profiling in single cells. *M.F.Z. Wang¹, M. Saikia^{1,2}, P. Burnham¹, S.H. Keshavjee¹, P. Moral-Lopez², M.M. Hinchman², C.G. Danko², J.S.L. Parker², I. De Vlaminck¹.* 1) Meinig School of Biomedical Engineering, Cornell University, Ithaca, NY 14853, USA; 2) Baker Institute for Animal Health, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA.

Background: High-throughput single-cell RNA sequencing (scRNA-seq) technology has provided important insights into cellular complexity and transcriptome dynamics. However, current implementations of this technology are limited to capturing information from the ends of A-tailed messenger RNA (mRNA) transcripts. Here, we describe a versatile technology, Droplet Assisted RNA Targeting by single cell sequencing (DART-seq), that surmounts this limitation allowing investigation of all regions of the polyadenylated transcriptome, as well as measurement of other classes of RNA in the cell. **Method:** Barcoded primer beads that capture the poly(A)-tailed mRNA molecules in Drop-seq are enzymatically modified using a tunable ligation chemistry to target specific RNA sequences. The resulting DART-seq primer beads are capable of priming reverse transcription of poly(A)-tailed transcripts as well as any other RNA species of interest. **Results:** We first applied DART-seq to simultaneously measure transcripts of the segmented dsRNA genome of a reovirus strain, and the transcriptome of the infected cell. DART-seq captured significantly more viral RNA in T3D reovirus-infected L cells, where the T3D mRNA lack poly(A)-tails, compared to the standard Drop-seq technique. Additionally, DART-seq increased the mean coverage of the S2 gene segment in T3D virus by 430-fold compared to Drop-seq leading to the observation of a high rate of G-to-A mutations in S2. We also found a significant increase in the proportion of viral transcripts in a subset of host cells with an upregulation of mitotic related genes demonstrating DART-seq's utility in studying the single cell heterogeneity of viral genotypes and cellular phenotypes during viral infections. In a second application, we used DART-seq to simultaneously measure natively paired, variable region heavy and light chain amplicons and the transcriptome of human B lymphocyte cells. DART-seq captured the paired heavy and light variable regions in B cells more frequently than standard Drop-seq without compromising transcriptome data for other cell types. **Conclusion:** DART-seq is an easy-to-implement, high-throughput scRNA-seq technology that overcomes the limitation of 3' end focused transcriptome measurements. We have demonstrated the versatility of DART-seq with distinct assays that probe host-virus interactions during RNA virus infection and that measure the endogenously paired antibody repertoire and the transcriptome of B cells.

1721F

Deciphering the cellular and molecular heterogeneity of the human lung airway epithelium in smokers and non-smokers with single cell transcriptomics. *M.A. Seibold^{1,2,3}, N.D. Jackson¹, S.P. Sajuthi¹, C. Rios¹, J.L. Everman¹, K.C. Goldfarbmuren¹.* 1) Center for Genes, Environment, and Health, National Jewish Health, Denver, CO, USA; 2) Department of Pediatrics, National Jewish Health, Denver, CO, USA; 3) Division of Pulmonary Sciences and Critical Care Medicine, Department of Medicine University of Colorado-AMC, Aurora, CO, USA.

Single cell transcriptomics is a valuable tool for inferring cell type-specific expression profiles that can be used to deconvolute the various cell phenotypes present in a system, such that the functional and genetic heterogeneity underlying these phenotypes can be accurately studied. The human proximal airway epithelium plays a vital role in respiratory health and homeostasis as well as in the pathogenesis of chronic pulmonary disease. In addition to cell types, the airway epithelium is composed of many distinct cell states whose identity, molecular phenotype, and relative prevalence in the airway remain poorly described. In this study, we characterize cellular heterogeneity in the tracheal epithelium of non-smoker and smoker donors using single cell RNA-seq of >10,000 cells. Examining these data, we find three states of basal airway stem cells, several states within two types of secretory cells, four states of ciliated cells, and various transitional cell states. We also reveal previously undescribed transcriptional signatures of rare cell types within the human airway epithelium, such as pulmonary neuroendocrine cells and ionocytes, which are likely critical to asthma and cystic fibrosis. Moreover, we characterize various immune cells that contribute to epithelial inflammation and immune responses present in many chronic lung diseases. We show that ciliated cell populations are both strongly re-programmed and proportionally reduced in smoker airways. Smoking also induces cell type-specific transcriptome shifts in other cell populations that alter airway function. In addition to functional characterization, we have developed panels of diagnostic cell surface biomarkers for described cell types and states that can be used for histological identification, isolation with flow cytometry, and deconvolution of cell type proportions in bulk RNA-seq datasets. These results provide functional information and molecular tools that will aid in our understanding and further study of the distinct and interacting transcriptional programs found within the proximal airway epithelium, both in health and in chronic lung disease.

1722W

Translational evaluation of a genome-wide CRISPR/Cas9 screen of acetaminophen-induced hepatocyte toxicity reveals mechanistic insights and druggable candidate genes. K. Shortt¹, D.P. Heruth¹, L.Q. Zhang², S.Q. Ye².

1) Children's Mercy Kansas City, Kansas City, MO; 2) University of Missouri Kansas City School of Medicine, Kansas City, MO.

Acetaminophen (APAP), a commonly used analgesic, is the leading cause of acute liver failure (ALF) in the United States and Great Britain. The drug is responsible for more than 56,000 overdose-related emergency room visits annually. APAP overdose has long asymptomatic period and limited treatment options, leading to a high rate of liver failure and frequently resulting in either organ transplant or mortality. Our aim is to identify new genetic risk factors to better understand the mechanisms of hepatotoxicity and to further identify gene and pathway candidates suitable for diagnostic and therapeutic targeting of APAP overdose and ALF. We previously described a genome-wide CRISPR/Cas9 screen, which evaluated genes that were protective against or increase susceptibility to APAP-induced liver injury. The HuH7 human hepatocellular carcinoma cell line containing CRISPR/Cas9 gene knockouts were treated with 15mM APAP for 30 minutes to 4 days to simulate partial to near total cellular death. Top candidate genes were identified based on their enriched knockout (pathogenic gene) or depleted knockout (protective gene) rankings ($p < 0.05$) and were further filtered by significant differential expression in gene expression data of various APAP exposures ($p < 0.05$). We now describe and extended assessment of the CRISPR/Cas9 screen data and gene expression data of 3 APAP-overdosed datasets of human patients and mice ($p < 0.05$), evaluation of genes containing 147 known APAP-associated polymorphisms, and evaluation of 48 genes involved in NAD metabolism. We additionally evaluate the therapeutic potential of 53 top candidate genes, both novel and known, identified by these approaches. This study has resulted in identification of numerous genes previously not linked to liver injury. Our translational analysis approach incorporates analysis of CRISPR/Cas9 screen data, gene expression data, and gene sets identified from the literature to build a more complete understanding of the genetics of APAP-induced ALF. We propose further evaluation of candidate genes for risk assessment, diagnostic, and therapeutic utility in clinical practice as well as continued functional studies to better characterize our findings. Collectively, this line of research highlights the power of genome-wide CRISPR/Cas9 screening technology to identify novel genes and pathways involved in APAP-induced hepatocyte toxicity and to provide potential targets to develop novel or repurpose existing therapeutics.

1723T

Transcriptome meta-analysis reveals dysregulated molecular signatures in cystic fibrosis patients with normal lung function. J.E. Ideozu^{1,2,3}, X. Zhang^{1,2,3}, V. Rangaraj¹, M. Kandpal³, M. Kaldunski⁴, A. Pan⁴, R. Davuluri³, P. Simpson⁴, M. Hessner⁴, H. Levy^{1,2,3}. 1) Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL; 2) Human Molecular Genetics Program, Stanley Manne Children's Research Institute; 3) Northwestern University, Chicago, IL; 4) Children's Hospital of Wisconsin, Medical College of Wisconsin, WI.

Introduction: In cystic fibrosis (CF), chronic lung infection with *Pseudomonas aeruginosa* (*Pa*) is the major cause of morbidity and mortality. Prior transcriptional profiling of peripheral blood mononuclear cells (PBMCs) suggested that impaired immune cell responses are a common feature in CF patients with lung disease. However, it remains unclear whether the immune system response is dysregulated at baseline, where there is normal lung function and no evidence of mucoid *Pa* lung infection. The identification of molecular targets dysregulated at baseline could hold therapeutic potential for treatment of CF. **Methods:** Gene expression dataset generated from PBMCs stimulated with plasma from CF and healthy control (HC) subjects were retrieved from the Gene Expression Omnibus repository (GSE71799). Sixteen CF subjects (mean age 7.48 ± 1.53 years) and 5 age-matched healthy controls (8.71 ± 1.78 years) were selected and analyzed to identify differentially expressed signatures at baseline. All CF subjects were F508del homozygotes, had forced expiratory volume in 1s $> 90\%$ predicted and were negative for mucoid *Pa* lung infection (± 6 months). Array data were normalized with Robust Multi-array normalization (RMA) and differentially expressed genes were identified with Partek GSA tool. Biological pathway and gene ontology enrichment were assessed from top differentially expressed genes ($FDR < 0.05$; $\log_2$ foldchange $> \pm 2$). Further in silico analysis using the gene expression data was performed to identify dysregulated immune cells. **Results:** We identified 1045 genes differentially expressed between CF and HC subjects ($FDR < 0.05$; $\log_2$ fold-change $> \pm 2$). Top dysregulated immune pathways (adjusted $p < 0.05$) include TREM1 signaling, agranulocyte adhesion and diapedesis, and communication between innate and adaptive immune cells. Upstream regulator analysis revealed downregulation of several cytokines ($p < 0.001$) including CCL5, IL1B, IL17A, and IL1A. The top molecular and cellular functions impaired ($P < 0.001$) were cell movement of myeloid cells, cell death and survival, cellular development, cellular growth and proliferation, and cell-to-cell signaling and interaction. Cell-type enrichment analysis highlighted dysregulation ($P < 0.05$) of several immune cell types including lymphocytes, monocytes, and dendritic cells. **Conclusion:** Our results suggest that CF subjects with normal lung function and without airway infection have dysregulation of their immune responsiveness at baseline.

1724F

Investigation of gut microbiome in ulcerative colitis and Crohn's disease patients. N.V. Ivanov¹, S.R. Abdulkhakov¹, T.V. Grigoryeva¹, A.A. Rizvanov¹, R.I. Faizullin¹, N.A. Danilova¹, A.R. Valeeva¹, S.Yu. Malanin¹, M.N. Siniagina¹, D.R. Khushnutdinova¹, I.I. Salafutdinov¹, A.V. Laikov¹, E.A. Boulygina¹, M.I. Markelova¹, V.V. Solovyova¹, L.G. Tazetdinova¹, D.S. Chulpanova¹, K.V. Kitayeva¹, Ye.E. Garanina¹, K.Ya. Ivanov¹, K.S. Kletenkov¹, A.R. Kayumov¹, G. Lo Sasso², C. Poussin², J.N.D. Battey², N. Sierrro², R.M. Gadaleta², A.P. Kiyasov¹, J. Hoeng², M.C. Peitsch². 1) Kazan Federal University, Kazan 420008, Russian Federation; 2) PMI R&D, Philip Morris Products S.A., Neuchatel, Neuchatel, Switzerland.

Ulcerative colitis (UC) and Crohn's disease (CD) are the two major phenotypes of Inflammatory Bowel Disease (IBD). IBD is a chronic gastrointestinal disorder characterized by periods of relapse and remission. UC and CD are often difficult to distinguish during diagnosis due to a number of common phenotypic and pathogenic features. Reliable, non-invasive biomarkers to aid diagnosis and prognosis of flares are not currently available. To fill this gap, new molecular and microbiome-based biomarkers need to be identified. Toward this end, we collected stool and blood serum samples from 40 patients diagnosed with UC, 40 patients diagnosed with CD, and 40 healthy control subjects. We investigated the gut microbiome by 16S RNA and shotgun sequencing in the three subject groups, and we measured the serum levels of (i) perinuclear anti-neutrophil cytoplasmic antibodies (pANCA) and anti-*Saccharomyces cerevisiae* antibodies (pASCA), (ii) C-reactive protein (CRP), and (iii) lipopolysaccharide (LPS). We also quantified the stool levels of calprotectin, lactoferrin, neopterin and S100A12. Finally, we evaluated the inflammation status of each subject's serum using Luminex multiplex profiling of cytokines and chemokines. Preliminary results reveal differences in microbiome profile of UC and CD subjects and other known biomarkers, which may contribute to the identification of novel molecular and microbiome-based signatures and approaches for the clinical management of IBD. .

1725W

Characterizing podocyte cell cultures and genetic markers of podocyte maturity. T.W. Soare, W. Zhang, A.D. Westerling-Bui, H. Hoang, A. Tebbe, L. Walsh, P. Mundel, T. Tibbitts. Goldfinch Bio, Cambridge, MA.

Background: Podocytes are terminally differentiated kidney cells that are vital to blood filtration. Podocyte injury and loss leads to excess protein in the urine, kidney scarring, and ultimately renal failure. Conditionally-immortalized differentiated podocytes provide an in vitro model to test hypotheses regarding mechanisms of disease and responses to treatment. However, podocytes in culture lack slit diaphragms (a morphological marker of maturity) and may lack expression of certain podocyte-specific proteins, creating doubts as to cultured podocyte maturation. **Objective:** To assess maturity of cultured podocytes, we characterized transcriptomic changes of podocyte differentiation at the single-cell resolution. **Methods:** Undifferentiated and differentiated mouse podocyte cultures were disassociated, fixed, suspended, and sequenced (Illumina HiSeq 26x40). To assess transcriptomic changes associated with differentiation and passage number, cells were clustered with principal component analysis (PCA) and visualized with t-Distributed Stochastic Neighbor Embedding (tSNE). Cell clusters were labeled by visual inspection of expression of published cell-type specific markers. We leveraged publicly-available scRNAseq data of mouse kidney to identify differentially expressed genes associated with podocyte maturation and to infer the trajectory of differentiation along the podocyte lineage. **Results:** We analyzed 1.8e9 reads in 17467 cells, identifying marker genes expressed in differentiated mouse podocytes. Differentiated podocytes in culture expressed markers of S-shaped bodies and early podocytes (*Col4a1*, *Col4a2*, *Col4a5*, *Pax2*) as well as markers of mature podocytes (*Synpo*, *Podxl*, *Nupr1*, *Plce1*, *Foxd1*). Other markers of mature podocytes were not expressed in cultured podocytes (*Wt1*, *Nphs1*, *Nphs2*, *Col4a3*, *Col4a4*). We also assess the functional importance of genes associated with podocyte differentiation by leveraging summary statistics from recent GWAS of kidney function and disease. **Conclusion:** Though differentiated podocytes in culture share transcriptomic signatures with mature podocytes in vivo, important differences remain. Future work will continue to differentiate mature mouse podocytes in culture as well as mature human podocytes in kidney organoids. Because podocytes are critical to healthy kidney function, a robust in vitro model of mature podocytes is vital to assessing mechanisms of kidney disease and responses to treatment.

1726T

Extracting expression-modulating genetic variants from Hidradenitis Suppurativa clinical trials using massively parallel reporter assays.

M. Liu¹, C. Lu², J. Degner², M. Hu³, M. Domanus³, A. Grundstad², K. Idler², J. Davis², J. Waring¹, C. Lu. 1) Pharmacogenetics and Human Genetics, AbbVie Inc., North Chicago, IL; 2) Computational Genomics, Genomics Research Center, AbbVie Inc., North Chicago, IL; 3) Genomic Technologies, Genomics Research Center, AbbVie Inc., North Chicago, IL.

With the development of large genetic biobank data, genome-wide association studies (GWAS) and whole-genome sequencing have identified numerous loci linked to human diseases and response to drug therapy. However, the correlation between nearby polymorphisms within individual associations often leaves tons of hundreds of potential causal variants to be interrogated. Additionally, disease-associated variants in noncoding regions have been estimated to explain a greater portion of the heritability for some disorders than variants in coding regions. This implies that it is absolutely crucial to have an efficient and systematic approach for identifying clinically actionable disease-causing genetic variants and target genes for drug development. In order to identify potentially functional single-nucleotide variants (SNVs) from a large pool of highly linked SNPs, we performed and optimized a high-throughput Massively Parallel Reporter Assay (MPRA) against nearly 30,000 SNVs in strong linkage disequilibrium from a previously published hidradenitis suppurativa (HS) GWAS (2016 ASHG poster abstract 1340T), aiming to detect subtle differences in how each allele drives expression. Specifically, this library contains 507 non-coding variants from GWAS ($p < 10^{-4}$) as well as more than 80 random sequences serving as negative controls to assess inter-experimental noise. We chose a primary human outer root sheath hair follicle cell line that is highly relevant to HS as a model cell line for pooled MPRA screening. Next-generation sequencing was performed using the Illumina NextSeq500. Out of 622 tested variants, MPRA analysis conservatively identified 17 potentially functional variants, which were then validated by individual luciferase reporter assays in a larger genomic context (~300 bp nucleotides with the SNP sequences, followed by the promoter of a nearby gene). Interestingly, of the top SNPs with allelic function identified by MPRA, none were the top SNPs (based on p-value) from the previously reported GWAS study. This result was confirmed by individual luciferase reporter assays. In summary, we were able to successfully employ MPRA to identify 17 potentially functional SNVs from a large number of candidate variants identified from previous GWAS results. Our study clearly highlights the limitation of relying on P values alone from GWAS studies and validates MPRA as a viable method to pinpoint causal variants in disease-relevant tissue.

1727F

Somatic mutations in focal cortical dysplasia. V.S. Almeida^{1,5}, S.H. Avansini^{1,5}, M. Borges^{1,5}, F. Rogerio^{2,5}, B.S. Carvalho^{3,5}, A.M. Canto^{1,5}, E. Ghizoni^{1,5}, H. Tedeschi^{1,5}, C.L. Yasuda^{4,5}, F.R. Torres^{1,5}, F. Cendes^{4,5}, I. Lopes-Cendes^{1,5}.

1) Department of Medical Genetics, School of Medical Sciences, University of Campinas (UNICAMP), Brazil; 2) Department of Anatomical Pathology, School of Medical Sciences, University of Campinas (UNICAMP), Brazil; 3) Department of Statistics, Institute of Mathematics, Statistics and Scientific Computing, University of Campinas (UNICAMP), Brazil; 4) Department of Neurology, School of Medical Sciences, University of Campinas (UNICAMP), Brazil; 5) Brazilian Institute of Neuroscience and Neurotechnology (BRANN), Campinas, SP, Brazil.

Introduction: Focal cortical dysplasia (FCD) is often associated with the occurrence of refractory seizures. FCD is characterized by alterations in the brain cytoarchitecture. Recently, mosaic mutations in genes of the mTOR pathway were detected in FCD; however, it is still unclear the proportion of patients with FCD who present somatic mutations. **Methods:** Deep sequencing of the mTOR pathway genes was performed on genomic DNA extracted from brain tissue resected by surgery (BTRS) and blood samples of 12 patients with FCD type II. Samples were sequenced following a 150bp paired-end protocol in a Miseq (Illumina), to achieve at least 600x of average coverage. We aligned sequences using BWA-MEM and performed realignment around SNPs and indels, quality recalibration and variant calling using the Genome Analysis Toolkit (GATK). We evaluated mosaicism using Mutect2. Variants were classified as mosaic mutations when less than 10% of reads are not aligned to the reference genomes and are present only in BTRS. Variants were filtered prioritizing frameshift, missense, nonsense and splicing site mutations that were localized in coding regions or exon-intron boundaries. In addition, we also focused on variants not described previously or variants whose minor allele frequency (MAF) is ≤ 0.01 . Effect of variants was evaluated using Variant Effect Predictor (VEP). **Results:** We identified somatic mutations in 8/12 patients (65%), including a previously reported somatic mutation c.4379T>C/p.Leu1460Pro (rs1057519779, NM_004958.3) with an allele frequency of 1.8% in the *MTOR* gene. In addition, we identified, in the same patient, a *TSC2* somatic mutation c.3781G>A/p.Ala1261Thr (not reported previously; NM_000548.3) with an allele frequency of 1.7% and an *AKT1* somatic mutation c.349_351del/ p.Glu117del (rs768025881, NM_005163.2) with allele frequency of 2.2%. **Conclusion:** Somatic mutations in genes of the mTOR pathway seems to be relatively common in patients with FCD type II since we found potentially deleterious variants in 65% of our patients. One of the patients studied had two somatic mutations in genes of the mTOR pathway (*TSC2* and *AKT1*) supporting the two hits hypothesis in FCD.

1728W

Cell type-specific responses in motor neuron disease: Analysis of the cell type-specific effects of C9orf72 repeat expansion in the motor cortex of mouse models of fALS at single cell resolution. L.A. Goff^{1,2,3}, K. Lee², J. Anderson², G. Stein-O'Brien^{1,2,4,5}, B. Winer¹, D. Heo², D. Bergles^{2,3}, S.P. Brown^{2,3}.

1) Institute for Genetic Medicine, Johns Hopkins University, Baltimore, MD; 2) Department of Neuroscience, Johns Hopkins University, Baltimore, MD; 3) Kavli Neurodiscovery Institute, JHU, Baltimore, MD; 4) Institute for Data Intensive Engineering and Science, Johns Hopkins University, Baltimore, MD; 5) Department of Oncology, Johns Hopkins University, Baltimore, MD.

Neurodegenerative disorders (ND) are characterized by neuronal loss often diverging based on the selective vulnerability of different neuron types. Amyotrophic Lateral Sclerosis (ALS) is a progressive and selective loss of motor neurons in the motor cortex or spinal cord. Several pathological features of ALS, including neuronal hyperexcitability, are observed in many different cortical cell types in mouse models. While bulk gene expression studies across models have revealed both distinct and common features among NDs, suggesting commonalities in the underlying pathological processes, these studies do not describe the distinct pathological responses of discrete cell types; a hallmark of selective vulnerability. Here we use single-cell RNA-Seq (scRNA-Seq) to examine the cell-type-specific transcriptional responses in the motor cortex of two different mouse models of C9orf72 repeat expansion, a causal variant for familial ALS. We have developed a suite of tools for analysis of patterns of co-regulated gene expression in scRNA-Seq. Using these tools, we can identify distinct and shared molecular pathological features across discrete cell types in the motor cortex. These patterns represent both common cellular responses to the pathogenic C9orf72 repeat expansion as well as responses that are unique to subpopulations of cells, and can be used to enhance understanding of the selective vulnerability to this pathogenic repeat. Finally, we use these data to relate cell type-specific gene expression to genetic analysis of disease association. The most commonly used tool for studying complex disease is the genome-wide association study (GWAS). Although GWAS is agnostic to the context in which disease-risk variation acts, the biological impact of common functional variation has been shown to be cell context dependent. We have established a framework to prioritize candidate genes across GWAS loci using cell type-specific gene expression and use this approach to refine the genetic architecture of ALS. This analysis exploits the relationship between context-specific gene expression and genetic association to both identify affected cell types, and select candidate genes within GWAS loci for each cellular context. Collectively, these data demonstrate the importance of cellular context for understanding the molecular pathology of disease and selective vulnerability of upper motor neurons in fALS, and provide a template for single cell-resolution analysis of complex NDs.

1729T

Generation and characterization of a conditional mouse model that uses SUMO1 to direct the BirA-mediated tagging of cell-type-specific nuclei *in vivo*. M. Hudson^{1,2}, X. Yuan^{1,2,3}, M. Liang^{1,2}, V. Gomes², A. Tasneem⁴, J.L. Lefebvre^{1,4}, M.D. Wilson^{1,2}. 1) Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada; 2) Genetics and Genome Biology Program, Hospital for Sick Children, Toronto, Ontario, Canada; 3) Developmental and Stem Cell Biology Program, Hospital for Sick Children, Toronto, Ontario, Canada; 4) Neuroscience and Mental Health Program, Hospital for Sick Children, Toronto, Ontario, Canada.

Knowing the roles each cell type plays within a given tissue at a given time is essential for understanding disease phenotypes. While genomic assays that capture gene expression, DNA methylation and open chromatin can now be deployed on single cells, there is a need for methods to rapidly isolate distinct cell populations from a tissue in a cost-effective manner. Recently, the powerful INTACT method (Isolation of Nuclei Tagged in specific Cell Types; Deal and Henikoff, 2011) was established for isolation of nuclei from genetically-labeled cell types and has now been applied to several model organisms including mice. Here we describe the generation and characterization of a new SUMO1-driven conditional INTACT mouse (CONTACT; CONditional Tagging of Cell-Type specific nuclei). This conditional mouse model expresses *SUMO1* fused to GFP and a biotin ligase recognition peptide that can be specifically biotinylated by the *E. coli* biotin ligase BirA. As SUMO1 is a cytoplasmic component of the nuclear pore complex, nuclei containing this construct can be rapidly purified from bulk populations using streptavidin-coated magnetic beads, similar to other INTACT methods. These nuclei can be used for genomic assays including ATAC-seq, ChIP-seq, or RNA-seq. A novel feature of our model is that it permits the detection of cell-type-specific SUMO1-modified proteins *in vivo*. Here I will present results using our SUMO1 CONTACT mouse to profile Purkinje cells, which are the sole output from the cerebellar cortex where motor signaling takes place. Purkinje cells have been implicated in motor neuron and neurodegenerative diseases and comprise ~0.03% of cells in the cerebellum. I will compare gene expression, open chromatin, and histone modifications of Purkinje cells (using the Purkinje-cell-specific *Pcp2-Cre*) to bulk cerebellar tissue at two postnatal ages. In summary, starting with Purkinje cells, this transgenic mouse tool will enable detailed investigations of cell-type-specific gene expression, epigenetic modifications, protein expression and post-translational modification.

1730F

High efficiency neurological panel design using Twist Bioscience double-stranded custom targeted enrichment probes and capture optimization for aged dry blood spot samples. . R. Pellegrino¹, C. Kaminski Vaccaro¹, J. Garifallou¹, F. Mafrà¹, J. Smiler¹, M. Gonzalez¹, L. Tian¹, F. Wang¹, M. Consugar², L. Kulikowski³, A. Dias³, H. Hakonarson¹. 1) The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Twist Biosciences; 3) FMUSP.

Introduction: Targeted gene sequencing panels are cost-effective and valuable tools for understanding the causative novel or inherited mutations in a specific disease. Custom targeted sequencing is ideal for exploring genes in specific pathways as for follow-up experiments larger studies such as from whole-exome and whole-genome sequencing. A critical component for successful sequencing analyses involves the design capability and the quality of the capture reaction, which drives the uniformity of the reads. A more uniform capture leads to more uniform sequencing (higher targets coverage, limiting wasted reads). Twist Bioscience's capture efficiency and coverage uniformity is the result of the probe design algorithm, dsDNA probes, and probe boosting based on GC composition. The oligo synthesis platform and amplification methodology yields highly uniform dsDNA capture probes that result in greater targeted enrichment capture and sequencing efficiencies. This innovation significantly permits the capture of both strands of the target, resulting in an increase in the number of unique molecules sequenced and allowing for detection of strand-specific mutations such as deamination events in ancient samples or FFPE specimens. **Methods:** CAG with the support of Twist, we created a curated design containing ~1120 probes covering 55 genes linked to Epilepsy. We generated gDNA libraries using the Kapa Hyper preparation kit and capture libraries using IDT universal adapters and Twist capture probes and hybridization reagents. Additionally we evaluated Twist's beta Custom Targeted Enrichment Core Kit in parallel for the same samples. Genomic DNA was obtained from 46 dry blood spots from very longstanding collections and presented very challenging QC for DNA concentration. **Results:** Capture libraries were successfully generated according to CAG's quality standards/metrics for all the samples in question. Libraries were sequenced to ~200X average depth of coverage on a HiSeq 2500. Capture data demonstrated good coverage, high uniformity and even distribution across the genes of interest. The predicted cost savings associated with Twist's greater dsDNA oligo probe's capture efficiency and resulting sequencing uniformity will markedly increase the attractiveness of future custom targeted enrichment projects at CAG.

1731W

Clinical and genetic implications for long-read sequencing after testing the ONT MinION and PacBio Sequel across the C9orf72 'GGGGCC' repeat expansion in human carriers. J. Fryer¹, S. Farrugia¹, J. Sens¹, K. Jansen-West¹, I. McLaughlin², B. Bowman², M. Seetin², M. van Blitterswijk¹, R. Rademakers¹, L. Petrucelli¹, M. Ebbert¹. 1) Neuroscience, Mayo Clinic, Jacksonville, FL; 2) Pacific Biosciences, Menlo Park, CA, USA.

Background. Structural variants, and repeat expansions specifically, are increasingly recognized for their role in human disease. Repeat expansions are a nucleotide sequence insertion (e.g., 'GGGGCC'; G_nC₂) that is repeated many more times than normal. They are known to cause many CNS diseases including Huntington's disease, spinocerebellar ataxias (SCA), frontotemporal dementia (FTD), and amyotrophic lateral sclerosis (ALS). A repeat expansion in ABCA7 associated with Alzheimer's disease was also recently reported. As researchers continue to investigate repeat expansions, it is important to know how well long-read sequencing technologies such as PacBio and Oxford Nanopore Technologies (ONT) perform when sequencing across repeats. Many repeat expansions are GC-rich, making them difficult to sequence. We previously presented data assessing how well PacBio and ONT sequencing platforms sequenced through repeat expansions cloned into plasmids for SCA36 (GGCCTG) and ALS/FTD (G_nC₂). We also presented preliminary data from whole-genome sequencing across the C9orf72 G_nC₂ repeat expansion in a human carrier. Since then, we have refined our analyses for the whole-genome data, and we performed targeted sequencing across the C9orf72 repeat expansion using PacBio's new no-amplification (No-Amp) targeted sequencing method, based on CRISPR-Cas9. **Methods.** We identified two C9orf72 G_nC₂ expansion carriers, extracted HMW DNA from cerebellar tissue, and performed whole-genome sequencing on the MinION (15 flowcells) and Sequel (5 SMRT-cells), and targeted sequencing on the Sequel. **Results.** Whole-genome data from the MinION and Sequel correctly estimated 8 repeats for the unexpanded allele. Throughput was a challenge for the MinION, but newer chemistries and hardware (e.g., PromethION) from ONT are likely to address this issue. The Sequel was able to sequence through the entire repeat at 1324 repeats (7941 bases). No-Amp sequencing on the Sequel generated 828 CCS reads across both the unexpanded and expanded alleles. We obtained a distribution of reads (mosaicism) that spanned the repeat and were able to accurately estimate the expansion size. We were also able to assess G_nC₂ content in this individual. We cannot rule out repeat interruptions. **Conclusion.** Our results suggest ONT and PacBio long-read sequencing technologies are capable of traversing challenging repeat expansions. Both present exciting opportunities in clinical, genetic counseling, and genetic discovery efforts.

1732T

Humanized STXBP1 animal model for detecting pathogenicity in clinical variants. C. Hopkins, G. Hu, K. McCormick, T. Brock. Nemamatrix, Salt Lake City, UT.

There is growing demand for functional data on clinical variants. Genomic data reveals large numbers of clinically-observed gene variants remain not specified or designated as variants of uncertain significance (VUS). High quality functional assessments (PS3 and BS3) are need. 20 genes ranked for highest submitters via ClinVar data (<https://clinvarminer.genetics.utah.edu>) were used in a dbSNP database query (<https://www.ncbi.nlm.nih.gov/variation/view>), which indicated average distribution between variant categories is 12.49% pathogenic, 2.94% likely pathogenic, 22.11% VUS, 7.54% likely benign, 0.69% benign, and 54.23% not provided/specified. Uncertainty increases as more genes are examined. Humanized (ClinPhen) animal models are created when a human cDNA is inserted a replacement for the endogenous coding sequence and the cDNA sequence exhibits capacity to restore normal function. As example, cDNA sequence for early infantile epileptic encephalopathy 4 gene (STXBP1) was recoded for optimal expression in *C. elegans*. CRISPR-based editing inserted hSTXBP1 sequence as gene replacement of the coding sequence in the unc-18 locus. Rescue of activity observed at 100% by electrophysiology (up 4x from unc-18 KO) and 50% in a chemotaxis assay (up from 0% in the unc-18 KO). Two known pathological variants, R292H and R406H, inserted into the humanized hSTXBP1 backbone exhibit electrophysiology and chemotaxis deficiencies. Because all of the known pathogenic variants in STXBP1 are autosomal dominant, heterozygotes were examined for deficiencies in a modified chemotaxis assay. 5 known pathogenic variants all showed chemotaxis deficiencies as heterozygotes. One of the variants, p.R388X, creating a two-thirds truncation of STXBP1 protein, presented with a dominant negative phenotype in the haploinsufficiency assay - approx. 3 fold more severe than heterozygote with full gene KO. The hSTXBP1 ClinPhen platform is being extended to analyze all of the gene's known variants. Drug screening is planned for all variant systems exhibiting activity defects consistent with pathogenicity in hope of finding new or existing AEDs with likeliness to be therapeutic for the specific variant patient. ClinPhen is being expanded to address variant profiles in KCNQ2, ATP1A3, LMNA and MSH2 genes. Approximately 2000 Rare Disease genes are expected to be addressable when converted into a ClinPhen system..

1733F

10x Genomics Chromium linked-reads for genotyping short tandem repeats in whole-genome sequence data. I. Rajan-Babu¹, R. Chiu², I. Biro², J.M. Friedman¹. 1) Department of Medical Genetics, University of British Columbia, and Children's & Women's Hospital, Vancouver, British Columbia, Canada; 2) Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver, British Columbia, Canada.

Background: 10x Genomics Chromium linked-read (LR) technology outperforms short-read sequencing in resolving complex structural variants; however, its utility in genotyping short tandem repeats (STRs) has not been examined. STRs are tandem arrays of repeat motifs composed of 1 to 6 base pairs (bp), and repeat expansions (increase in the number of copies of repeat units) at ~40 loci are associated with disease. We assessed the ability of 10x Genomics Chromium LR technology to characterize the *frataxin* (*FXN* [MIM 606829]) GAA and the *fragile X mental retardation 1* (*FMR1* [MIM 309550]) CGG repeat expansions that cause Friedreich ataxia (FRDA [MIM 229300]) and fragile X syndrome (FXS [MIM 300624]), respectively. **Methods:** Four FRDA and six FXS reference lymphoblastoid cell lines were obtained from the Coriell Cell Repository. High molecular weight genomic DNA specimens derived from the cell lines were used for Chromium library preparation. Paired-end whole-genome sequence (WGS) reads of 150-bp were generated on an Illumina HiSeq X instrument and processed using the Long Ranger™ pipeline. Seventeen disease-associated STRs, including the *FXN* GAA and *FMR1* CGG repeats, were genotyped in each sample using ExpansionHunter. **Results:** ExpansionHunter generated genotype calls for nearly all STRs, with repeat sizes that were accurate for the *FXN* normal alleles and in the normal range for most other STR loci tested. The heterozygous and homozygous expanded *FXN* alleles in the FRDA samples were clearly differentiated, but their repeat sizes, which actually ranged from ~650 to ~1030 GAA, were genotyped as 54 to 73 GAA. No genotype was obtained for the *FMR1* CGG repeat locus whether it was normal or expanded in eight of the 10 samples tested. Our inability to genotype the *FMR1* CGG repeats reflects very low or no sequence coverage of this region, in contrast to adequate coverage at the other STR loci genotyped in these Chromium LR WGS data. **Summary:** 10x Genomics Chromium LR WGS is a promising tool that has the potential to evolve into the "single test" for detecting all classes of genomic variants, but resolving issues related to poor coverage of GC-rich regions is critical to improve diagnostic sensitivity. Additional investigations of "difficult-to-genotype" STRs and of other analytical alternatives to ExpansionHunter for accurate repeat size estimation of full expansions in 10x Genomics Chromium LR WGS data are underway.

1734W

Whole genome sequencing of amyotrophic lateral sclerosis discordant monozygotic twins identifies thousands of false positive somatic mutations. K.L. Williams¹, E.P. McCann¹, N.A. Twine², D.B. Rowe³, G.A. Nicholson⁴, D.C. Bauer², I.P. Blair¹. 1) Centre for Motor Neuron Disease Research, Faculty of Medicine and Health Sciences, Macquarie University, NSW, Australia; 2) Commonwealth Scientific and Industrial Research Organisation, Sydney, NSW, Australia; 3) Department of Clinical Medicine, Faculty of Medicine and Health Sciences, Macquarie University, NSW, Australia; 4) Molecular Medicine Laboratory, Concord Repatriation General Hospital, Sydney, NSW, Australia.

ALS is a fatal late-onset disease characterised by rapid progressive degeneration of upper and lower motor neurons. Sporadic and hereditary forms of ALS are clinically and pathologically indistinguishable, yet substantial variation is seen in age of onset and rate of disease progression, even among patients with identical causal mutations. This suggests a role for modifying factors influencing disease onset and progression. Disease-discordant identical twins provide a powerful study design to uncover modifiers of disease, as the unaffected co-twin is the best-matched control available for their affected co-twin, reducing known confounding factors (such as age, sex) and unknown confounding factors. This co-twin study investigated the genomes of n=3 monozygotic twin sets and n=1 MZ triplet set discordant for ALS. One MZ twin set both harbor a pathogenic C9orf72 hexanucleotide repeat expansion and the MZ triplet set all carry a pathogenic SOD1 mutation. Whole-genome sequencing of the 9 twins/triplets identified 20902, 23946, 35332 and 7380 discordant variants between the affected twin/triplet and their unaffected co-twin/triplets thus being putative modifier variants. This large number of discordant variants was unanticipated and reflects false positive variants. Validation 1 using any discordant variants present on a SNP microarray identified all selected variants as concordant homozygous within twin/triplet sets. Validation 2 involved variant filtering (private genotype, private location, quality scores, MAF, population controls) to reduce the number of discordant variants to 167, 269, 295 and 6 respectively. Sanger sequencing of selected filtered variants confirmed all as concordant between twin/triplet sets. Validation 3 repeated WGS of one twin set at a second sequencing service provider using the same library prep method and sequencing instrument as the original sequencing service provider. Original and new WGS data for this twin set were processed through the same bioinformatic pipeline to identify discordant variants. No discordant variants overlapped between the different service provider sequencing results, hence confirming WGS-identified discordant variants as false positive results. Therefore, somatic *de novo* mutations are not causing ALS. This study highlights the necessity to either improve alignment and variant calling algorithms to significantly reduce the number of false positives, or to perform all WGS experiments in duplicate or triplicate.

1735T

Using Agilent SureSelect^{XT} Human Methyl-Seq to further study gene regulation through methylation modifications. C. Vaccaro¹, J. Smiler¹, L. Tian¹, D. Li¹, E. Bhoj¹, S. Sheppard¹, M. Filipek², H. Hakonarson¹, R. Pellegrino¹. 1) The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Agilent Technologies, Santa Clara, CA.

Introduction: In the era of next generation sequencing, it is not only important to examine mutations within genes, but to also focus on regulatory elements. DNA methylation is one form of epigenetic regulation. Some regions within DNA that are prone to methylation include CpGs, promoters, and differentially methylated regions (DMRs). Agilent has released an 84 Mb design that targets these methylated regions in their SureSelect^{XT} Human Methyl-Seq protocol. The SureSelect^{XT} Human Methyl-Seq protocol was constructed to capture both the design of SureSelect target enrichment and bisulfite treatment to target methylated regions. Its single base pair resolution and detection of individual CpGs make SureSelect^{XT} Human Methyl-Seq an attractive alternative to microarrays (only target about 5-10%). More specifically, SureSelect^{XT} Human Methyl-Seq covers 3.7 million CpGs (~91% of UCSC annotated CpGs) including CpG islands, shores, shelves, as well as detecting methylation in undermethylated regions, GENCODE promoters (~141,000), DNaseI hypersensitive sites, and cancer tissue-specific DMRs (~23,000). Additional improvements of the SureSelect^{XT} Human Methyl-Seq in comparison to microarrays are increased throughput, reducing sequencing costs, an increase in library complexity, and a reduction in bias. **Materials/Methods:** Our study consists of nine samples processed with the SureSelect^{XT} Human Methyl-Seq. Amongst this cohort, there were patients with *de novo* heterozygous mutations in *KMT5B*, a histone methyltransferase that is associated with a neurodevelopmental disorder, and age/sex matched controls. The SureSelect^{XT} Human Methyl-Seq required DNA shearing via sonication on the Covaris LE220, library preparation (end repair, a-tailing, and ligation), hybridization with the 84 Mb Human Methyl-Seq, bisulfite treatment, PCR, and sequencing on the Illumina HiSeq2500. **Results:** All samples passed sequencing QC. On average, 158 million reads were produced per sample. All samples exhibited a Q30 score of 88% and cluster density of 1041K/mm². **Conclusion:** The SureSelect^{XT} Human Methyl-Seq protocol offers a new alternative to studying gene regulation in areas that are modified by methylation. This protocol from Agilent will give a more cost-effective method for labs and increase their data at the single base pair resolution.

1736F

Exome analyses in subfamily trios from large family tree in the South-Eastern Moravia (Czech Republic) population with high incidence of parkinsonism. R. Vodicka¹, R. Vrtel¹, K. Kolarikova¹, M. Prochazka¹, K. Mensikova², P. Kanovsky². 1) Institute of Medical Genetics, University Hospital and Palacky University Olomouc, Olomouc, Czech Republic; 2) Department of Neurology Faculty of Medicine and Dentistry, Palacky University, University Hospital, Olomouc, Czech Republic.

There has been previously described higher prevalence of parkinsonism in small isolated region with well assembled 11 generation large pedigree from the South-Eastern Moravia. Aim of the study was to analyze whole exomes in selected subfamily trios from the pedigree. We used NGS Ion AmpliSeq Exome method (IonTorrent) for two (A and B) subfamily trios. Each trio comprised of two affected and one healthy person. DNA exome libraries were sequenced on IonPI chips. Variants were predicted using Torrent Suite and IonReporter softwares. Aligned reads (BAM files) were then analyzed using IonReporter Whole Exome Trio workflow. Final filtering was done with respect to population frequency, variant effects and with respect to the presence of variants in Parkinsonism disease responsible genes. Last filter was done with respect to the segregation of the disease. Almost whole exome was sequenced with coverage 1-20 and 90 % of exome was covered more than 20x in all the samples. Together more than 70.000 variants with average base coverage depth 75 were analyzable in both trios before filtering. After filtering there were found 99 and 96 variants in trio A and B respective. The most potentially associating variants with Parkinsonism are following: Trio A: SLC18A2:p.Gly195Ser;c.583G>A DRD1:p.Ala353Val;c.1058C>T AP2A2:p.Asn401Ser;c.1202A>G CCDC88C:p.Leu1696Pro;c.5087T>C ZFHX3:p.Met2102Thr;c.6305T>C ARAP2:p.Pro159Ser;c.475C>T CYP4F11:p.Trp29Ser;c.86G>C MRPS15:p.Thr252Ile;c.755C>T MRPS28:p.Arg48Pro;c.143G>C PRELID2:p.Val62Met;c.184G>A FAM171A1:p.Ser844Leu;c.2531C>T CAPRIN2:p.Arg373His;c.1118G>A FAM186B:p.Ala727Val;c.2180C>T CROT:p.Glu118Asp;c.354A>C MPDZ:p.Cys119Ser;c.356G>C Trio B: TENM4:p.Asn965Ser;c.2894A>G MON2:p.Gln531Arg;c.1592A>G MTCL1:p.Ala482Val;c.1445C>T NEPRO:p.Val297Ile;c.889G>A FAM131A:p.Leu280Val;c.838C>G ADH1C:p.Arg48His;c.143G>A SYNE1:p.Lys3729Asn;c.11187G>T RXFP2:p.Thr222Pro;c.664A>C AKAP11:p.Ile183Met;c.549A>G ZNF19:p.Pro216Ser;c.646C>T LRRK2:p.Arg1514Gln;c.4541G>A OSBPL1A:c.115_116insAATT SACS:p.Met1359Thr;c.4076T>C ZFHX3:p.Met2102Thr;c.6305T>C COL18A1:p.Ala1381Thr;c.4141G>A Detailed whole exome analyses in genetic isolated parkinsonism patients could contribute to further understanding of molecular-genetic mechanism and background of the disease. This study was supported by MH CZ – DRO (FNOI, 00098892).

1737W

Single-cell RNA-sequencing identifies multiple human microglial subtypes associated with neurodegenerative disease. V. Menon¹, M. Olah¹, M. Taga¹, C. Yung¹, R. Sarkis¹, J. Schneider², D. Bennett², W. Elyaman¹, E. Bradshaw¹, P. De Jager¹. 1) Neurology, Columbia University Medical Center, New York, NY; 2) Alzheimer's Disease Center, Rush University, Chicago, IL.

Our understanding of the biology of human microglia is advancing rapidly as modern transcriptome-wide assessments – along with proteomic and epigenomic studies of bulk microglia – have provided new insights into the role of this cell type in brain development, central nervous system tissue homeostasis and neurodegenerative diseases. However, our understanding of the diversity of cell states that microglia can adopt in the human brain remains limited; microglia are highly plastic and have multiple different roles, making the extent of intra-individual heterogeneity a central question as we begin to consider the development of therapies targeting this cell type. Here, we investigated the population structure of human microglia by single cell RNA-sequencing. Using surgical and autopsy-derived cortical samples, we isolated >15,000 cells and identified 14 distinct human microglial populations. We note substantial intra- and inter-individual heterogeneity, suggesting that microglial diversity in humans may be affected by both genetic background and disease state. The overall heterogeneity comprises both graded variation in a large, putatively homeostatic population as well as discrete subtypes that are associated with proliferative and neurological disease states. These microglial subclasses are distinguished by distinct transcription factor and cell surface combinations and display divergent associations with Alzheimer's pathology, dementia and multiple sclerosis. Upon comparison with previous studies, we find similarities with disease-associated microglial states reported in mouse, but the greater diversity of states in human suggests a more varied role of microglia in the brain.

1738T

Single-cell RNA-sequencing in facioscapulohumeral muscular dystrophy disease etiology and development. A. van den Heuvel¹, A. Mahfouz^{2,6}, S.L. Kloet¹, J. Balog¹, B.G.M. van Engelen³, R. Tawil⁴, S.J. Tapscott⁵, S.M. van der Maarel¹. 1) Human Genetics, Leiden University Medical Center, Leiden, Zuid-Holland, Netherlands; 2) Leiden Computational Biology Center, Leiden University Medical Center, Leiden, Zuid Holland, Netherlands; 3) Department of Neurology, Donders Institute for Brain Cognition and Behaviour, Radboud University Medical Center, Nijmegen, Gelderland, the Netherlands; 4) Department of Neurology, University of Rochester, Rochester, NY 14642 USA; 5) Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, WA 98109 USA; 6) Bioinformatics Lab, Delft University of Technology, Delft, Zuid-Holland, Netherlands.

Facioscapulohumeral muscular dystrophy (FSHD) is one of the most common inherited muscle diseases worldwide, with patients suffering from progressive wasting and weakness of the muscles in the face and upper extremities. FSHD has a unique etiology being caused by the de-repression of the cleavage-stage transcription factor DUX4 in only a small number of myonuclei of the affected muscle. This activates a cascade of muscle disrupting events ultimately leading to muscle atrophy and apoptosis. Yet, with approximately 1:500 nuclei expressing DUX4 in primary myotube cultures, how such sporadic expression leads to the severe muscle wasting observed in FSHD patients remains unclear. With this, essential insights in the order of- and interaction between the cascade of events implicated in FSHD are currently limited. Transcriptome analyses in FSHD muscle cell cultures or biopsies have systematically been challenged by the sporadic nature of DUX4 expression, with the majority of non-affected cells masking the FSHD transcriptome signature. Consequently, the analysis of the FSHD-transcriptome has thus far largely been facilitated by DUX4-expression or DUX4-reporter systems, limiting the analyses to DUX4-associated effects, and leaving non-cell-autonomous effects underrepresented in the FSHD transcriptome. We used single-cell RNA-sequencing, combined with pseudotime trajectory modeling, to study FSHD etiology and cellular progression. We show that we can identify a small FSHD-specific cell population in all tested patient-derived cultures, and we detect new expression signatures that are likely related to both direct and indirect effects of DUX4. Furthermore, capitalizing on the cellular heterogeneity in FSHD cell cultures, we performed pseudotime trajectory modeling to generate a detailed FSHD cellular progression model. We show that our model faithfully reflects both the early burst-like DUX4 expression, as well as the programmed downstream activation of various pathways previously implicated in FSHD. Altogether, our data highlights the potential for single-cell transcriptomics in studying disease etiology and cellular progression of genetic diseases, even when hallmarked by cellular heterogeneity as strong as in FSHD. We expect that pseudotime trajectories like our FSHD pseudotime model holds valuable information not only for further unraveling the disease etiology and progression, but also for biomarker and target selection guiding future therapy development.

1739F

Amplification-free targeted SMRT sequencing using CRISPR/Cas9 for studying repeat expansions in Huntington's disease. I. Höjjer¹, Y. Tsai², I. Jonasson¹, E-L. Stattin¹, M-L. Bondeson¹, T.A. Clark², U. Gyllenstein¹, A. Ameur¹. 1) SciLifeLab, Dept. of immunology, genetics and pathology, Uppsala University, Uppsala, Uppsala, Sweden; 2) Pacific Biosciences, Menlo Park, CA, USA.

Amplification of DNA is required as a mandatory step during library preparation in most targeted sequencing protocols. This can be a critical limitation when targeting regions that are highly repetitive or with extreme guanine-cytosine (GC) content, including repeat expansions associated with human disease. Here we used an amplification-free protocol for targeted enrichment utilizing the CRISPR/Cas9 system (No-Amp Targeted sequencing) in combination with Single Molecule, Real-Time (SMRT) sequencing for studying repeat elements in the huntingtin (*HTT*) gene, where an expanded CAG repeat is causative for Huntington's disease (HD). A guide RNA was designed in proximity to the *HTT* CAG repeat and was used to target and digest whole genome PacBio libraries (SMRTbells). Capture adapters were subsequently ligated to the digestion site and magnetic beads used to enrich for SMRTbells containing the targeted sequence. We applied the No-Amp method to eleven HD patient samples and found that the CAG repeat count from our method agreed with previous results based on fragment analysis for all 22 alleles. Furthermore, polymorphic CCG repeats in the flanking regions of the CAG repeats could be analyzed in detail. We also observed somatic variability of the CAG repeats, with greater variability seen in the expanded alleles. This is a known phenomenon but has been difficult to investigate using PCR-based protocols. In addition to on-target reads, off-target effects by the CRISPR/Cas9 could be identified in sequences with high homology to the guide RNA. The No-Amp protocol allows multiplexing of targets and barcoding of samples. It also enables detection of epigenetic modifications in the data. In conclusion, with the No-Amp Targeted sequencing method, we are able to accurately sequence through repeat expansions that are difficult or even impossible to investigate using PCR-based targeted sequencing protocols. .

1740W

Investigating changes to the gut microbiome before and during symptom progression in the MeCP2-e1 mouse model of Rett syndrome. T.E. Grant¹, T.S. Totah¹, A. Vogel Ciernia⁴, B.P. Durbin-Johnson³, D.H. Yasui¹, M.R. Rolston¹, M. Settles³, J.M. LaSalle². 1) Medical Microbiology and Immunology, University of California, Davis, Davis, CA; 2) Department of Medical Microbiology and Immunology, Genome Center, MIND Institute, UC Davis School of Medicine, CA, 95616, USA; 3) Division of Biostatistics, Department of Public Health Sciences, UC Davis School of Medicine, Sacramento, CA, 95817, USA; 4) Department of Medical Microbiology and Immunology, MIND Institute, UC Davis, CA School of Medicine, 95616, USA; 5) Genome Center, University of California, Davis, CA, 95616, USA.

Rett syndrome (RTT) is an X-linked dominant neurodevelopmental disorder caused by mutations in the MECP2 gene encoding methyl CpG binding protein 2 (MeCP2), an abundant neuronal chromatin factor. Girls with RTT have typical early development, followed by regression in developmental milestones and the progression of gastrointestinal symptoms and oxidative stress. While most RTT research has focused on the role of MeCP2 in the central nervous system, MeCP2 is expressed in all cell; including microglia implicated in neuroimmune interactions in RTT. Microglial maturation is influenced by the gut microbiome, and the gastrointestinal RTT symptoms such as constipation suggest that MeCP2 mutation may interact with the gut microbiome and enteric nervous system during symptom progression. To examine this relationship, we utilized a mouse model with an RTT targeted mutation of *MeCP2* exon 1 (*MeCP2-e1*). *MeCP2-e1* deficient male and female mice develop progressive motor impairments, behavioral changes, and metabolic symptoms similar to those observed in RTT patients. Therefore, we hypothesized that *MeCP2-e1* mutation may contribute to a perturbed gut microbiome, potentially exacerbating the severity and progression of the neurological symptoms observed in RTT. To test this hypothesis, we performed weekly timed collections of fecal pellets, disease progression scoring, and body weights from 4 to 20 weeks in *MeCP2-e1* deficient and control littermate mice of both sexes. DNA was isolated from fecal pellets and amplified using 16S ribosomal primers to make sequencing libraries for the Illumina MiSeq. Bioinformatic and statistical analyses (FDR<0.05) demonstrated several microbial genera significantly altered by time and genotype, with mutant males showing more significant differences than females. Interestingly, significant microbial differences were also observed before the progression of hindlimb clasping or weight gain. *Bifidobacteria* were consistently lower in *MeCP2-e1* mutant males and females across time. Furthermore, several butyrate producing *Clostridia* class members were significantly elevated at early time points in *MeCP2-e1* mutant males, suggesting a potential interaction of the histone deacetylase inhibiting the action of butyrate with increased histone acetylation levels observed in RTT brain. Future work will examine the effects of probiotic therapy using *Bifidobacteria*.

1741T

An integrated deep mutational scanning approach to defining the *PTEN* genotype-phenotype map. T.L. Mighell^{1,2}, J.L. Mester³, K.S. Hruska³, K.A. Matreyek⁴, D.M. Fowler^{4,5,6}, B.J. O'Roak². 1) Neuroscience Graduate Program, Oregon Health & Science University, Portland, OR; 2) Department of Molecular & Medical Genetics, Oregon Health & Science University, Portland, OR; 3) GeneDx, Gaithersburg, MD; 4) Department of Genome Sciences, University of Washington, Seattle, WA; 5) Department of Bioengineering, University of Washington, Seattle, WA, USA; 6) Genetic Networks Program, Canadian Institute for Advanced Research, Toronto, Ontario, Canada.

Empirically testing genetic variants is the gold standard for predicting pathogenicity, but for most variants these data are lacking. *PTEN* Hamartoma Tumor Syndrome (PHTS) includes a set of clinical presentations sharing benign overgrowths and cancer susceptibility caused by germline mutations. *PTEN* mutation carriers may also display macrocephaly with autism spectrum disorder (ASD) or developmental delay with or without symptoms of PHTS. Limited availability of clinical data and the scale of experimental data on specific mutations has hampered our understanding of *PTEN* genotype-phenotype relationships and over half of ClinVar missense variants are still characterized as variants of uncertain significance (VUS). Deep mutational scanning (DMS) datasets that test the effects of thousands of variants in parallel have been generated for a few human genes. For *PTEN*, DMS data describe effects of variants on lipid phosphatase activity (LPA) (7,244/8,463 variants, 86%, Mighell *et al.*, 2018) as well as cellular stability (CS, 4,112 variants, 49%, Matreyek *et al.*, 2018). LPA alone can predict pathogenicity with high positive predictive value (PPV, 93%), but modest sensitivity (SEN, 83%). ASD-associated variants retain hypomorphic LPA, while PHTS-associated variants are generally loss-of-function. We hypothesized that integrating multiple *PTEN* DMS datasets would improve pathogenicity prediction and form a more complete genotype-phenotype map. Using random forest models trained on 12-15 biochemical and evolutionary features, we first imputed missing LPA and CS scores with high accuracy ($r=0.80$ and 0.76 , respectively, 10x cross validation). We then developed a two-step integrated pathogenicity model, which was evaluated on 79 pathogenic (ClinVar) and 74 putatively benign (gnomAD) missense variants. This model increased SEN (90%) and maintained high PPV (87%) relative to LPA alone. Applying this model to VUS supports reclassification of 49 (37%) as pathogenic. We also found 617/1,941 (32%) of variants with low LPA had high CS, which may indicate a dominant negative effect. Finally, integrating published missense variants with a new set of 80 recurrent clinical variants revealed that variants found in both PHTS and ASD trend toward intermediate LPA scores that fall between PHTS or ASD alone. These data support the use of integrated DMS datasets for dissecting complex disorders and lay the groundwork for personalized care for *PTEN* mutation carriers.

1742F

A copy number variation map of pediatric neuropsychiatric traits. M. Zarre^{1,2}, C. Burton³, W. Engchuan^{1,2}, D. Merico^{1,4}, E. Higginbotham^{1,2}, M. Woodbury-Smith^{1,5}, J. Wei¹, J. Xiao², J. Shan², J. MacDonald¹, B. Thiruvahindrapuram¹, A. Chan^{1,2}, G. Pellecchia¹, T. Nalpathamkalam¹, R. Davies^{1,2}, M. Faheem^{1,2}, R. Manshaei⁶, S. Lamoureux¹, S. Walker¹, L. Strug^{1,2,7}, R. Yuen^{1,2}, C. Marshall¹, R. Wintle¹, P. Arnold^{8,9}, J. Crosbie³, R. Schachar³, S. Scherer^{1,2,10}. 1) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Ontario, Canada; 2) Program in Genetics and Genome Biology, the Hospital for Sick Children, Toronto, Canada; 3) Neurosciences and Mental Health, Psychiatry, The Hospital for Sick Children, Toronto, Ontario, Canada; 4) Deep Genomics Inc., Toronto, Ontario, Canada; 5) Institute of Neuroscience, Newcastle University, Newcastle Upon Tyne, UK; 6) Ted Rogers Centre for Heart Research, Cardiac Genome Clinic, The Hospital for Sick Children, Toronto, Ontario, Canada; 7) Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, Canada; 8) Mathison Centre for Mental Health Research and Education, Hotchkiss Brain Institute, University of Calgary, Alberta, Canada; 9) Departments of Psychiatry and Medical Genetics, Cumming school of Medicine, University of Calgary, Alberta, Canada; 10) Department of Molecular Genetics and McLaughlin Centre, University of Toronto, Toronto, Ontario, Canada.

'Spit for Science' is a cohort of over 17,000 youth (6-17 years of age) with associated psychiatric phenotype measurements and saliva DNA collection. Quantitative phenotype data were collected for attention deficit hyperactivity disorder (ADHD), obsessive compulsive disorder (OCD), and response inhibition using Stop-Signal Reaction Time (SSRT). We genotyped unrelated Caucasians (n=4,817) using Illumina HumanCoreExome beadchips. Participants with self-reported diagnosis of neurodevelopmental disorders (NDD; e.g. autism, OCD, ADHD), comprised 1,100 (23%) individuals. A total of 1,907 (40%) of participants had either a diagnosis of NDD or having neuropsychiatric trait(s). We performed global and functional gene-set burden analyses, and locus association test on rare CNVs and SNV/indels using linear regression against different traits. We identified CNVs linked with neuropsychiatric disorders in 191 (4.0%) participants with elevated standardized z-scores, for all traits. These variants were enriched in participants with a diagnosis of NDD ($p=0.002$), but not in those only on a trait spectrum ($p=0.617$). These CNVs included i) sex-chromosome and autosomal aneuploidies (e.g. 45,XO, 47,XXY, 47,XXX, 47,YYY, 48,XXXY, and 47,XX,+21) in 12 (0.25%) individuals, ii) 13 unique genomic loci previously implicated in NDD (e.g. 1q21.1, 15q11.2, 16p12.1, 17p12, 22q11.2) in 52 (1.1%) different participants, and iii) large CNVs affecting genes with known functions in the human brain such as *ASTN2*, *CNTN4*, *DLGAP2*, *MACROD2*, *NRXN1*, and others. ADHD traits were associated with increased deletion burden ($p=0.014$), whereas there was an increased burden of duplication for SSRT ($p=0.034$). We found no global CNV burden for OCD traits. ADHD traits showed significant enrichment for deletions in genes associated with mouse neurophenotypes ($p=0.009$) and with human brain gene expression ($p=0.04$). OCD and SSRT traits showed enrichment for duplications in genes expressed in human brain ($p=0.004$ and 0.016 , respectively). Adding the effect of SNVs/indels increases the significance level in many neurofunction and brain expression gene sets for ADHD traits. The following genes were driving signals: ADHD: *TUBGCP5*, *GALC*, *GJB6*, *TERT*, *BTN3A3*, *GCNT2*, *GPC5*, *RBFOX1*, *PLEC*, *MACROD2*, *PRKN*; OCD: *PRKCE*, *SRBD1*, *TNR*, *DISC1*; SSRT: *KIF1A*, *LDOC1*, *RB1CC1*. SSRT traits showed enrichment for duplications in eight loci: 1q21.1, 1q42.3, 2q37.3, 9q22.2, 12q13.13, 18p11.21, and 22q11.21.

1743W

Phased SLC6A4 promoter haplotype determination using long-read single molecule real-time (SMRT) sequencing. Y. Yang^{1,2}, M.R. Botton^{1,2}, A. Salomatov², E.R. Scott^{1,2}, R. Sebra^{1,2}, R.J. Desnick¹, L. Edelmann^{1,2}, E.E. Schadt^{1,2}, S.A. Scott^{1,2}. 1) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029; 2) Sema4, a Mount Sinai venture, Stamford, CT 06902.

The *SLC6A4* gene at chromosome 17q11.2 encodes the serotonin (5-HT) transporter (5-HTT), which mediates 5-HT reuptake and is the target of many tricyclic antidepressants and selective serotonin reuptake inhibitors (SSRI). 5-HTT expression influences 5-HT transporter function, which has been implicated in psychiatric disorder risk and interindividual SSRI response variability. The polymorphic *SLC6A4* promoter is composed of a variable number of homologous 20-24 bp repeats (5-HTTLPR), which is strongly correlated with 5-HTT expression. The long (16 repeats) and short (14 repeats) alleles of the insertion/deletion variant (rs4795541) are associated with higher and lower expression, respectively. However, two additional *SLC6A4* promoter variants (rs25531, rs25532) also influence expression but are most informative when considered with the insertion/deletion as a promoter haplotype. Given the clinical relevance and polymorphic nature of the *SLC6A4* promoter, we developed a novel long-read single molecule real-time (SMRT) sequencing method using the Pacific Biosciences (PacBio) platform. Overlapping *SLC6A4* promoter amplicons up to 2 kb had complete concordance between all identified variants across six control samples. Importantly, whole genome sequencing (WGS) data from 70 GeT-RM Coriell controls (~45X) indicated that short-read sequencing does not detect the short *SLC6A4* promoter allele due to the homologous tandem repeats; however, SMRT sequencing of 1.0 kb amplicons unambiguously identified the long and short alleles and clearly phased the polymorphic promoter in all 70 samples, including complex compound heterozygous haplotypes. Moreover, six additional promoter variants were detected by SMRT sequencing, including two samples with expanded 5-HTTLPR haplotypes that were also not detected by short-read sequencing. Targeted enrichment and high-depth (~330X) short-read sequencing also could not resolve the *SLC6A4* promoter architecture in an additional 96 control samples. A final cohort of 52 samples were subjected to *SLC6A4* SMRT sequencing, which clearly detected phased promoter haplotypes that were confirmed by orthogonal genotyping and Sanger sequencing in a subset of samples. These data indicate that long-read *SLC6A4* promoter SMRT sequencing is an innovative and validated method for phased resolution across complex repetitive regions, which will likely improve *SLC6A4* interrogation for both candidate gene research studies and clinical pharmacogenetic testing.

1744T

Cell-type specific differential gene expression in post-mortem snRNA-seq of the depressed brain. M. Maitra¹, C. Nagy¹, M. Suderman², J.F. Thérault³, I. Ragoussis³, G. Turecki¹. 1) McGill Group for Suicide Studies, Douglas Institute, Verdun, QC, Canada; 2) Bristol Medical School - Population Health Sciences, Bristol, UK; 3) McGill University Centre for Molecular and Computational Genomics, Montreal, QC, Canada.

Background: Measurement of gene expression changes in homogenized tissues with very heterogeneous cell-type composition, as seen in the brain, can mask subtle cell-type specific effects. Multiple high-throughput protocols for single-nucleus RNA-sequencing (snRNA-seq) have been established in archived post mortem human brain tissue. Here we exploit the enhanced ability for genetically defining cellular subgroups on a large scale to identify cell-type specific differentially expressed genes in major depressive disorder. **Methods:** We used 10X Genomics' droplet based snRNA-seq technology and a custom, optimized wet-lab protocol, to characterize single-nucleus transcriptomic profiles of > 70,000 cells from the prefrontal cortex (Brodmann areas 8 and 9) of 34 individuals - including 17 depressed patients who died by suicide and 17 psychiatrically healthy subjects. A customized bioinformatic approach using the Seurat and lme4 R packages was used for data analysis. **Results:** Graph-based clustering within Seurat was followed by automatic and manual quality control of clusters. Annotation of cell-types based on previously established markers, allowed us to identify 10 excitatory neuronal clusters, 7 inhibitory neuronal clusters, and 9 non-neuronal clusters. Major glial cell-types including astrocytes, oligodendrocyte lineage cells, and microglial cells were readily identified. Cortical layer specific markers and interneuron subtypes markers also demarcated sub-groups of cells as expected. Expression profiles were in good agreement with previously published human brain snRNA-seq datasets. Within each cluster, further selection of high-quality cells followed by differential expression analysis with lme4 revealed 16 significantly upregulated and 80 significantly downregulated genes across 16 out of the 26 clusters (FDR < 0.1). **Conclusion:** Our sample preparation and custom analysis protocols allowed us to identify specific cell-types which are dysregulated in the prefrontal cortex in depressed individuals who died by suicide and genes with altered expression in those cell-types. Further analysis of the interconnections between the dysregulated cell-types and between the identified genes should yield insights into the cellular and molecular pathways that contribute to depression.

1745F

Neural stem cells from MAND patients reveal a significant role for MBD5 in neurodevelopment, circadian rhythm, Wnt and leptin signaling, and autism. S.V. Mullegama¹, S.R. Williams², S. Klein³, J.W. Innis⁴, F.J. Probst⁵, C. Haldeman-Englert⁶, J.A. Martinez-Agosto², T. Ezashi⁶, S.H. Elsea². 1) Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX, 77030, USA; 2) University of Virginia, VA; 3) Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California, USA; 4) Departments of Human Genetics and Pediatrics and Communicable Diseases, University of Michigan, Ann Arbor, MI 48109, USA; 5) Department of Pediatrics, Section on Medical Genetics, Wake Forest University Health Sciences, Winston-Salem, NC, 27157, USA; 6) Division of Animal Sciences, University of Missouri, Columbia, Missouri 65211, USA.

MBD5-Associated Neurodevelopmental Disorder (MAND) is an autism spectrum disorder (ASD) characterized by intellectual disability, motor delay, severe speech impairment and autism behavioral problems. MAND serves as an umbrella term that encompasses 2q23.1 deletion syndrome, *MBD5* variant cases and 2q23.1 duplications. The importance of *MBD5* in neurodevelopmental function and autism remains largely undefined. Thus, to understand the role of *MBD5* in neurodevelopmental function and autism, we created neural progenitor stem cells derived from patients with *MBD5* deletions and conducted RNA-seq to identify the contributory altered gene pathways. We identified 499 genome-wide significant differentially expressed annotated genes. To understand the neurodevelopmental functions affected when *MBD5* is deleted, we used DAVID to perform GO analysis of significantly differentially expressed genes from each analysis. Most of the GO clusters were related to three groups: neuron development, brain development and cellular processes. We utilized the Reactome FIViz application in Cytoscape to identify the top signaling pathways enriched in our genome-wide significantly differentially expressed genes. We found significant enrichment for Hippo, CREB, leptin, MAPK, circadian rhythm, and Wnt signaling pathways. The transcriptional impact of *MBD5* deletions on expression of known autism-associated genes was significant. We show that 60 genes which are differentially expressed in MAND neural progenitor stem cells are known ASD associated genes according to SFARI. Utilizing ToppCluster, we compared differentially expressed genes in MAND to the publically available SFARI syndromic autism genes to identify common biological processes. This analysis revealed a striking significant overlap between the two gene lists in biological processes which are in keeping with the neurological phenotype and ASD, such as neuronal differentiation, brain development, and regulation of neurogenesis. Overall, our study reveals that deletion of *MBD5* disrupts important processes involved in neurodevelopment and ASD, identifying specific pathways that may be targets for therapeutic intervention. .

1746W

A splice-site mutation and overexpression of CCNT2 in autosomal recessive intellectual disability patients. K. InanlooRahatloo^{1,2}, F. Peyman², K. Kahrizi², H. Najmabad². 1) School of Biology, College of Science, University of Tehran, Tehran, Iran; 2) Genetic Research Center, University of social welfare and Rehabilitation Sciences, Tehran, Iran.

Intellectual disability (ID) is a most prevalent neurodevelopmental disorder. Although ID is a clinically important disorder, the etiology and pathogenesis are poorly understood. Transcriptome profiling is a very important tool for understanding how genetic variants alter cell expression. Exome sequencing revealed a homozygous splice site mutation (chr2: 135677463, T>G) in the *CCNT2* gene in one ID family. *CCNT2* gene belongs to the cyclin family and gene ontology annotation related to this gene includes protein kinase binding. RNA-Seq results indicated overexpression of normal isoform and non-protein coding isoform of *CCNT2* gene in *CCNT2* mutant patients. Furthermore, *CCNT2* mutation was found to disrupt the transcriptional regulatory function of *CCNT2* in both patients. In silico sequence analysis using PROMO 2.0 web based tool predicted the creation of two binding sites for transcription factor YY1 in the mutant G allele. TRAP also predicted the creation of binding sites for several transcription factors in the mutant sequence such as YY1, SRF and, ELK. MISO (Mixture of Isoforms) results indicated the retention of intron 2 in *CCNT2* gene in patients. Allele-specific expression study of the *CCNT2* gene showed that the expression of mutant allele was higher than wild-type allele in heterozygote control. It seems likely that the mutation leading to an increased expression in mutant cases. To obtain insights into the biological pathways influenced by overexpression of *CCNT2*, we applied our RNA-Seq data to several gene ontology programs. The most significantly dysregulated pathways were relevant to neural development and function, such as "Tyrosine protein Kinase SRC", "Nervous system disease", "Gene expression", and "actin cytoskeletal organization". Pathway enrichment analysis using Genomatix revealed a down-regulation of genes linked with nervous system development, regulation of axonogenesis, actin cytoskeleton pathways and over-representation of gene expression pathway in *CCNT2* mutant patients. The findings reported here provide new insights into the molecular pathways underlying intellectual disability and highlight the importance of this transcription factor in neuronal functions.

1747T

Single cell sequencing of iPSC neural cells from Down syndrome patients uncovers perturbed cell differentiation. J. Klar, J. Schuster, M. Sobol, L. Laan, A. Fatima, N. Dahl. Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala, Sweden.

Background: Down syndrome (DS) is a leading genetic cause of intellectual disability. Approximately 95% of patients with DS have a full trisomy 21 (T21) resulting in a global transcriptional dysregulation. Furthermore, studies suggest that the distribution of neuronal and glial cells is altered in DS. **Aim:** To investigate transcriptional profiles of single neural cells with T21 to uncover alterations in cell fate specification, commitment and differentiation relevant for DS brain development. **Methods:** We established induced pluripotent stem (iPS) cells from a patient with DS (full T21) and a healthy control individual. The iPS cells were further differentiated into neural cells for 30 days using an undirected protocol followed by immunostaining for neural markers and RNA extraction. Sequencing libraries for single cell sequencing was prepared using the Chromium Single Cell 3'Reagent Kits v2 (10X Genomics). Sequencing was performed on a HiSeq 2500 PE (Illumina) and the data was analyzed using Cell Ranger V2.0 (10X Genomics) and Monocle 2 (Trapnell Lab). **Results:** Cluster analysis of transcriptomes from 3734 euploid single neural cells shows that 63% correspond to radial glia, 6% to intermediate progenitor cells, 20% to neurons and 5% to a cluster of glia cells expressing *PDGFRA*. A pseudo-time analysis of euploid cells confirms an early formation of radial glia followed by mature glia and neuronal cells. In contrast, the cluster analysis of 4169 T21 cells shows that 24% correspond to radial glia, 5% to intermediate progenitor cells and only 4% neuron-like cells. Importantly, 56% of trisomic cells express the glia cell marker *PDGFRA*, making this cell type predominant in trisomic cells. Furthermore, pseudo-time analysis of T21 cells reveals a continuous formation of glial cells along differentiation as well as a relative reduction in the formation of neuronal-like cells. **Conclusion:** In our iPSC model of T21 neurogenesis we observe a reduced proportion of neuronal cells and an increased proportion of glia cell population when compared to euploid control cells. In T21 cells, the formation of glia cell populations appear to be continuous and with a differentiation trajectory distinct from that of euploid cells.

1748F

Scaling up genome engineering for modeling reciprocal genomic disorders. X. Nuttle^{1,2,3}, D.J.C. Tai^{1,2,3}, C.E. de Esch^{1,2,3}, K. Mohajer^{1,2,3}, J.F. Gusella^{1,4}, M.E. Talkowski^{1,2,3,5}. 1) Center for Genomic Medicine and Department of Neurology, Massachusetts General Hospital, Boston, MA, USA; 2) Department of Neurology, Harvard Medical School, Boston, MA, USA; 3) Program in Medical and Population Genetics and Stanley Center for Psychiatric Research, Broad Institute, Cambridge, MA, USA; 4) Department of Genetics, Harvard Medical School, Boston, MA, USA; 5) Departments of Pathology and Psychiatry, Massachusetts General Hospital, Boston, MA, USA.

Reciprocal genomic disorders (RGDs) arise from recurrent deletions and duplications of specific chromosomal segments and account for 5-10% of autism spectrum disorder, schizophrenia, and intellectual disability. Performing genome engineering in induced pluripotent stem cells (iPSCs), we previously recapitulated microdeletions and microduplications at 16p11.2 and 15q13.3, some of the most common pathogenic rearrangements associated with abnormal neurodevelopment. Extending this approach, we generated cellular models of Prader-Willi syndrome and Angelman syndrome in two iPSC lines, including ~6 Mbp type I deletions and ~5.3 Mbp type II deletions on both parental homologs. To streamline the expensive, time-consuming process of creating and validating RGD cellular models, we developed a parallelized genome engineering approach incorporating libraries of CRISPR/Cas9 guide RNAs (gRNAs) into a piggyBac (PB) transposon system. We designed 59 gRNAs for producing deletions and reciprocal duplications at 19 RGD loci and cloned them into a plasmid-encoded PB transposon harboring doxycycline-inducible Cas9 and a positive/negative selection marker. We co-transfected iPSCs with this library and a plasmid encoding PB transposase, evaluating 12 different combinations of library and transposase DNA input. Leveraging cell-population and single-cell amplicon sequencing, we are currently identifying conditions yielding 0-1 genomic PB integrations per cell with complete library representation. After optimizing transfection parameters to ensure each puromycin-resistant cell expresses one gRNA and each gRNA gets expressed in several cells, we will induce Cas9 expression to perform efficient, temporally controlled genome editing. Using fluorescence-activated cell sorting, we will then isolate single cells, expand resultant colonies, and screen for deletions and duplications of interest using molecular inversion probes. To remove our PB construct from genomes of successfully edited colonies, we will transfect them with a plasmid encoding excision-only PB transposase. Because this transposase achieves >99.7% scarless excision, the resulting cellular models will be near-isogenic, genetically differing only by targeted modifications. The ability to rapidly establish high-quality cellular models at scale will position us well to investigate functional consequences of RGD rearrangements and disrupt specific genes within RGD regions to elucidate their roles in pathogenesis.

1749W

Integrated functional characterization of a non-coding GPC3 allele in a family with recurrent cases of suspected Simpson-Golabi-Behmel syndrome. A. Vitobello¹, J. Thevenon², P. Callier^{1,3}, T. Jouan¹, L. Duplomb¹, M. Bordesoules¹, Y. Duffourd^{1,3}, E. Tisserand¹, A. Boland⁴, R. Olasso⁴, J.F. Deleuze⁴, D. Sanlaville⁵, C. Thauvin-Robinet^{1,3,6}, L. Faivre^{1,3,6}. 1) Université Bourgogne Franche-Comté, Inserm UMR 1231 GAD team, Genetics of Developmental disorders, Dijon, Burgundy, France; 2) Service de Génétique, CHU de Grenoble, Grenoble, France; 3) CHU Dijon, FHU TRANSLAD, Dijon, France; 4) Centre National de Recherche en Génomique Humaine (CNRGH), CEA, Evry Cedex, France; 5) Service de Génétique, Laboratoire de Cytogénétique, CHU de Lyon, HCL, France; 6) CHU Dijon, Centre de référence Anomalies du Développement et Syndromes Malformatifs, Centre de Génétique, F-21000 Dijon, France.

Simpson-Golabi-Behmel syndrome (SGBS) is an X-linked condition occurring primarily in males, characterized by pre- and postnatal overgrowth, coarse facial features, variable intellectual disability, congenital heart defects, supernumerary nipples, hepatomegaly and increased risk to develop neoplasia such as Wilms tumors. The major cause of SGBS is mutation/deletion in the gene *GPC3* encoding glypican-3, an outer membrane glycoprotein implicated in WNTs, Hedgehogs (Hh), fibroblast growth factors (FGF), and bone morphogenetic proteins (BMP) signaling modulation. Yet, a variable fraction of convincing SGBS are negative to routine *GPC3* testing, indicating either genetic heterogeneity or clinical overlapping with different overgrowth conditions. Targeted approaches as well as whole exome sequencing (WES) analysis of single-nucleotide variants (SNVs) and copy-number variants (CNVs) failed to detect the molecular cause responsible for recurrent incidence of suspected Simpson-Golabi-Behmel syndrome within a family referred to our center. An integrated analysis deploying whole genome sequencing (WGS) and total RNA sequencing (RNA-seq) was performed in patient-derived samples, in order to interrogate non-coding mechanisms potentially accounting for dysregulation of known and novel candidate genes implicated in overgrowth syndromes. WGS allowed us to identify the presence of a complex intrachromosomal rearrangement resulting in the insertion of a 150 Kb region within the second intron of *GPC3*. Furthermore, RNA-seq analysis allowed the identification of SGBS-specific transcriptional pathways compatible with glypican-3 knockout. We provide evidence of the utility of genome-wide genetic and expression analyses as complementary approaches to interrogate the transcriptional readout of non-coding variants in an ultra-rare genetic condition with negative or inconclusive routine laboratory results in undiagnosed overgrowth syndromes.

1750T

A next-generation mapping approach for the assembly and detection of structural variation within genomic regions containing complex segmental duplications. F. Yilmaz^{1,2}, Y. Mostovoy², E.A. Geiger², N.J.L. Meeks², K.C. Chatfield², C.R. Coughlin II², P.Y. Kwok^{3,4,5}, T.H. Shaikh². 1) Integrative and Systems Biology Program, University of Colorado Denver, Denver, CO; 2) Department of Pediatrics, University of Colorado Denver, School of Medicine, Aurora, CO; 3) Cardiovascular Research Institute, University of California, San Francisco, San Francisco, CA; 4) Institute for Human Genetics, University of California, San Francisco, San Francisco, CA; 5) Department of Dermatology, University of California, San Francisco, San Francisco, CA.

Segmental duplications (SDs) are a class of low copy repeats, which promote genomic rearrangements associated with numerous genomic disorders via nonallelic homologous recombination (NAHR). In addition, SD-mediated structural variations (SVs) also play a role in genomic variation within the general population. However, the true extent of SD-mediated variation in the genome remains to be fully elucidated, due to the limitations in accurate mapping and sequencing of these complex regions using conventional techniques, being especially recalcitrant to short-read sequencing technologies. We used next-generation mapping with Bionano Genomics' Irys and Saphyr systems to analyze SD-containing regions within the human genome. We focused on three genomic regions, 7q11.23, 15q13.3 and 16p12.2, containing large (>300 Kb), paralogous SDs associated with known genomic disorders. We first analyzed optical mapping data from 154 control individuals from five super populations within the 1000 Genomes Project: African, American, European, East Asian and South Asian. Using a combination of single molecules and *de novo* assembled contigs from multiple overlapping molecules, we were able to accurately map the SD-containing regions on 7q11.23, 15q13.3 and 16p12.2. We confirmed previously known large inversions of these regions, which were observed in all populations. Notably, we observed several novel SVs within specific paralogs of SDs. These included copy number variations of a ~16 kb duplicated module within 7q11.23 SDs, ranging between one copy, which was the most common, and six copies, which was the rarest. In addition, we also observed several novel inversions within 7q11.23 SDs. In 15q13.3, a large (~500kb) novel deletion was observed within one of the SDs in all populations, along with multiple inversions, both previously known and novel. Thus, we present a more refined map of SD-mediated variation in these three genomic regions. We were also able to accurately identify the hybrid junction fragments resulting from NAHR-mediated microdeletions in multiple patient samples, allowing an unprecedented level of breakpoint localization within highly paralogous SDs. The high level of variation observed within SDs may contribute to variation in the degree of their instability and their subsequent predisposition to rearrangements associated with genomic disorders. Overall, we demonstrate the power of NGM for better mapping and characterization of SD-containing genomic regions.

1751F

Longitudinal aging-associated gene expression profiling in African green monkey (*Chlorocebus aethiops*). J. Huh^{1,2}, J. Lee¹, S. Choe^{1,2}, H. Cho^{1,2}, S. Park¹, Y. Kim^{1,2}. 1) National Primate Research Center, Korea Research Institute of Bioscience and Biotechnology(KRIBB), 30 Yeongudanji-ro, Ochang-eup, Cheongwon-gu, Cheongju-si, Chungcheongbuk-do, Korea; 2) Department of Functional Genomics, KRIBB School of Bioscience, Korea University of Science & Technology(UST), 217 Gajeong-ro, Yuseong-gu, Daejeon, Korea.

Chlorocebus aethiops is an old world monkey, has been traditionally used as an experimental model for biomedical research. In addition, recent technological advances of next generation sequencing is useful for unravel the genetic mechanisms of senescence, aging and age-related diseases. The blood transcriptome of healthy aged *C. aethiops* from nine individuals ranging in age from 17 to 25 years old, were analyzed in each of the last three years by the Illumina sequencing platform. We identified 910 to 1399 accumulated differentially expressed genes (DEGs) in each nine individuals, tend to increase with age. Aging-related DEGs were sorted in total three time points of each nine individuals. A major proportion of the aging-related DEGs were belonged to gene ontology (GO) categories involved in translation, rRNA metabolic process. Next, we sorted aging-related common DEGs in total three time points of all nine individuals. Aging-related common DEGs mainly were belonged to GO categories involved in translation, cellular component biogenesis, rRNA metabolic process, cellular component organization or biogenesis, and RNA metabolic process. Furthermore, we identified 29 aging gene which upregulated through the time, by time-series analysis. These aging genes were linked to protein synthesis and mitochondrial dysfunction. This study reported the gene expression changes during aging in *C. aethiops* longitudinal transcriptome, and aging genes identified have the potential as the particular target of aging.

1752W

Simons Foundation data through the GPF (Genotypes and Phenotypes in Families) system. *I. Iossifov^{1,2}, L. Chorbadjiev², N. Dimitrov³, A. Chernae³, E. Gotsev³, N. Volfovsky⁴.* 1) CSHL, Cold Spring Harbor, NY; 2) NYGC, New York, NY; 3) Seqpipe Ltd, Sofia, Bulgaria; 4) Simons Foundation, New York, NY.

In recent years, we witnessed the impressive success in the genetics of autism and other disorders through the analysis of sequence datasets generated from deeply phenotyped large family collection like the Simons Simplex Collection (SSC). There is an enormous amount of work that needs to follow the early success in order to develop effective treatment and early diagnostic strategies. A variety of future research projects will study in detail the effects of hundreds genetic variants and genes at molecular, cellular and organismic levels. Such projects will greatly benefit from the accumulated datasets but their large size and complex structure create a major obstacle for their efficient use. We addressed such difficulties by building a system called Genotype and Phenotype in Families (GPF) that enabled simple and intuitive interaction with the data that was generated by us and data generated independently by other groups. The system allows the integration and the joint analyzes of diverse phenotypic and genotypic data and the efficient distribution of the valuable resources to the wider scientific community through the GPF's intuitive web interface. An instance of the GPF system was deployed at SFARI (<http://gpf.sfari.org>) where it manages a number of large genotypic and phenotypic datasets built through SAFARI's support. These datasets include the phenotypic data gathered from SSC, Simons Variation in Individuals Project (SVIP), and Simons Foundation Powering Autism Research for Knowledge (SPARK) collections and the genotypic data from these collections generated through whole-exome and whole-genome sequencing. .

1753T

Platform for single-cell transcriptomics data analysis and management. *S. Ramachandrala, A. Pal, H. Adil, D. Chohan, S. Cherukuri, M. Nagarajan.* Strand Life Sciences Private Limited, Bengaluru, India.

StrandNGS is a next-generation sequencing data analysis software designed for the biologists. By following the well-established computational best practices, StrandNGS doubles up as flexible data management system as well. While there are innumerable open source pipelines that help in the analysis of single-cell transcriptomics sequencing studies, data management and results verification remains a challenge. Also, given that most of these are standalone approaches, results from various approaches cannot be easily compared against each other. By allowing the use of R-scripts and Jython-scripts well within the framework of the software, StrandNGS provides a platform to not only organize all the data, analysis, and results but also provides immaculate visualization options that help in comparing and understanding the results from various approaches. Using data from a mouse pancreatic single-cell study (GSE87375), we illustrate these capabilities of StrandNGS. The raw data was imported and aligned against known RefSeq transcripts. Differential expression analysis across the various cell types and developmental stages was performed using the statistical tests that are packaged in the software, and the tests executed through R-scripts. Also, resultant gene lists from externally run tests were imported into the software. The confounding effects in the data set due to varied cell collection methods were identified and corrected for using the in-built surrogate variable analysis; the corrected data matrix was used for all subsequent analysis. Expression profile clustering with sample metadata framework and t-SNE visualization options were extensively used for pattern identification and refinement of results. Moreover, using NLP framework available in StrandNGS, developmental models from peer-reviewed publications were mined and converted into networks. Data from the current study was overlaid on those networks as quilt plots to assess conformity with previous studies. .

1754F

Omic-analysis to discovery genetic biomarker of severe forms of incontinentia pigmenti by using IPGB biobank samples. *F. Fusco¹, V. Valente¹, D. Fergola¹, A. Pescatore¹, MB. Lio², MV. Ursini².* 1) Institute of Genetics and Biophysics "A. Buzzati-Traverso" CNR, Italy; 2) University of Basilicata, Potenza 85100, Italy.

The phenotypic heterogeneity of Incontinentia Pigmenti (IP, MIM308300), a rare X-linked dominant disorder affecting neuroectoderm, suggests to search for molecular genetic markers to aim personalized therapeutic strategies. Although significant progress has been made over the last few years in the diagnosis of this rare condition, the genetic defect and underlying pathological abnormality remain unknown. Moreover, to date no definitive cure is available for IP, nor are there sufficiently reliable and specific biomarkers to predict the disease severity. This is in part due to the rarity and genetic and phenotypic heterogeneity of IP disease and the lack of access to patient samples. The availability of the genetic Biobank for IP has addressed this bottleneck and supported IP research. The Incontinentia Pigmenti Genetic BioBank Project (IPGB, <http://www.igb.cnr.it/ipgb>) was launched in 2015 with the aim of providing evidence for the implementation of personalized medicine by constructing a large, patient-based biobank and to enhance research on IP, related to pathophysiology, biomarkers and therapeutic approaches. The cohort includes 0–60 years old, encompassing 436 clinically-confirmed IP cases (IP females and IP males) from 200 cooperative medical institutes worldwide including USA. The Biobank contains blood samples, peripheral blood mononuclear cells (PBMC), red blood cells/granulocyte pellet (totaling 1370 samples). We will describe the IPGB infrastructure and we will present preliminary results of the project to identify what variations/genes/pathways may affect the severity of IP, we used an OMIC-strategy: Target Exome Sequencing of Metabolic Pathway in 80 IP cases (40 IP-severe, 40IP-mild); Whole Exome and Transcriptome Sequencing in 3 IP-severe. More than 30,000 high quality variants were found across the haloplex samples, with an average of 430 variants/sample, and more than 300,000 across the exome samples, with an average of 40,900 variants/sample. Statistical analysis and gene validation are in progress. All IP samples belonged to our historical collection and to IPGB biobank. This research was supported by grant from the CNR-DSB Progetto-Bandiera "InterOmics".

1755W

Cell-free DNA analysis by automated electrophoresis. *M. Liu, S. Borg.* Agilent Technologies, La Jolla, CA.

Sequencing of cell-free DNA (cfDNA) extracted from blood specimens or other body fluids is possible due to the establishment of low input library protocols for next-generation sequencing workflows. Accurate quantification of cfDNA samples is essential to determine suitable input amounts for cfDNA library preparation prior to sequencing. The main component of cfDNA samples is the mononucleosome with a size around 170 bp, sometimes with additional species representing nucleosome multimers. Further, cfDNA samples may contain larger DNA fragments dependent on the donor's health status, preanalytical sample treatment, or extraction method. High molecular weight material can negatively influence library preparation and subsequently result in lower sequencing depth. Therefore, reliable quantification of cfDNA requires a method that separates DNA fragments by size, such as electrophoresis. This poster shows the use of a new cell-free DNA assay for an automated electrophoresis system. The cell-free DNA assay enables automated cfDNA quantification with pre-defined regions. Moreover, the results include sample purity as a score to qualify cfDNA samples according to their contamination level with high molecular weight material. The features of the cfDNA assay are described with examples of typical sample patterns.

1756T

BGI employee multi-scale and multi-omics cohort and analysis. *R. Li^{1,2}, P. Li¹, R. Guo¹, L. Chen¹, X. Liu¹, C. Nie¹.* 1) BGI-Research, Shenzhen, China; 2) Harbin Institute of Technology Shenzhen Graduate School, Xili University Town, Shenzhen, China.

Body check data of BGI employee were collected annually in BGI to get multi-omics data of health people since 2015. In 2017, multi-scale and multi-omics data of ~4000 employee were collected, including whole genome sequences, immune repertoire and gut metagenome sequences; blood biochemistry and routine test, hormones, amino acids and vitamins and micro-elements concentrations in blood; facial skin image test, body composition, physiological and psychological test and life-style survey. Using 2017 data, we built baselines and setup an aging (biological ages) and healthy assessment system of different aspects of health, like skin, immune, body composition, metabolism, Gut microbiota etc. We constructed a correlation network of different data and communities within reveal groups of closely correlated biomarkers. We found Body composition data were the most correlated data type, indicating that body composition can be general marks in health improvement practices. Personal network relative to the whole cohort gives candidates of abnormal biomarkers pairs, which were in accordance with our health and aging assessment results. For the first time, we did GWAS analysis so many phenotype, most are new, like skin and immune. GWAS-PheWAS together with lifestyle analysis separated genetic and environmental risk of specific health problem, which gave clues to guide health intervenes, for example we found smoking is the key factor determined Cadmium concentration in blood. Gender differences of phenotypes changes with age were also analyzed, simultaneously we found several biomarkers changes with age at different pace for different genders, including physiological and nutritional biomarkers. Our results show measurement of multi-scale and multi-omics data can increase our knowledge of health and help to improve effect of health intervenes.

1757F

A rapid library prep with enzymatic shearing for an improved next generation sequencing workflow for WGS and target enrichment. *J. Lenhart, B. Lahann, R. Stedtfeld, S. Sandhu, L. Kurihara, V. Makarov.* SwiftBiosciences, Ann Arbor, MI.

Cost efficient and rapid NGS library preparation protocols will further enable high-throughput next generation sequencing for inherited and somatic variant detection. To this end, a rapid library preparation method was developed that employs a combination of enzymatic fragmentation, unique DNA repair mechanisms, and an exceptionally efficient adapter ligation strategy to generate highly complex libraries for Illumina sequencing. Geared to increase NGS library complexity and optimize sequencing output, a sub two-hour workflow (not including PCR) provided libraries for WGS and targeted sequencing. Libraries were prepared from a range of DNA inputs including FFPE and sequenced on MiSeq and NovaSeq. Validated using genomic DNA from human samples, the workflow provided unbiased coverage and consistent fragmentation across multiple input quantities. Libraries prepared for target enrichment were further validated with the IDT xGen Lockdown Exome and Pan Cancer hybridization capture panels. The rapid library preparation workflow provided unbiased target coverage, high complexity, high percent bases on target, and maintained uniform coverage. In conclusion, Swift's new library preparation workflow enables a rapid workflow for whole genome and targeted sequencing of samples.

1758W

Towards precision medicine: Accelerating diagnostics for infectious and inflammatory diseases using blood gene expression biomarkers. M. Kaforou, D. Habgood-Coote, J.A. Herberg, A. Cunningham, M. Levin. Imperial College London, Medical School Building, Norfolk Place, London, W2 1PG, UK.

Infectious diseases continue to be amongst the most important causes of illness and mortality accounting for 17% of deaths worldwide and are responsible for a major component of the workload and cost of medical care. Available pathogen-based microbiological and laboratory tests are not sufficiently rapid or accurate to allow antibiotics to be safely withheld. Instead of trying to detect the pathogen, an alternative approach measuring gene expression patterns in the patient's blood is being introduced. However, until now, diagnostic signatures attempt to address the patient classification problem in a simplistic dichotomous way limiting the opportunities for clinical use (bacterial vs. viral, malaria vs. other diseases etc.). Here, we address gene expression based disease assignment, as a hierarchical multi-label classification problem mimicking the actual clinical decision-making processes. Using a cohort of well-phenotyped patient transcriptomic data (n=1350) with a range of infectious and inflammatory diseases, including bacterial and viral infections, as well as healthy controls we implemented a hierarchical multi-label classifier while rebalancing the classes according to cost of misclassification. We employed variable selection methods to identify the best set of biomarkers, which can, in different combinations, optimally discriminate between diseases. The biomarker set was discovered for all conditions jointly to maximize usage of the information captured by individual genes. Multi-label classification assigns each patient to a set of diagnostic labels, which are not mutually exclusive, and allows for the assignment of non-infectious diseases, co-infections and severity. Hierarchical multi-label classification methods applied on patient transcriptomic data can provide greater diagnostic information than simplistic binary classifiers or conventional clinical classification. An accurate and robust classifier based on a parsimonious set of genes has the potential to be measured using a single blood test and could enhance disease diagnosis and treatment decision-making, enabling the transition to the precision medicine era. [1] Herberg, J.A. et al. JAMA, 316 (8): p. 835-45 (2016) [2] Zaas, A.K. et al., Cell Host Microbe, 6 (3): p. 207-17. (2009) [3] Colborn, J.M. et al., J Immunol Res, 162639 (2015) [6] Kaforou, M., et al., PLoS Med, 10 (10): p. e1001538 (2013) [7] Anderson, S.T., et al., N Engl J Med, 370 (18): p. 1712-23 (2014).

1759T

Development and validation of the CESeq and WESeq test systems. S. Burns, M. Soucy, J. Uvalic, D. Bergeron, M. Peracchio, M. Preggo, G. Ananda, A. Hesse, P. Panneer Selvam, J. Davey, K. Kelly, H. Reddi. The Jackson Laboratory, Farmington, CT.

Introduction The Jackson Laboratory has introduced a new reflex testing algorithm for inherited disorders designed to efficiently address clinical utility from a diagnostic perspective while maintaining a short turn-around time. CESeq is a targeted panel allowing genotypic identification of causal variants in clinically significant coding exons spanning about 4,800 genes. If CESeq turns up no disease causing variants, the sample is reflexed to WESeq, which provides 45 Mb of exome coverage. **Methods** Both CESeq and WESeq were developed and validated using Illumina library preparation kits. Both assays were optimized for DNA input, batch size, PCR conditions, and loading concentration. For the validation of CESeq, 24 gDNA samples were blinded and prepared in 3 batches. WESeq validation included the evaluation of 21 blinded gDNA samples prepared in 3 batches. Both assays were assessed for accuracy, sensitivity, specificity, precision, limit of detection, and reportable range. Variants were called by the Genome Analysis Toolkit (GATK) HaplotypeCaller, annotated using Alamut Batch against the reference genome build HG19/GRCh37 and further computationally annotated by the gene/variant prioritizer software Exomiser using best-matched controlled Human Phenotype Ontology (HPO) terms. **Results** Table 1 demonstrates the results for each performance characteristic.

Performance Characteristic	CESeq (%)	WESeq (%)
Accuracy	99.88	99.97
Sensitivity	96.5	99.76
Specificity	99.98	99.97
Precision	97.98	95.92
(Repeatability)		
Precision (Reproducibility)	96.12	95.9
Limit of Detection SNVs	minDP 15, AF width 25%	minDP 15, AF width 20%
Limit of Detection Indels	minDP 10, AF width 25%	minDP 10, AF width 20%
Reportable Range	Variants detected/not detected above the LOD within the region of interest.	Variants detected/not detected above the LOD within the region of interest.

Performance characteristics for the validation of CESeq and WESeq are all greater than 95%. **Conclusion** Based on performance characteristics for the validation of CESeq and WESeq, the assays were incorporated into the clinical menu at The Jackson Laboratory. If CESeq fails to uncover a disease causing variants, the sample is reflexed to WESeq without any additional requirements from the physician. JAX's reflex exome algorithm provides an efficient solution for clinicians and their patients. .

1760F**Genome manipulation of induced pluripotent stem cells using CRISPR/**

Cas9. A.M. Resch¹, J.S. Hsiao², N. Turan¹. 1) Coriell Institute for Medical Research, Camden, NJ; 2) Novartis Institutes for BioMedical Research, Cambridge, MA.

Induced pluripotent stem cells (iPSCs) are unique tools that have been used to model genetic disorders and unravel underlying disease mechanisms. Clustered regularly interspaced short tandem repeats (CRISPR) and CRISPR associated protein 9 (Cas9) have made genome engineering in mammalian cells possible for laboratories with basic molecular biology capabilities. Many cell types are relatively easy to manipulate. However, engineering pluripotent stem cells (PSCs) is still insufficient and difficult to achieve. Here, we provide a method to efficiently obtain genome-edited human PSC lines, minimizing the number of clones to be screened. Many types of genome engineering including knock-in, knock-out, and point mutations can be achieved with this method. Genomic knock-out can be achieved with two CRISPR guide RNAs flanking the target region, while genomic knock-in can be achieved using a plasmid donor with 800-1000 bp homologous arms flanking the knock-in sequence. Point mutations can be achieved using a single-stranded oligonucleotide (ssODN) as a donor template. Temporary antibiotic selection greatly increases the success rate of editing events by eliminating cells without the CRISPR/Cas9 plasmid. We demonstrate that this method is suitable for most laboratories to perform gene-editing in human PSCs to sufficiently obtain modified PSC clones. This technology is being used to create isogenic control lines for disease research. In the past, comparisons between disease and healthy controls were performed using cell lines from separate individuals. This approach is constrained because of differing genetic backgrounds between individuals. The creation of gene-edited isogenic lines that control for differences in genetic background between individuals will provide new tools for disease research.

1761W**Variant identification from whole genome sequencing at the UPMC**

Genome Center. A. Berndt¹, S. Mentch¹, M. Zhuo¹, L. Luo¹, C. Sims¹, J.M. Xavier², P. Empey², A.V. Lee². 1) UPMC Genome Center, Pittsburgh, PA; 2) Institute for Precision Medicine, University of Pittsburgh, Pittsburgh, PA.

Falling costs for high-throughput genome sequencing have shifted our attention from using whole genome sequencing solely in research settings to embedding the services into clinical care. In parallel, cloud and software vendors provide secure and compliant environments to protect personal health information, therefore providing an infrastructure platform that can scale with the demands of high-throughput, whole genome data volumes (storage) and throughput (compute). We will show how UPMC has stood up its own in-house, industrial scale, and clinical-grade genome center over the last 12 months. The UPMC Genome Center integrates with a cloud-only data infrastructure that allows for processing and securing genomic data for clinical processing. We will present data from our extensive validation set for whole genome germline sequencing. We will also present results from two clinical use cases: i) comparing whole genome sequencing and genotyping data for pharmacogenomics variants and ii) rare variant identification from patients in the Neonatal Intensive Care Unit (NICU).

1762T

Enhanced targeted resequencing by optimizing the enrichment technology and DNA insert length. M. Delledonne¹, R. Sirica¹, L. Marcolungo¹, B. Iadarola^{1,2}, E. Zago², A.E. Laria², C.G. Avanzato², A. Mori², S. Benfatto², E. Cosentino¹, M.G. Maturo¹, S. Maestri¹, E. Fortunati¹, M. Rossato¹, L. Xumerle².

1) University of Verona, Verona, Italy; 2) Personal Genomics srl, Verona, Italy.

Whole exome sequencing (WES) is a mature, quite standardized methodology where the most important parameter for WES quality is the coverage (read depth) in order to properly call variants. WES data are typically stored as variant call format (VCF) files, which was developed primarily to represent human genetic variation and not to make causal links between variants and human disease. Standard variant calling pipelines based on VCF are in fact limited in their ability to call the patient genotype in loci homozygous or hemizygous for the reference allele. Absence of a variant in a certain position may be due to the absence of the variant but it may also be due to the failure to detect it because of poor read depth in that region or because of poor mapping quality of the reads due to multiple alignment of those reads on the genome. By taking advantage of the gVCF file format, an extension to the standard VCF format that stores genotypes of variant sites and invariant genomic regions with associated calling scores, we compared coverage and base calling (genotypability) of different whole exome enrichment technologies and found that in many cases there is not a direct relationship with the coverage. Apart from repetitive regions, so far rarely object of investigation in a clinical setting, the lack of a direct relationship is particularly evident - as expected - for genes having pseudogenes, for duplicated genes and for gene families, because of the uncertain read alignment on the genome that determine a low mapping quality of the reads. We observed that increasing the average length of the DNA fragments increases the mapping quality of some regions and improves the base calling for a significant number of genes. The reads mapping near the target (padded regions) expand with longer DNA fragments, but the increased sequencing depth necessary to maintain a good coverage of the targets is negligible for gene panels and, for WES, it is compensated by the high uniformity of some enrichment platforms, which reach base calling saturation at low coverage.

1763F

POP 1: A new sieving matrix for capillary electrophoresis that supports a wide range of applications with a single instrument set-up and with extended stability at ambient temperature. J. Romero. Thermo Fisher Scientific, South San Francisco, CA.

The sieving polymer formulations (POP-4™, POP-6™, and POP-7™) that Thermo Fisher Scientific manufactures for capillary electrophoresis (CE) are designed to provide optimal performance when installed on a CE instrument for up to one-week at ambient temperature (15-30 °C). To assure optimum electrophoretic performance, these formulations must be shipped to customers under refrigerated conditions to prevent the breakdown of urea, which causes an increase in electric current and joule heating and eventually a loss in resolution performance. In an investigation to develop a new universal formulation that could replace POP-4™, POP-6™, and POP-7™ we have explored methods to stabilize performance for an extended period at ambient temperature. There is not an electrophoretically compatible substitute for urea, and urea breakdown cannot be prevented at elevated temperatures. The new POP formulation, POP 1, has been designed to compensate for urea breakdown, allowing a longer lifetime for the reagent on the instrument. POP 1 has also been designed to capture most the electrophoretic properties of POP-4™, POP-6™, and POP-7™ so that a single POP formulation can now be used for multiple applications. For example, POP 1 provides better resolution for fragment analysis applications than POP-4™, with slightly faster run times. Sequencing run times are substantially faster with POP One than with POP-6™, but resolution of short sequencing fragments is significantly better for POP One than for POP-7™. For sequencing applications, with either a 36 cm or 50 cm capillary array, a set of three run modules were developed that provide (1) the longest possible basecalling read-length without consideration of run time, or (2) the best compromise between basecalling read-length and run time, or (3) the fastest run time but with a shorter read-length (> 500 bases). For fragment analysis applications, run modules have been developed to support a range of applications from very fast analysis of short fragments (e.g. single-base extension genotyping) to high-resolution analysis of long fragments up to 1kb long. We will compare the resolution performance of POP 1 to previous POP formulations, provide examples of sequencing and fragment separations with the different run modules. We will also compare the electrophoretic performance of freshly installed POP 1 versus polymer that has resided on the CE instrument for multiple weeks.

1764W

The design of Taiwanese-specific SNP array by Taiwan Biobank. P.E.

Wu^{1,2}, C.Y. Shen^{1,2}. 1) Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan; 2) Taiwan Biobank, Academia Sinica, Taipei, Taiwan.

To understand the relationship between genetics, environmental exposure, diet, and the etiology/progression of chronic disease, the Taiwan Biobank is establishing a scientific infrastructure accessible to biomedical researchers. Through the recruitment and follow-up of a cohort of 200,000 individuals from the general population and another of 100,000 patients with chronic diseases of public health importance from medical centers, the Taiwan Biobank aims to improve the health of future generations and facilitate genomic research in the post-genomic era in Taiwan. Currently, more than 100,000 participants of general population have been recruited, and more than 1,500 participants with chronic diseases have been recruited. Whole-genome genotyping information of more than 24,000 individuals using TWBv1.0 chip (653,291 SNPs, specifically for the Han Chinese in Taiwan) were obtained. The work of whole genome sequencing of 1,500 individuals has been completed. Based on this, Taiwan Biobank has designed a Taiwanese-specific SNP array, which is called TWB v2.0 chip to differentiate the SNP chip (i.e. TWBv1.0) we currently use. There are 717,510 variants included in this chip. In this chip, we use the whole genome sequencing data to select representative 415,000 variants among Taiwanese population, serving as the backbone of this chip. The SNPs used in backbone are the markers showing minor allele frequency (MAF)>5% and covering 0.968 of our genome. Additionally, the coverage rate is 0.702 for the markers showing MAF between 1% and 5%. In TWB v2.0, we also include 105,000 variants with clinical relevance, 81,000 functional variants, 31,000 cancer-associated variants, 50,000 pathogenic variants, the variants showing GWAS significant and the variants of copy number variation. We aim at making the chip both representative to Taiwanese population, and valuable for clinical usage. This chip will be used in the participants of Taiwan Biobank, and summary genetic information will be released in our website (i.e. Taiwan View website) for the public. Taiwan Biobank creates one of the largest publicly accessible databases that is combine genetics, environmental exposure, diet, and clinical information. Based on these data, human geneticists can make use of this valuable and ample information for further research.

1765T

Covaris and Twist fully validated NGS target enrichment library preparation with LE220-plus and oneTUBE. P. Kocjan¹, J. Simon¹, H. Khoja¹, J. Laugharn¹, M. Consugar², C. Thompson², R. Gantt².

1) Covaris, Woburn, MA; 2) Twist, San Francisco, CA.

Covaris, known as the Gold Standard for DNA shearing in NGS, partnered with Twist Bioscience for an exome enrichment library prep workflow. For NGS applications such as WGS and WES, accurate DNA fragmentation is key to achieve high quality libraries ready for sequencing. Covaris produces Focused-ultrasonicators such as the ME220 and LE220-plus which utilize Adaptive Focused Acoustics® (AFA)® technology to mechanically shear nucleic acids and aid in extraction of other biological sample types. Pairing AFA technology with Twist's uniquely designed double-stranded DNA probes offers a solution for a targeted exome enrichment library prep. The LE220-plus, which has high-throughput processing capabilities for up to 96 samples is compatible with the newly developed 96 oneTUBE-10 AFA Plate. This specially engineered polymer allows for low impedance and better transmission of acoustic energy producing tight size distributions necessary for NGS. Designed for use as typical labware, the 96 oneTUBE-10 AFA plate can simplify library prep workflows improving turnaround time and throughput. Utilizing this new consumable, we sheared human genomic DNA evaluating various size distributions to determine the optimal AFA conditions for Twist library construction. For lower throughput applications, the ME220 is the ideal solution processing up to 8 samples in one batch. The 8 strip formats of the oneTUBE-10 or microTUBE-50 allow the flexibility necessary for Twist workflows. Covaris and Twist in combination offer a fully validated library prep solution for targeted exome sequencing. Producing highly accurate and reproducible results for DNA fragmentation with AFA®, Covaris offers a wide variety of Focused-ultrasonicators such as the LE220-plus and ME220. Paired with Twist's superior enrichment efficiency, Covaris and Twist produce the highest quality results for targeted exome library preparation. .

1766F

NEBNext Ultra II FS DNA: A robust, enzyme-based DNA library preparation method. *K. McKay, L. Apone, V. Panchapakesa, B. Langhorst, T. Shtatland, C. Song, P. Liu, C. Sumner, C. Rozzi, M. Am, D. Munafo, F. Stewart, E. Dimalanta, T. Davis.* New England Biolabs, Inc., Ipswich, MA.

Next Generation Sequencing technology has advanced our knowledge of the molecular mechanisms involved in disease development and progression. This understanding has led to advancements in target identification and drug discovery in therapeutic medicine. The extremely high data output of current sequencing instruments has shifted the bottleneck from sequencing to optimal, high-throughput sample preparation in NGS workflows. To meet this challenge, we have developed a robust library construction method that integrates enzyme-based DNA fragmentation with end repair and dA-tailing in a single tube. This method eliminates the need for expensive equipment to fragment DNA; moreover, the optimized workflow reduces the numerous cleanup and liquid transfer steps, reducing time, cost and errors associated with library construction. Genomic DNA with various GC contents was used to construct Illumina libraries. DNA (100pg - 500ng) was fragmented, end repaired and dA-tailed in a single step followed by adaptor ligation in the same tube. PCR amplified and PCR-free libraries were sequenced, reads aligned to the appropriate reference genome, and quality metrics generated using Picard tools. The high library yields obtained with enzyme-based shearing enable PCR-free workflows from as little as 50ng starting material. Sequencing quality of these libraries generated show similar coverage uniformity and fragment size distribution, with minimal GC bias. To demonstrate that genome integrity is maintained with the Ultra II FS, libraries were constructed from human genomic DNA spiked with DNA reference standards containing multiple common cancer mutations (Horizon Discovery, Inc.) at defined frequencies. Target enrichment and deep sequencing of a panel of cancer genes were performed, and variant frequency of input DNA and enriched FS libraries were quantified using digital droplet PCR. Our data showed that Ultra II FS maintains mutation abundance in samples without introducing bias. In addition, Ultra II FS enables high-quality library construction from a broad range of DNA quantities and qualities. Sequencing metrics of FS libraries are similar or superior, to those produced with mechanically sheared DNA. These advances will allow greater use and adoption of NGS technologies in clinical and diagnostic settings.

1767W

Rapid isolation (<2 h) of ultra-high molecular weight genomic DNA from fresh/frozen human blood and cultured cells/frozen cell pellets without the use of agarose plugs. *H. Sadowski, C. Proskow, A. Files, K. Pham, G. Pljevaljcic, M. Borodkin.* Bionano Genomics, San Diego, CA.

Successful optical mapping of genomic DNA on the Bionano Genomics Saphyr® system for genome assembly or structural variation detection relies on starting with ultra-high molecular weight (UHMW) DNA. To achieve this, we have established methods for the isolation of genomic DNA from various sources that involves the embedding of material in agarose plugs and overnight lysis *in situ* with detergent and proteinase K. After extensive washing on day 2, the agarose plug is melted, treated with beta agarase to digest agarose polymers to monomers and the liberated DNA is finally subjected to drop dialysis. The UHMW DNA is ready for quantification and labeling on day 3. This "plug lysis" method is extremely robust, but it is also labor intensive, difficult to automate, lengthy and expensive. To address these shortcomings, we have developed a solution based lysis method which we have coupled with a purification step that leverages a novel process to bind, wash and elute UHMW genomic DNA. This entire protocol can be conducted in less than two hours on a batch of samples. The eluted material contains high quality DNA that is clean enough for enzymatic labeling using nick-label-repair-stain (NLRS) or direct labeling and staining (DLS). The resulting metrics of this labeled DNA on a Saphyr Chip® are comparable to labeled DNA isolated by the traditional "plug lysis" protocol. We have validated protocols for fresh and frozen blood as well fresh cultured cells and frozen cell pellets and are developing protocols for plant and animal tissue. This protocol is automatable providing the needed solution for researchers needing to purify DNA from hundreds to thousands of individuals per year.

1768T

Single-cell protein and gene expression profiling of stem memory T cells by Ab-seq. M. Nakamoto¹, C. Chang¹, E. Shum¹, J. Martin¹, J. Ghadiali¹, M. Corselli², D. Jensen¹, J. Hu¹, D. Rosenfeld¹, C. Fan¹. 1) BD Biosciences; 2) BDTI.

High-throughput single-cell RNA sequencing recently emerged as a powerful tool to profile complex and heterogeneous cell populations and dynamics. However, the lack of information on protein expression can make identifying cell types that have conventionally been defined by cell-surface markers challenging, as mRNA and protein expression are often not tightly correlated. T cells in particular contain relatively low abundance of transcripts, and different T-cell subsets often exhibit highly similar transcriptional profiles. Here we utilized BD Ab-seq, a novel protein sequencing technology enabled by oligo-conjugated antibodies, to study stem memory T cells (Tscm), a rare long-lasting memory T-cell population with stem cell-like properties. An oligo-conjugated antibody panel against immune-relevant cell-surface markers was created and used for protein profiling alongside gene expression profiling in single cells. From human PBMCs, we enriched for Tscms, naïve T cells, central memory T cells and effector memory T cells by FACS. The sorted samples were labeled with antibodies barcoded with unique sample IDs, which allowed us to pool multiple samples into one single-cell experiment, and de-multiplex the samples after sequencing. The antibody-specific oligos and sample ID-specific oligos were captured, amplified and sequenced alongside mRNAs in a single workflow on BD Rhapsody™, a massively-parallel single-cell analysis system. The resulting sequencing data provided a combined output of gene expression, protein expression and sample identity. The detection of protein expression by BD Ab-seq was highly sensitive and specific, and we observed consistent protein expression patterns when comparing BD Ab-seq data to FACS data of the same samples. The addition of protein marker expression provided more distinct and robust clustering of single-cell data compared to clustering by gene expression alone. We were also able to correlate gene expression differences between the four T-cell subsets with protein marker expression at the single-cell level to examine heterogeneity within the Tscm subset. Our study demonstrates the power of combining BD Ab-seq with RNA-seq to gain a more comprehensive understanding of cell lineage and function at the single-cell level.

1769F

Deep mutational scanning of the β_2 -adrenergic receptor. E.M. Jones, N.B. Lubock, D. Cancilla, M. Satyadi, J. Davis, R. Jajoo, S. Kosuri. Dept. of Chemistry and Biochemistry, Molecular Biology Institute, University of California Los Angeles.

G protein-coupled receptors (GPCRs) are principal mediators of cell signaling in humans. GPCRs are widely implicated in disease and targeted by roughly 30% of all FDA approved drugs. How their genetic variation affects drug/ligand binding and signaling remains unknown on a comprehensive scale. The difficulty to interrogate GPCR structure through traditional methods further obfuscates this relationship. Deep mutational scanning presents a way to probe the fitness effect of every missense mutation possible for a protein and understand the contribution of each individual residue to function. We have leveraged advances in DNA synthesis, gene editing, and next-generation sequencing to create a platform able to measure the difference in drug-induced G-protein signaling of all beta-2 adrenergic receptor missense variants (~8000) in multiplex. Our assay enables us to understand broad structural rules for regions of the protein, comprehensively understand the fitness of previously known and novel motifs, and relate this data to genetic variation in humans. Additionally, we found simple algorithms such as k-means clustering groups our mutations into distinct classes that map onto the structure in a functionally relevant manner.

1770W

CleanPlex® UMI with unique dual index increases the overall usable reads available for ultralow-frequency variant calling. *L. Lee, Z. Liu, L. Y. Liu, G. Liu, J. Liu, L.J. Zhang, L. Lin.* Paragon Genomics, Hayward, CA.

With recent advances in next generation sequencing, larger capacity sequencers are pushing for higher multiplexing of samples per sequencing run. With this increase in paralleling power, there is a growing need for not only additional indices, but also robust and high performing indices that minimizes errors associated with certain sequencing chemistries. It was recently discovered that high capacity sequencers are susceptible to "index hopping" that can increase mis-assignment and also lead to an increase in false positives. With increasing interest in ultra-low allele frequency detection, such as in liquid biopsy, it is crucial that the index itself does not contribute to errors that will interfere with accurate calling. To mitigate these issues, Paragon Genomics created unique dual indices (UDI) or distinct pairs such that show a significant improvement in usable double-stranded barcoding reads compared to libraries with the traditional overlapping i5 and i7 combination indexing method. The CleanPlex® UMI technology's fast and simple libraries made with Paragon Genomics's UDI allows allele frequency detection down to 0.1%, utilizing less read depth.

1771T

Taiwan biobank surveys of gut microbiota. *M. Su, H. Chu, T. Yu, C. Chang, C. Chang, C. Lin, C. Hsiung, C. Shen.* Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan.

The gut microbiota is a hidden organ because it serves many biological functions are equivalent or more important than other organs. It can affect human behavior and complex diseases such as inflammatory bowel diseases, type 2 diabetes, and colorectal cancer. Various clinical studies have demonstrated that fecal microbiota transplantation (FMT) could become a viable therapeutic option. Biobanking gut microbes along with their genetic and metabolic profiles will become important to a wide range of diseases. We have observed many confounding issues through sample collection to data analysis in biobanking microbes. In this study, we evaluate the effect of sample storage condition, sample collecting time, sample collecting kit, sequence target region and data analysis pipelines. Results show that sample storage condition is the most important confounding factor of a microbiota study. If fecal samples cannot be frozen at -80 degree centigrade immediately, store the samples in a stabilization reagent is optimized for sample collection. The standard approach of 16S data analysis is clustering reads by sequence similarity into operational taxonomic units (OTUs). Denoising approaches have been proposed to improve the resolution of the 16S analysis. A comparative analysis of clustering and denoising methods showed that clustering method impacts on the number of OTUs and denoising algorithm influence more on taxonomic affiliations. The discrepancy in the number of OTUs between partial 16S rRNA and full-length 16S rRNA need further validations to determine whether the full-length bacterial 16S rRNA will provide a better bacterial classification.

1772F

Evaluation of methods for the extraction of microbial DNA from vaginal swabs used for microbiome studies. S. Tomei, V. Mattei, S. Murugesan, M. Al Hashmi, P. Singh, R. Mathew, N. James, A. Lakshmanan, A. Terranegra, S. Al Khodor. Sidra Medical and Research Center, Doha, Qatar.

Background. The composition of the microbiome in various human body sites plays an important role in health and disease. The advancement of sequencing technologies has made the assessment of the composition of the microbiota possible through microbial DNA extraction and sequencing. Therefore, it is of a paramount importance to select a sensitive and reproducible DNA extraction method, that facilitates isolation of microbial DNA with a sufficient quantity and purity, from microbial species living in the vaginal environment. In this study, we have evaluated four different DNA extraction protocols from self-collected vaginal swabs. **Methods.** Five healthy female volunteers were enrolled in the study. Each donor was asked to self-collect 4 samples using Copan ESwab. DNA was extracted using Qiagen DNeasy kit and three modified protocols of the MoBio PowerSoil kit. DNA quantity and integrity was checked through Nanodrop and LabChip GX. DNA was further tested through quantitative real-time PCR (qPCR) and 16S sequencing. Vaginal microbiota diversities were determined using MiSeq-Illumina high-throughput sequencing of bacterial 16S rDNA V1-V3 fingerprint. Sequencing data were analyzed using QIIME pipeline. **Results.** Qiagen DNeasy protocol resulted in the highest DNA yield as compared to the modified protocols of MoBio Powersoil kit. The size of the DNA extracted using each protocol was ~40kb. Qiagen DNeasy protocol gave the highest Genomic Quality Score (average $\pm$ standard deviation: 4.24 ± 0.36), followed by the different MoBio Powersoil protocols. A substantial variability in microbial DNA abundance was found across the different protocols tested. As expected, the vaginal microbiota of the healthy volunteers was dominated by *Lactobacillus species*. MoBio Powersoil kit provided a significantly higher alpha diversity as compared to the Qiagen DNeasy kit. On the other hand, beta diversity measures did not reveal any significant cluster changes among the four protocols tested. **Conclusion.** Using all four DNA extraction protocols described in this study, we were able to isolate microbial DNA from the vaginal swabs. However, the results vary based on the downstream application needs. We show that, the Qiagen DNeasy is the best extraction protocol to use in terms of DNA yield and quality. However, the modified MoBio PowerSoil protocols showed higher microbial diversities as compared to the standard protocol.

1773W

Streamlining single-cell next-generation sequencing of small non-coding RNA. J.M. Henderson¹, D. Keys¹, S. Shore², R. Khnouf^{3,4}, C.M. Han^{3,5}, R.I. Hogrefe¹, A.P. McCaffrey¹, S.A. Munro^{3,6}, H. Shintaku⁷, J.G. Santiago³. 1) Research and Development, TriLink BioTechnologies, San Diego, CA; 2) Singular Genomics, San Diego, CA; 3) Department of Mechanical Engineering, Stanford University, Stanford, CA; 4) Department of Biomedical Engineering, Jordan University of Science and Technology. Irbid, Jordan; 5) National Institute of Standards and Technology (NIST), Gaithersburg, MD; 6) Department of Bioengineering, Stanford University, Stanford, CA; 7) RIKEN Cluster for Pioneering Research, Saitama Japan.

Differential expression of small RNA (sRNA) is crucial for many cellular processes such as development and regulation, however current techniques to study such events via massively parallel sequencing are limited by difficulties associated with tagging low amounts of short sequences and often rely on bulk cell analysis. It is becoming more apparent that variations in gene expression within a single cell can go undetected within a pooled population and may lead to false negative results, especially if assaying for rarities such as circulating tumor cells. Therefore, the need for improved small RNA next generation sequencing (sRNA-Seq) of single cells is imperative. One of the main challenges in sRNA-Seq occurs during library preparation when fixed-sequence adapters ligate directly to each other instead of to the small RNA within the sample. Because the adapters are complementary to primers used for reverse transcription and PCR, empty libraries known as "adapter dimer" will also amplify. Adapter dimer may be only 20 nucleotides shorter than the tagged small RNA, making it difficult to separate away from the library of interest. This formation of adapter dimer becomes more prevalent when lower amounts of sRNA are available in the reaction. If this contaminant is not efficiently removed then adapter sequences will monopolize NGS reads. We reasoned that reducing adapter dimer would lead to improved sensitivity at low RNA input levels, allowing for sRNA-Seq of single cells. In this work we present a combination of novel technologies; chemically modified CleanTag adapters for dimer suppression and single-cell microfluidic on-chip isotachopheresis (ITP) for separation of nucleus and cytoplasm. We compared CleanTag sRNA libraries prepared from bulk K562 cells, single K562 cells lysed in Triton X-100, and ITP separated cytoplasmic RNA from single K562 cells. Our results indicate that combining the streamlined methods of microfluidic chip-based single cell isolation with CleanTag library preparation will be invaluable in automating high throughput research and diagnostics in the small RNA field. .

1774T

Returning unanticipated genomic results in a hospital-based research biobank. R.C. Green^{1,2,3,4}, C.L. Blout¹, M.S. Lebo^{2,3,4,5}, J.B. Krier^{1,2,3}, J. Smoller^{2,6}, E.W. Karlson^{1,3}, L. Mahanta², K. O'Brien², K.D. Christensen^{1,3}, N. Boutin², S.T. Weiss^{2,3}. 1) Department of Medicine, Brigham and Women's Hospital, Boston, MA; 2) Partners HealthCare Personalized Medicine, Cambridge, MA; 3) Harvard Medical School, Boston, MA; 4) Broad Institute, Cambridge, MA; 5) Department of Pathology, Brigham and Women's Hospital, Boston, MA; 6) Department of Psychiatry, Massachusetts General Hospital, Boston, MA.

Introduction: Many healthcare systems are analyzing genetic data from biobanks along with patients' electronic health records (EHRs). We present early results from Partners Biobank on return of genomic results (gRoR). **Methods:** We elected to return pathogenic/likely pathogenic (P/LP) variants in ACMG59 genes from genotyping (GT) on the Illumina MEGA (N=15,000 to date) or by next-generation gene sequencing (GS) (N=2500 to date). Participants with P/LP variants first received orientation to the non-CLIA finding from a study GC and were offered CLIA confirmation paid by the study. Those electing to confirm were transitioned to a clinical work stream to learn the exact variant from a clinician trained in genetics and the result entered into the EHR and communicated to their provider. **Results:** GT quality control yielded a false positive rate of 37%, 2/3 of which were caused by multi-allelic sites. Of 15,000, GT yielded reportable results in 108 participants (0.72%), while GS of 2,500 samples yielded reportable results in 58 subjects (2.3%), the difference reflecting probes in GT limited to previously known variation and difficulties genotyping rare variation. Together 166 individuals were identified, with results related to cancer (88), cardiology (44), cholesterol (27) or other phenotypes (7). Of these, 42 already had EHR documentation of the result, 9 were deceased, 11 were lost to follow-up and 34 are in process. Of 59 participants so far reached and offered confirmation, 11 (18.6%) declined to proceed. For an additional 6 subjects, P/LP variants in the ACMG59 genes were determined to be somatic due to previously diagnosed cancer (1) or clonal hematopoiesis (5). **Conclusions:** GT is less expensive but less sensitive or specific than GS. A substantial number of biobank participants already had the detected variant in their EHR or were unreachable, and among those contacted nearly one-fifth declined confirmation. Some variants classified as P/LP required deep sequencing and clinical correlation to recognize they were somatic rather than germline. Transition to a specialized clinical workstream as a requirement for confirmatory disclosure may help avoid misunderstanding or mismanagement around results delivered directly to the participant or their provider. Additional data on technical considerations and participant outcomes will be presented. Lessons learned from these experiences may help inform gRoR for other biobanks including the All of Us Research Program.

1775F

Genetic profiling of ADME markers using PharmacoScan™ solution across blood, saliva, and buccal cell-derived genomic DNA. D. Lizarraga¹, J. Law¹, K. Tsai¹, L. He², S. Larocca², X. Fang², R. Setterquist², C. Bruckner¹, R. Varma¹, J. Gollub¹, D. Black¹, M. Shapero¹. 1) Thermo Fisher Scientific, 3450 Central Expressway, Santa Clara, CA 95051; 2) Thermo Fisher Scientific, 2130 Woodward Street, Austin, TX 78744.

Genetic variation in drug absorption, distribution, metabolism, and excretion (ADME) genes forms the basis for each individual's clinical response to xenobiotics. The Applied Biosystems™ PharmacoScan™ Solution enables comprehensive, accurate genotyping on a DNA microarray of ~4,600 markers in over 1,000 genes involved in drug-metabolizing enzymes and transporters research. As personalized medicine becomes increasingly important in patient care, this study set out to expand the application of this platform to various DNA sample types and extraction methodologies. To this end, the characterization and verification of genotyping performance of saliva and buccal cell sample types with PharmacoScan™ Assay 96-Array Format and PharmacoScan™ Assay 24-Array Format plates was carried out. The study verified the use of both magnetic, bead-based high-throughput DNA isolation and manual, precipitation-based DNA extraction methods. Both extraction methods provided high-quality genomic DNA (gDNA) that is compatible with PharmacoScan Assay. This study demonstrated similar genotyping performance between DNA from oral samples versus whole blood. All sample types analyzed passed the PharmacoScan Assay in-process quality control (QC) checks and met the genotyping and copy number performance metrics. Results showed that saliva and buccal-derived DNA perform as well as whole blood in both the PharmacoScan™ Assay 96-Array Format using the Beckman Biomek FXP automated target preparation workflow and the PharmacoScan™ Assay 24-Array Format using manually prepared DNA target. This study supports the use of gDNA derived from oral samples, such as saliva and buccal cells, in the PharmacoScan™ Assay. For Research Use Only. Not for use in diagnostic procedures.

1776W

Longitudinal individualized saliva omics profiling. *G.I. Mias^{1,2}*. 1) Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI; 2) Institute for Quantitative Health Science and Engineering, East Lansing, MI.

Precision individualized wellness can now utilize state-of-the-art sequencing technology and computational capabilities to monitor hundreds of thousands of molecular components. As we move closer to clinical implementations, a realistic medical approach must supplement static genomic data (the current focus of most investigations) with the dynamics of disease, as reflected by longitudinal monitoring of patients. Integrating multi-omics dynamics (previously transcriptomes, proteomes, and now including extracellular vesicles and microbiomes) can reveal individualized deviations from a healthy baseline as these occur. We will present updates from our individualized pneumococcal vaccination response project, where we integrated longitudinal profiling in saliva and peripheral blood mononuclear cells (PBMCs), including extracellular vesicle-like (EV) composition in saliva. We focus on detecting adverse events or gradual changes, that may lead to early treatment or prevention, and additionally on the utility of saliva for continuous non-invasive diagnostic monitoring. We have expanded our saliva-based dataset on an individual to include 104 timepoint evaluation of EV RNA, and 12 timepoints on other individuals undergoing the same immunization (pneumococcal vaccine, PPSV23). The small RNA-seq from saliva EVs supplements our proteomics (mass spectrometry) and mRNA-sequencing in saliva and PBMCs at various timepoints (100+ for saliva, 30 for PBMCs). The EVs provided information on miRNAs, and endogenous or exogenous genome contributions (microbiome components), as these vary with vaccination response, indicative of both innate and adaptive responses: from 104 timepoints we detected on average (in reads per timepoint) 2×10^5 miRNA, 25×10^3 miRNA precursors, 26×10^3 piRNA, 64×10^3 exogenous (non-human) miRNA, and 4×10^6 exogenous genome reads. All omics were integrated using MathlOmics to identify corresponding longitudinal behavior across the different modalities in time-frames that match the phenotype (including fever within hours of vaccination, and innate/adaptive responses over multiple days). The full dataset is being made publicly available and can be used to model longitudinal personalized omics with saliva. This research continues to be to our knowledge the only systematic implementation of personalized wellness monitoring that employs the same controlled immune activation, pneumococcal vaccination, across all subjects with non-invasive saliva monitoring.

1777T

Combined mRNA and microRNA NGS library prep enables a more complete characterization of cell-free RNA. *K.D. Allen, A.R. Morris, S.N. Piehl, K. Kushiro, A. Kothandaraman*. Applied Genomics, PerkinElmer, Austin, TX.

Historically, most RNA-seq experiments have been performed with the goal of studying mRNAs. Recently, deep sequencing of small RNAs, in particular miRNAs, has gained in popularity as researchers have begun to appreciate that miRNA expression can also be used to create gene signatures and that miRNAs have significant impact on regulation of gene expression. Researchers are increasingly interested in profiling both mRNA and miRNA content of a sample, but no simple method currently exists for a combined library preparation. The methods currently in existence typically require splitting a sample into two fractions, isolating mRNA in one fraction, fragmenting the mRNA, treating the fragmented mRNA with T4 PNK, then recombining the fractions and using a small RNA library preparation protocol with direct ligation of adapters to RNA molecules. Here, a combined mRNA/miRNA library preparation protocol is described that is streamlined, does not require rRNA depletion or poly(A) selection, and does not require splitting the sample into separate fractions. It relies upon a novel technique in which poly(A)-tailed RNA species are selectively reverse transcribed, then sheared by RNase H into fragments of useful length. Sequencing results show high concordance of RNA abundance between this method and traditional mRNA-Seq protocols, as well as high replicability. This method will be especially valuable to researchers studying cell-free, or circulating, RNA. Both mRNA and miRNA biomarkers are known to be present in biofluids, however due to their low concentration, no practical method yet exists for sequencing both RNA species simultaneously. As a proof of principle, we show that this previously presented method can also be employed to generate combined mRNA/miRNA libraries from human serum and plasma. We expect that this method will become a powerful new tool for researchers interested in biomarker discovery in this first report on data generated using this method.

1778F

Encapsulating single cells with barcoded beads on the Nadia Instrument.

M. Breteau, J. Fischer, N. Shakir. Dolomite Bio, Royston, England, United Kingdom.

Single-cell RNA-sequencing (scRNA-seq) is a cutting edge innovative technique that provides unprecedented insight into the expression patterns of thousands of single cells. Using droplet microfluidic technology, thousands of single cells can now be individually analyzed enabling the discovery of previously unidentified or cryptic cell types in a tissue of interest and the characterization of individual cells within biologically important processes. Using cutting edge microfluidic technology this abstract reports the results obtained by encapsulating single cells with barcoded mRNA beads. Using the Drop-Seq protocol developed by Macosko et al., single cells are encapsulated alongside a single bead within a droplet. These beads have surfaces coated with DNA-oligos containing a (dT)30 stretch allowing for the capture of polyadenylated mRNA which is subsequently reverse transcribed into cDNA. Each bead also carries a unique DNA barcode that identifies not only individual cells but also individual mRNA transcripts derived from any given cell. It is desirable that a single cell be encapsulated alongside a single bead within a droplet. This can be achieved through the dilution of cell and bead suspensions, but there is a fine balance to be struck between quality (fewer cells but data of higher quality) and quantity (higher throughput but data of lower quality). In this protocol, cell and bead suspensions are adjusted to achieve encapsulation in which 10% of droplets will contain a bead (i.e. 1 in 10 droplets) and 5% of droplets will contain a cell (i.e. 1 in 20 droplets). This translates to 10% of cells being encapsulated with a bead and very low cell doublet rates. A 15-minute run on the Nadia instrument, with 250 µl of bead suspension and 250 µl of cell suspension will result in the production of 6,000 individual cellular transcriptomes. Cell encapsulation is fast becoming a mature technology having transitioned from early prototypes into robust instruments suitable for use in biological research that is at the cutting-edge. The results indicate that using the drop-seq protocol on this single cell microfluidic platform high quality single cell libraries were produced in a high throughput manner. This reports results obtained from the encapsulation of cells with barcoded beads for single-cell RNA sequencing using the Dolomite Bio Nadia Instrument.

1779W

Analysis of single cell transcriptome from PBMCs using ddSeq with different sequencing conditions.

H. Kim¹, E. Cho¹, H. Lee², S. Yun³, A.K. Cashion¹, J. Gill¹. 1) National Institute of Nursing Research, Bethesda, MD; 2) College of Nursing, Yonsei University, Seoul, Korea; 3) Yotta Biomed, Bethesda, MD.

Use of well characterized and reproducible biomarkers can help monitor and improve clinical outcomes across a variety of treatments and populations. However, the validity of biomarkers cannot be easily evaluated, especially when subtle changes in small populations of cells are masked. Single cell assays can provide critical information for the development of biomarker assays for clinical applications. However, reliable methodologies need to be developed and validated to use this novel technique. In this study, we demonstrate how to prepare single cell libraries of PBMCs by using the ddSeq, a single cell isolator based on droplet technology followed by sequencing with different conditions, and we analyze the data with BaseSpace and Seurat. We used cryopreserved human PBMCs from a clinical patient at 3 different time points with ~75% of cell viability. Concentration of cells was adjusted to 3,000 cells/µl for PBMCs. One same sample was loaded on all 4 wells of a ddSeq cartridge. They were merged before tagmentation considering relatively low quantity of transcriptome of PBMCs. Based on QC, final concentration of the sample was 0.43 nM to 0.89 nM. Libraries of the sample were sequenced on both Illumina HiSeq 2500 and NovaSeq. FASTQ files were uploaded and processed on Illumina's BaseSpace to get UMI count table using SureCell RNA Single-Cell (version 1.2.0). Background cells were removed by filtering with a knee plot. We obtained 1,446 to 5,425 cells with 16 to 170 million reads, which is ~11,000 reads to ~42,000 reads per cell. Detected variable genes expressed in PBMCs ranged from 285 to 568. Seurat was used to perform further quality control and clustering analysis. Seven to 8 different cell types were clustered per sample by tSNE, which is comparable to another study with 1-2 million reads per cell. Our results suggest that 1) library preparation of ddSeq single cell assay using PBMCs needs to be modified from other types of cells, 2) a relatively low read number per cell along with a small number of detected genes are enough for classification of PBMCs based on their transcriptome profile, and 3) increasing read number per cell detects more genes, which may be critical in biomarker findings.

1780T

High-throughput single nuclei RNA-seq from limited amounts of frozen human and mouse tissues. A.M. Raman¹, S. Preissl², Z. Cheng³, S. Chen³, Z. Ye⁴, S. Kuan¹, K. Zhang⁵, C. Barr¹, C. Chen⁵, B. Ren^{1,2,6}. 1) Ludwig Institute for Cancer Research, 9500 Gilman Drive, La Jolla, CA 92093, USA; 2) Center for Epigenomics, University of California, San Diego, 9500 Gilman Drive, MC 0653, La Jolla, California 92093 -0653; 3) Department of Bioengineering, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA; 4) Krembil Research Institute University Health Network, The Hospital for Sick Children, The University of Toronto, Krembil Discovery Tower, 60 Leonard Ave. 8KD-412, Toronto, ON M5T 2S8, Canada; 5) Department of Neurosurgery, University of Minnesota, Minneapolis, D429 Mayo Memorial Building, 420 Delaware Street S.E., MMC 96MN 55455, United States; 6) Department of Cellular and Molecular Medicine, University of California, San Diego, School of Medicine, 9500 Gilman Drive, La Jolla, CA 92093, USA.

While single cell RNA-seq is a powerful tool to dissect tissue heterogeneity, it is difficult, and sometimes even impossible, to isolate intact cells from frozen tissue samples. Additionally, current approaches to analyze such samples are limited by throughput or requirement of high input material. To overcome these hurdles, we have optimized nuclei preparation from frozen tissues, and combined it with the droplet-based single cell RNA-seq platform from 10x Genomics. Nuclei contain only 10% of a cell's mRNA; therefore, single nuclei RNA-seq (snRNA-seq) data are exceptionally sparse with a median number of 260-770 detected genes per nucleus. To overcome this, we optimized a computational analysis workflow to enable robust analysis and clustering of snRNA-seq profiles. In total, we profiled 33,252 single nuclei from a wide variety of tissue types, including mouse forebrain, human frontal cortex and hippocampus, human cardiac left and right ventricles, and glioblastoma sections from a human brain. Using our analytical framework we digitally dissected cellular composition and identified: differences between ventricle specific cardiomyocytes in human ventricles, a neural stem-cell like population in the mouse forebrain; and grey and white matter specific astrocyte populations, as well as hippocampal and frontal cortex specific OPCs (oligodendrocyte progenitor cells) in the human brain. Coupling our analysis of these normal human frontal cortex and hippocampus samples with inferred CNV calls from our snRNA-seq data, we were then able to differentiate tumor from normal cells in our glioblastoma sections, and characterize intratumoral heterogeneity. The presented strategy enables the characterization of cellular heterogeneity from limited amounts of archived snap frozen tissue samples.

1781F

Single cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. C. Roco¹, A. Rosenberg¹, R. Muscat¹, A. Kuchina¹, P. Sample¹, Z. Yao², L. Gray², D. Peeler¹, S. Mukherjee¹, W. Chen¹, S. Pun¹, D. Sellers¹, B. Tasic², G. Seelig¹. 1) University of Washington, Seattle, WA; 2) Allen Institute for Brain Science, Seattle, WA.

To facilitate scalable profiling of single cells, we developed Split Pool Ligation-based Transcriptome sequencing (SPLIT-seq), a single-cell RNA-seq (scRNA-seq) method that labels the cellular origin of RNA through combinatorial barcoding. SPLIT-seq is compatible with fixed cells or nuclei, allows efficient sample multiplexing and requires no customized equipment. We used SPLIT-seq to analyze 156,049 single-nucleus transcriptomes from postnatal day 2 and 11 mouse brains and spinal cords. Over 100 cell types were identified, with gene expression patterns corresponding to cellular function, regional specificity, and stage of differentiation. Pseudotime analysis revealed transcriptional programs driving four developmental lineages, providing a snapshot of early postnatal development in the murine central nervous system. SPLIT-seq provides a path towards comprehensive single-cell transcriptomic analysis of other similarly complex multicellular systems. We will present on new developments to SPLIT-seq, including significant increases to sample multiplexing and gene detection as well as new applications where we have applied the technology in fields including oncology, cardiology, and immunology.

1782W

Improving transcriptome profiling for single cell and low input RNA. J.C. Sanford¹, K. Krishnan¹, Y. Bei¹, J.G. Borgaro¹, S. Guan¹, V. Panchapakesa¹, K. McKay¹, L. Apone¹, T. Shtatland¹, B.W. Langhorst¹, M. Arn¹, C. Sumner¹, D.R. Pattabiraman², T.C. Evans, Jr.¹, E. Dimalanta¹, N.M. Nichols¹, T. Davis¹. 1) New England Biolabs, Ipswich, MA 01938; 2) Geisel School of Medicine at Dartmouth, Hanover, NH 03755.

RNA sequencing—determining gene expression profiles of diverse tissues, cell types, developmental stages and diseases—is a widely used method with utility in basic and clinical sciences. However, RNA sequencing studies often use heterogeneous populations of thousands of cells in order to adequately profile gene transcripts. As such, significant biological variations among individual cells are obscured by the overall RNA pool. To overcome this limitation, single-cell RNA-seq is emerging as a powerful approach to characterize gene expression heterogeneity within phenotypically identical or complex cell populations and in rare cell types. Here, we have developed a simple and robust single-cell, low input RNA-seq workflow to generate full-length cDNAs that can be easily prepared for Illumina sequencing when combined with the NEBNext Ultra II FS DNA Library Preparation Kit with enzymatic fragmentation. Using this approach, we generated libraries from a variety of input material including Universal Human Reference RNA (2.5pg – 200ng), as well as single cells from cultured cell lines and mouse primary cells, and sequenced on the Illumina Nextseq 500. High quality sequencing data was obtained from all samples, with excellent gene body coverage observed. The workflow is highly sensitivity as demonstrated by detection of a high number of transcripts and single copy of RNA spike-ins (ERCC). The data showed strong gene expression correlation (Pearson $r > 0.9$) between RNA inputs that span over five orders of magnitude and in cultured cells (single vs hundreds). Analysis of primary cells could successfully distinguish two developmental lineages from 8-week old mouse mammary glands, tracing them to the basal and luminal developmental lineages, highlighting the high sensitivity of the protocol. We have developed a highly robust and sensitive method that consistently generates high quality sequencing data from single cells or low input RNA. This workflow is streamlined, amenable to automation, and is accessible to high throughput studies. We envision this method facilitating novel discoveries in the area of low-input transcriptome applications across various platforms.

1783T

Supporting single-cell RNA-seq at scale in the genomics platform at the Broad Institute. C. Walsh, B. Granger, A. Hollinger, T. Desmet, T. Bowers, N. Lennon. Broad Institute, Cambridge, MA., USA.

Single-cell RNA-sequencing (scRNA-seq) is a powerful technique to study gene expression, cellular heterogeneity, and delineation of cell states. Interest in single-cell research has risen to an all-time high, as illustrated by the launch of the Human Cell Atlas project. Traditionally, single-cell sample preparation and sequencing is conducted by individual research labs, at relatively low scale, with little support for end-to-end processing and analysis. As projects become larger and more collaborative, the need for reproducibility and quality at scale becomes increasingly important. In response, we have developed a portfolio of single-cell services that utilize automated workflows, and integrated sample tracking to support 10X Genomics Chromium Single-Cell 3' single-cell, and a modified SMART-Seq2 mRNA library construction, sequencing, and analysis at scale. Major design challenges we encountered scaling SMART-Seq2 included managing the minimal input concentrations, high viscosity master mixes, low volumes and necessary dead volumes for full-scale automation. This was achieved using a variety of in-house created liquid classes, labware, and protocol design. Data delivery and analyses from single cell sequencing is made available through our cloud based platform, Firecloud. For each cell, we run a best practices QC and analysis workflow that mirrors the methods being made publically available via the Human Cell Atlas consortium, consisting of alignment with HISAT2, sequencing quality assessment with Picard tools, and determination of expression with RSEM. The outputs from these QC tools is aggregated and run through a second workflow to visualize the results at the level of the plate allowing for troubleshooting of lab processes and the identification of systematic biases.

1784F

Discovery of phylogenetically informative Y-chromosomal variants in Argentine Amerindians Q haplogroup using next-generation sequencing data. *P.B. Paz Sepúlveda¹, P. Hallast², J.J. Zaiat³, E.L. Alfaro⁴, J.E. Dipierri⁵, C.M. Bravi^{1,5}, M. Muzzio^{1,5}, G. Bailliet¹.* 1) Instituto Multidisciplinario de Biología Celular, La Plata, Buenos Aires, Argentina; 2) University of Tartu, Institute of Biomedicine and Translational Medicine, Faculty of Medicine, Tartu, Estonia. Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, UK; 3) Instituto de Química Biológica de la Facultad de Ciencias Exactas y Naturales (IQUIBICEN), Plataforma Bioinformática Argentina (BIA), CONICET, Universidad de Buenos Aires. Ciudad Autónoma de Buenos Aires, Argentina; 4) Instituto de Ecorregiones Andinas (INECOA), UNJu-CONICET, Instituto de Biología de la Altura, Universidad Nacional de Jujuy, San Salvador de Jujuy, Jujuy, Argentina; 5) Facultad de Ciencias Naturales y Museo, Universidad Nacional de La Plata, La Plata, Buenos Aires, Argentina.

The human Y-chromosomal Haplogroup Q is the most frequent haplogroup in Amerindian people of South America. Haplogroup Q-M3 and Q-Z780 are subclades of Haplogroup Q-L54. Haplogroup Q-M3 (Q1b1a1a) is defined by the presence of M3 Single Nucleotide Polymorphism (SNP) and Haplogroup Q-Z780 (Q1b1a2) by SNP Z780. The sub-haplogroups of Q-M3 and Q-Z780 have the highest frequency within the autochthonous lineages in Argentina. Currently the research of both sub-haplogroups is limited due to poor resolution of Y-chromosomal phylogenetic tree constructed from the classical SNP markers. Here, we analysed four amerindian whole Y chromosomes from Gran Chaco, Patagonia and Northwest Argentina generated using next-generation sequencing (NGS). We compare these with previously published Q-M3 and Q-Z780 sequences with the aim to discover new phylogenetically informative variants and increase the resolution of this branch of the Y Phylogenetic tree. Those new variants will improve our understanding of the geographic structuring of the genetic diversity of the Y chromosome in South America.

1785W

Genomic characterization of fluorescently tagged human induced pluripotent stem cells. *T.S. Grancharova, R.N. Gunawardane, Allen Institute for Cell Science.* Allen Institute for Cell Science, Seattle, WA.

The Allen Institute for Cell Science is using CRISPR/Cas9 to create a collection of fluorescently tagged clonal hiPSC lines targeting organelles and other major cellular structures in stem cells and cardiomyocytes. After extensive quality control, cell lines expressing endogenously tagged proteins are live imaged, and segmented images are used to create integrated models of cellular organization and behavior. As part of the quality control workflow, we used transcriptome and exome sequencing to assess the effects of gene editing, cloning, and extended culturing on the hiPSCs. We found that a set of ~50 genes enriched for developmental process gene ontology categories were down regulated in the edited lines relative to the unedited parental line (WTC). However, serially passaged parental cells (~15 passages after thaw) and mock edited cell lines (mock transfected, cloned, and subjected to same quality control as edited lines) displayed similar shifts in their transcriptional profiles suggesting that these changes are due to extended culturing and not to gene editing itself. A pilot single cell RNA-Seq experiment with the parental line and one edited cell line revealed more heterogeneity (multiple clusters with distinct profiles) in the parental population relative to the edited. The parental unedited line was whole genome sequenced, and exome sequencing was used to look for novel variants (absent from the parental line) in the edited clonal lines with a focus on mutations in genes involved in cancer susceptibility and cell proliferation. We found that most of the edited lines had a truncating mutation (NC_000017.11:g.60663160G>T; NP_003611.1:p.Glu476*) in the sixth exon of *PPM1D*, a known oncogene and regulator of *TP53* and *CHK2*. Although this mutation was not detected in the whole genome sequence of the early passage parental line, RNA-Seq data and a ddPCR SNP assay revealed that it was present at a low frequency in the parental population indicating that it arose sometime during the culturing of the parental line before it was used for editing. Preliminary ddPCR data also suggests that the allele frequency of the mutation increases in the parental population as it is serially passaged. The edited hiPSC lines are available to the non-profit scientific community, and their associated image and sequencing data will be openly available through the Allen Cell Explorer (allencell.org).

1786T

Improved sequencing depth metrics for variant calling error rate. *K.**Dunaway, S. White, W. Lee.* Helix, San Carlos, CA.

Sequencing depth (DP) is one major metric often used to determine validity of germline genotype calls. However, there are conflicting opinions about what the minimum DP should be. While a common held opinion is that 20x DP is the minimum acceptable DP for a confident call, many papers have cited differing opinions on the matter. Some of the confounding factors include type of experiment, sample preparation, and sequencing chemistry. This issue is further confused by conflating DP of a call with average coverage across the entire sample. Here we present the results of over 2000 replicates of control samples with known genomic sequence (NA12877 and NA12878). These were Exome+ sequenced in a CLIA/CAP controlled sequencing environment and analyzed using a proprietary pipeline that primarily follows GATK best practices. We have found that error rate does not decrease as DP increases above 20x. In fact, the DP threshold can move down even lower than 20x without significant increase in error rate. Moreover, adding additional thresholds to other metrics can yield more significant decreases in error rate. For instance, heterozygous calls with extremely low alternate or reference allele depth (AAD, RAD) will have higher error rate. We spell out these additional metrics to aid returning as much information as possible while still minimizing error rate.

1787F

Theoretical and practical considerations for validation of functional

assays in clinical variant interpretation applications. *S.E. Brnich^{1,2}, J.J. Se-kelsky¹, J.S. Berg¹.* 1) Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Curriculum in Genetics and Molecular Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC.

The increasing number of clinically identified genetic variants poses challenges to clinical variant interpretation, which considers patient phenotype, population data, and an appraisal of the literature for each variant in a disease-associated gene. Professional organizations such as the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) have developed standards and guidelines that weigh the strength of each piece of evidence for an overall classification along a spectrum from benign to pathogenic. The classification of "variant of uncertain significance" (VUS) applies in the case of conflicting or insufficient available evidence, as is common for rare, missense variants. VUS are numerous and cannot be used for diagnosis or clinical decision-making, thereby complicating patient counseling and management. The rarity of most individual VUS limits the collection of patient and family data, thus efforts to resolve VUS classifications have touted the potential of high-throughput functional approaches to pathogenicity assessment. We developed an algorithmic approach to examine evidence combinations within the ACMG/AMP classification framework and the effect of strong, functional evidence on overall variant interpretation. This method can be used to prioritize genes and variants for functional assay development based on the number of missense VUS that could be reclassified. Our initial analysis of two ClinGen variant curation expert panels indicates that 10/17 non-conflicting missense VUS would reclassify with strong functional evidence. Although there is great theoretical potential for functional assays to reclassify missense VUS, there are practical considerations related to how any given assay can qualify as "well validated" for application in the ACMG/AMP framework. First, it should consider the appropriateness to disease mechanism. The second consideration is validating with an adequate number of pathogenic and benign controls. Third is the ability to apply statistical rigor in estimating the sensitivity and specificity of the assay. Finally, the cost and throughput merit consideration, as they are rate-limiting steps in evaluating the growing number of VUS. In an ongoing process of establishing a homologous recombination assay for *PALB2* (MIM: 610355) missense variants, we developed real-world experience with each of these considerations, which will be presented here to help inform guideline development.

1788W

Long-read single-molecule maps of the hydroxymethylome. Y. Ebstein¹, T. Gabrieli¹, H. Sharim¹, G. Nifker¹, J. Jeffet¹, T. Shahal¹, R. Arielly¹, M. Levi-Sakin¹, N. Arbib², Y. Michaeli¹. 1) School of Chemistry, Tel Aviv University, Tel Aviv, Israel; 2) Meir Hospital, Kfar Saba, Israel & Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel.

We report on the development of a genome-wide analysis workflow for optical detection of fluorescent 5-hydroxymethylcytosine profiles along chromosomal DNA molecules. The epigenetic mark 5-hydroxymethylcytosine (5-hmC) is a distinct product of active DNA demethylation that is linked to gene regulation, development, and disease. In particular, 5-hmC levels dramatically decline in many cancers, potentially serving as an epigenetic biomarker. The noise associated with next-generation 5-hmC sequencing hinders reliable analysis of low 5-hmC containing tissues such as blood and malignant tumors. Additionally, genome-wide 5-hmC profiles generated by short-read sequencing are limited in providing long-range epigenetic information relevant to highly variable genomic regions, such as the 3.7 Mbp disease-related Human Leukocyte Antigen (HLA) region. We present a long-read, highly sensitive single-molecule mapping technology that generates hybrid genetic/epigenetic profiles of native chromosomal DNA. The genome-wide distribution of 5-hmC in human peripheral blood cells correlates well with 5-hmC DNA immunoprecipitation (hMeDIP) sequencing. However, the long single-molecule read-length of 100 kbp-1 Mbp produces 5-hmC profiles across variable genomic regions that failed to show up in the sequencing data. In addition, optical 5-hmC mapping shows a strong correlation between the 5-hmC density in gene bodies and the corresponding level of gene expression. The single-molecule concept provides information on the distribution and coexistence of 5-hmC signals at multiple genomic loci on the same genomic DNA molecule, revealing long-range correlations and cell-to-cell epigenetic variation.

1789T

A simple segue from Sanger to high-throughput SMRT Sequencing with an M13 barcoding system. C. Heiner, A. Larea, J. Gu, J. Harting, L. Aro, S. Ranade, J. Korlach. Pacific BioSciences, Menlo Park, CA.

High throughput NGS methods are increasingly utilized in the clinical genomics market. However, short-read sequencing data continues to remain challenged by mapping inaccuracies in low complexity regions or regions of high homology and may not provide adequate coverage within GC-rich regions of the genome. Thus, the use of Sanger sequencing remains popular in many clinical sequencing labs as the gold standard approach for orthogonal validation of variants and to interrogate missed regions poorly covered by second-generation sequencing. The use of Sanger sequencing can be less than ideal, as it can be costly for high volume assays and projects. Additionally, Sanger sequencing generates read lengths shorter than the region of interest, which limits its ability to phase allelic variants accurately. High throughput SMRT Sequencing overcomes the challenges of both the first and second generation sequencing methods. PacBio's long read capability allows sequencing of full-length amplicons with multiple passes per molecule for accurate genotyping with the phasing of allelic variants. In the present work, we have developed a universal M13 barcoding system for SMRT Sequencing that utilizes the M13 sequences commonly used to tail PCR amplicons for Sanger sequencing. We demonstrate both workflow efficiency and cost-efficiency of SMRT Sequencing for a broad range of clinically relevant targets >500 bp up to 5 kb. Our method provides a simple, cost-effective and highly accurate segue from Sanger to SMRT Sequencing for existing or new clinical sequencing assays to interrogate a spectrum of polymorphisms across a wider span of amplicon size, regardless of the genomic sequence context. .

1790F**SMRT Sequencing of clinically relevant targets using an amplification-free target enrichment approach.**

S. Ranade¹, P. Baybayan¹, S. Chakraborty¹, R. Hall¹, L. Aro¹, K. Varley², I. Vasenkova², T. Moore². 1) Pacific Biosciences, Menlo Park, CA, USA; 2) Kailos Genetics, Huntsville, AL, USA.

Simultaneous interrogation of genetic and epigenetic signatures in genomic DNA provide unique insights into clinically actionable variants of genes. However, isolation of genomic regions of interest without PCR amplification or complex bisulfite treatment for obtaining such information are still under-developed. In the present work, we have paired an amplification-free target enrichment method developed by Kailos Genetics with long-read SMRT® Sequencing technology. The circular consensus sequences (CCS) generated from contiguous long reads (average read-length >15-20 kb) for the control and methylated BRCA genes not only produced accurate genetic information but also detected all the expected epigenetic markers.

1791W**A new thermostable reverse transcriptase improves sensitivity, specificity and time to result for multiplex RT PCR-based detection and quantification of RNA.**

T.W. Schoenfeld, R.C. Heller, S. Chung, K. Dumas, K. Crissy, P.R. Barchard, D.M. Schuster. Research and Development, Quantabio, Beverly, MA.

Reverse transcriptase quantitative PCR (RT-qPCR) is often used to confirm, quantify and monitor transcripts, noncoding RNAs and viral targets that were initially identified by RNA-Seq projects. Despite their importance, the limited options among available reverse transcriptase enzymes can compromise performance in certain application. Most one step RT-PCR kits contain a mixture of two enzymes: a retroviral RT, usually an M-MLV derivative, and a thermostable DNA polymerase, usually derived from Taq polymerase. This reliance on two enzymes limits the temperature of cDNA synthesis and necessitates compromises in reaction conditions between those favorable for each respective component, which can reduce performance, especially on low-abundance and structured targets, and can limit storage stability. A molecular screening and directed evolution program has generated a new thermostable RT, active at up to 85°C and stable up to 95°C, that allows single-enzyme hot-start RT-PCR. Higher temperature cDNA synthesis and detection improves specificity and sensitivity and allows direct detection of RNA targets at near-single-copy sensitivity, notably on rare or highly structured targets. This new chemistry is compatible with singleplex and multiplex formats using most dye and probe-based detection chemistries. Detection facilitated by this enzyme is equally effective on RNA and DNA targets and can allow simultaneous detection of both targets without modification of the detection protocol. Because cDNA synthesis occurs rapidly during the initial stage of PCR, the elimination of a distinct RT step and inherently faster kinetics reduces detection times to less than 40 minutes. The inherent stability of the enzyme provides simplified formulation and improved liquid and dry storage. These improvements promise to simplify, accelerate and improve the reliability and flexibility of detection and analysis of mRNA, noncoding RNA and viral targets.

1792T

Harmonized sequencing based genetic testing in the eMERGE III

Network. D.M. Muzny¹, N. Lennon², H. Zouk³, E. Venner⁴, R.A. Gibbs¹, H.L. Rehm^{2,3,4}, *The eMERGE III Consortium.* 1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 2) Broad Institute of MIT and Harvard, Cambridge, MA 02141; 3) Laboratory for Molecular Medicine, Partners Healthcare, Cambridge, MA 02139; 4) Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA 02114.

Background: In eMERGE III, the network is for the first time conducting genetic testing via massively parallel sequencing. This necessitated coordination and harmonization of pre-analytical and analytical processes across two Genome Sequencing Centers (GSCs) and eleven clinical sites. **Methods:** Targeted capture panels of ~500 kb (including 56 'ACMG genes', 54 additional genes for discovery, and ~1400 individual single nucleotide variant sites (SNVs)) were designed on two different capture platforms (one per each GSC). These panels were applied for the analysis of 25,000 participant DNA samples. All steps, from collection of samples, panel clinical validation, sequencing, interpretation, reporting, and data distribution were harmonized. **Results:** We addressed issues at collection sites, in assay development, validation, primary analysis, variant classification, defining report content, data delivery, network access to interpreted variants and for progress reporting to build a complete and compliant network. For example, through data comparison and probe optimization, both sequencing centers achieved greater than 99.7% coverage of all bases targeted on the panels. In addition, through ongoing comparison of reportable variants between the two sequencing centers, we identified 51 variants with discordant interpretations. These were able to be resolved and enabled consistency in the interpretive process. Analysis of an interim data freeze representing 15,574 sequenced subjects, demonstrated the high quality of assay performance and showed that the rate of return of Likely Pathogenic and Pathogenic results was approximately 1.5%, but varied according to the collection criteria. **Conclusions:** We developed methods to coordinate genetic testing from multiple clinical sites and showed that two sequencing centers can harmonize both the technical and interpretive aspects of clinical sequencing tests, a critical achievement for the standardization of genomic testing.

1793F

Deciphering the methylome: Enzymatic Methyl-seq. L. Williams, C.V.K.

Ponnaluri, B. Sexton, L. Saleh, K. Marks, M. Samaranayake, L. Ettwiller, S. Guan, H.E. Church, N. Dai, E. Tamanaha, E. Yigit, B. Langhorst, Z. Sun, T.C. Evans, R. Vaisvila, E. Dimalanta, T.B. Davis. New England Biolabs, Ipswich, MA.

Cytosine (C) methylation is an important regulatory mechanism and identifying 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC) is essential in understanding gene regulation. Bisulfite sequencing (BS) is the gold standard for detecting methylated C's; however, DNA is commonly damaged and degraded by the chemical bisulfite reaction. Whole genome bisulfite sequencing (WGBS) demonstrates high GC-bias and enriched methylated regions. Here we describe an enzymatic approach to detect methylated C's: NEBNext Enzymatic Methyl-seq (EM-seq). EM-seq overcomes many of BS disadvantages as the DNA remains intact, thereby enabling longer sequencing reads. Additionally, libraries have reduced GC bias. EM-seq and WGBS libraries were prepared using 10 ng, 50 ng or 200 ng NA12878 genomic DNA. All libraries were sequenced using Illumina's NovaSeq 6000 and data were analyzed using a standard Bismark workflow. GC bias and insert size distribution were determined using Picard 2.17.2. MethylKit was used for correlation analysis. EM-seq libraries have longer inserts, lower duplication rates, a higher percentage of mapped reads and less GC bias compared to WGBS libraries. Global methylation levels are similar between EM-seq and WGBS libraries while CpG correlation plots demonstrate that EM-seq libraries are superior. EM-seq libraries identify ~2 million more CpG's than WGBS libraries and have a higher percentage of CpG's covered for all minimum coverage depths examined (up to 20X). Methylation profiles across genomic features (TSS, CpG island, etc.) were similar, however, normalized coverage was much higher for EM-seq compared to WGBS. Dinucleotide analysis showed EM-seq did not suffer from depletion of C containing dinucleotides nor enrichment of A/T containing dinucleotides as seen in WGBS. EM-seq is superior to WGBS: it is robust and works over a wide range of DNA inputs, has better sequencing metrics and detects more CpG's over a wide range of genomic features. .

1794W

Systematic comparison between different library preparation kits and sequencing machines for high-throughput whole genome bisulfite sequencing. L. Zhou¹, HK. Ng¹, B. Lehne², D. Moses³, S. Schuster⁴, CH. Kim⁴, JC. Chambers^{1,2,5,6}, M. Loh^{1,2,7,8}. 1) Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore; 2) Department of Epidemiology and Biostatistics, Imperial College London, London W2 1PG, UK; 3) Singapore Centre for Environmental Life Sciences Engineering, Nanyang Technological University, Singapore 637551; 4) MACROGEN, Inc., Seoul, Republic of Korea; 5) Department of Cardiology, Ealing Hospital, London North West Healthcare NHS Trust, Southall UB1 3HW, UK; 6) Imperial College Healthcare NHS Trust, London W12 0HS, UK; 7) Translational Laboratory in Genetic Medicine, Agency for Science, Technology and Research, Singapore (A*STAR), 8A Biomedical Grove, Immunos, Level 5, Singapore 138648; 8) Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117596.

Whole genome bisulfite sequencing (WGBS) is an emerging high-throughput technique for profiling DNA methylation at single nucleotide resolution across the entire genome in both fundamental and clinical research. Although there are extensive choices of WGBS library preparation kits and sequencing platforms available, their output and performance have not been systematically evaluated yet. We perform a comprehensive analysis on two leading library preparation kits (TruSeq and Swift Accel-NGS) that require low DNA input and two highest-throughput Illumina sequencing machines (X10 and NovaSeq) for their data quality and performance. We find that the average Q30 fraction and mapping rate in TruSeq sample is similar to Swift sample (95.7% vs 95.9%; $p=0.81$ and 79.4% vs 79.8%; $p=0.66$ respectively). However, the average median insert size of TruSeq sample is significantly shorter than that of Swift sample (142bp vs 276bp; $p<0.001$), with PCR duplication rate of TruSeq sample significantly higher than that of Swift sample (42.1% vs 23.4%; $p<0.001$). The percentage of methylated cytosine in CpG context of TruSeq sample is also significantly lower than that of Swift sample (76.8% vs 80.8%; $p<0.001$). Most importantly, the effective genome coverage (coverage normalized by raw reads) of TruSeq sample is significantly lower than that of Swift sample (9.6x vs 24.6x; $p<0.001$). When comparing the sequencing platforms, we find that although NovaSeq sample provides better sequencing quality than that of HiSeq X10 (Q30 fraction 96.9% vs 95.3%; $p<0.05$), neither the PCR duplication rate nor the effective genome coverage is significantly different from that of HiSeq X10 sample (35.5% vs 29.7%; $p=0.13$ and 18.2x vs 17.9x; $p=0.91$ respectively). In addition, we also compare our WGBS data with available array results and find that WGBS data show systematic bias. For beta values between 0 and 0.5, the WGBS sample methylation level is lower than the array beta value by 5.1% on average (SD: 1.5%), whereas for beta value between 0.5 and 1, the WGBS sample methylation level is higher than that from the array by 4.6% on average (SD: 2.3%). The bias is consistent across library preparation kits as well as sequencing platforms ($p>0.05$). In summary, our results demonstrate that Swift library kit outperforms the currently widely used TruSeq kit in various aspects and provide a potential solution for preparing WGBS library with better quality.

1795T

Comparison of whole exome capture products: Coverage & quality vs cost. B. Marosy, J. Gearhart, B. Craig, K.F. Doherty. Johns Hopkins School of Medicine, Baltimore, MD.

Next Generation Sequencing technology has been a driving force for reducing sequencing costs and enhancing genome technology; allowing investigators and clinicians the opportunity to better delineate an individuals' genetic makeup. For facilities providing genetic services, maintaining the balance of providing high quality data at the lowest cost is a continuous effort requiring recurrent evaluation and optimization of protocols. Here we have compared the performance between four different exome capture products: SureSelect Human All Exon v7 (Agilent Technologies), xGEN@Exome Research Panel v1.0 (Integrated DNA Technologies), SeqCap EZ Exome V3.0 (Roche) and Twist Human Core Exome (Twist Bioscience) to determine cost effectiveness and coverage improvements between products. Libraries were prepared from 50ng of HapMap DNA using the Kapa Hyper Prep kit (KAPABiosystems). Libraries were captured using either the Agilent, IDT, Roche or Twist exome capture product according to each manufactures protocol. Captured libraries were clustered and sequencing was performed on the Illumina HiSeq2500 platform. Sequencing data was down-sampled and the UCSC Exonic coding regions (35.2Mb) were used to ensure equal comparison between methods. Preliminary sequencing data analysis yielded 81.86%, 90.54%, 88.95% and 92.21% of targeted bases captured at $\geq 20\times$, respectively, when down-sampled to 4 Gb of raw sequencing data. Additional analysis of quality versus cost will be presented.

1796F

A re-designed and unified workflow for exome and custom targeted panels developed to support a high throughput clinical research environment. *E. Roth, J. Abreu, B. Blumenstiel, M. Cipicchio, M. DeFelice, M. Fleharty, J. Grimsby, A. Hollinger, S. Low, H. Lyon, A. Macbeth, G. Vicente, M. Costello, T. Desmet, N. Lennon, S. Gabriel.* Broad Institute, Cambridge, MA.

The Broad Institute Genomics Platform has developed a novel, unified workflow for whole exome and custom targeted gene panels. Our method utilizes Integrated DNA Technologies universal blocking reagents and capture method, resulting in exceptionally high specificity and even coverage, and compatibility with any DNA probe based panel design. By using this method in conjunction with panels synthesized by Twist Bioscience, we have demonstrated the ability to reach 97% of target bases at 20X coverage with only 5 gigabases of sequencing and a Fold 80 base penalty evenness calculation of 1.5 for a 12-plex exome pool, significantly decreasing the amount of sequencing lanes needed to reach product deliverables. In addition to developing a unified capture process, we have further optimized the workflow to be compatible with a wide range of sample input types and a variety of different DNA probe sets which can be selected on the same 96 well plate. This streamlining reduces turnaround time for custom panels and whole exomes in both research and clinical environments. Our Lab Information Management System (LIMS) provides tracking of individual libraries and allows for complex automated procedures and pooling techniques without loss of sample integrity. This novel workflow enables the optional use of unique molecular index (UMI) duplex adapters, allowing for highly efficient library construction that allows for consensus-based somatic variant calling for detection of low variant allele fraction events in deep coverage sequencing projects. For germline focused exome or custom projects, we have also developed a cost effective longer insert process to achieve more Gb per lane and minimize the amount of sequencing lanes needed to reach desired coverage. By combining highly efficient library construction techniques tailored to project needs with a novel capture method, we have achieved exceptionally even coverage for exome and custom panels with a minimal amount of sequencing.

1797W

Beyond the exome report: Approaches to additional analysis for undiagnosed genetic disease. *D.B. Zastrow^{1,2}, D.E. Bonner^{1,2}, L. Fernandez^{1,2}, J.N. Kohler^{1,2}, C. Reuter^{1,2}, M. Majcherska^{1,2}, C. McCormack^{1,2}, S. Marwaha^{1,2}, C. Curnin^{1,2}, C.M. Eng³, Y. Yang³, P. Ward³, D. Bick⁴, B. Korf⁴, E. Worthey⁴, E.A. Ashley^{1,2,5}, P.G. Fisher^{1,6}, J.A. Bernstein^{1,7,8}, M.T. Wheeler^{1,2}, U.D.N. Undiagnosed Diseases Network⁹.* 1) Stanford Center for Undiagnosed Diseases, Stanford University, Stanford, CA; 2) Division of Cardiovascular Medicine, Stanford University, Stanford, CA; 3) Baylor College of Medicine, Houston, TX; 4) HudsonAlpha Clinical Services Lab, Huntsville, AL; 5) Department of Genetics, Stanford School of Medicine, Stanford, CA; 6) Department of Neurology, Stanford School of Medicine, Stanford, CA; 7) Department of Pediatrics, Stanford School of Medicine, Stanford, CA; 8) Lucille Packard Children's Hospital Stanford, Palo Alto, CA; 9) NIH Undiagnosed Diseases Network, Common Fund, Office of the Director and the National Human Genome Research Institute, National Institutes of Health, Bethesda, MD.

Exome sequencing (ES) is often performed when a patient presents with a suspected genetic disorder that is genetically heterogeneous or when phenotype-directed genetic testing has not been diagnostic. Estimated yield of clinical ES is around 25% to 28% (Lee, 2014; Yang 2013, 2014; Retterer, 2016), still leaving a majority of patients undiagnosed. The Stanford Center for Undiagnosed Diseases has completed sequencing or ES reanalysis on 79 probands between Sep 2015 and May 2018: exome (n=51) or genome (n=21, 5 after ES reanalysis), ES reanalysis (n=12). Available family members were included to enhance sequencing analysis. The probands ranged in age from 4 months to 74 years and had disease suspicious for a genetic etiology in a broad range of systems (most common: neurologic, n=40; immune, n=11; and musculoskeletal, n=9). A genetic diagnosis was obtained from the initial ES report for 13/51 probands (25.5%), and further approaches were attempted to uncover a diagnosis for the remaining probands. Our ES reanalysis cohort had 3 diagnoses, and genome sequencing (after negative ES) identified 3 diagnoses. We present additional approaches for reaching a genetic diagnosis after an inconclusive clinical ES report, along with case examples. Additional analyses includes: multi-omics/RNA sequencing; clinical correlation with candidate variants (VUS, or autosomal recessive conditions with only 1 pathogenic variant identified); "research" variants identified by the clinical laboratory; data reanalysis; data sharing and collaboration (e.g. GeneMatcher); segregation analysis; additional data from patient/family (e.g. phenotype or family history); and assessment of mosaicism. These methods have resulted in 9 additional diagnoses in our cohort (9/79, 11.4%), and 10 probands with a candidate diagnosis where confirmation is ongoing (10/79, 12.6%), increasing our overall diagnosis rate from 16.5% after WES (13/79) to 44.3% (35/79). For patients receiving inconclusive exome sequencing results, these approaches may be utilized by the clinician to increase the chance of a genetic diagnosis.

Sequencing Diagnoses				
Sequencing	Probands	Diagnosis from initial report	Diagnosis after internal analysis	Candidate Diagnosis
Exome	51	13	3	5
ES Reanalysis	12	0	3	3
Genome	21*	3	3	2
Totals	79	16, 20.3%	9, 11.4%	10, 12.6%

*5 after ES reanalysis.

1798T

Nextera Flex enrichment with automation provides a rapid and panel agnostic NGS workflow. J. Zhao¹, S. Gross¹, R. Tim¹, S. Andrew², G. Schroth¹, S. Youting¹, E. Allen¹, Y. Martin¹, J. Rodriguez¹, S. Kevin¹, S. Deshmukh¹, A. Barr¹, M. Anthony¹, G. DeSantis¹. 1) Core Application Group, Illumina, Inc, San Diego, CA; 2) Biochemistry System, Illumina, Inc, The Watson Building 11 Granta Park Great Abington Cambridgeshire CB21 6GP United Kingdom.

The advance of Next Generation Sequencing (NGS) has greatly accelerated genomic studies, clinical diagnostics, microbiology and forensic science fields. Recent NGS technology advancements and cost reductions have democratized whole genome sequencing. However, in some cases, targeted sequencing assays are appropriate to achieve very high depth coverage of interesting regions. In such cases, highly multiplexed processes and simple workflows are necessary to reduce time and cost. Both amplicon and hybrid capture enrichment assays offer viable workflows. We have developed an innovative and simple fast hybridization reagent which, when paired with the advanced technology of Nextera Flex, can provide a sample to sequencer workflow to be completed in less than one day. Our assay was optimized to be panel agnostic and accept probe panels as small as 500 oligos or as large as 500,000 oligos, or greater. In addition, equivalent performance was demonstrated with both 120-bp and 80-bp capture oligos, achieving high padded read enrichment, high uniformity, low duplicate reads and sensitive minor allele frequency detection. The simple 6 hour workflow is compatible with direct blood or saliva input, and compatible with extracted FFPE. We have automated the workflow on common liquid handling systems to achieve hands-free preparations with minimal intervention. Data for several fixed and custom panels are presented that demonstrate padded read enrichment up to 90%, uniformity of coverage up to 97%.

1799F

Rapid and accurate quantification of Illumina NGS libraries using the Q real-time qPCR instrument. P. Bartholomew, H. Liu, E. Kolossovski, B. Komorous, D. Schuster. Quantabio, Beverly, MA.

Accurate quantification of the number of amplifiable library molecules is a critical factor for obtaining high quality read data with next-generation sequencing technologies. The high sensitivity, broad dynamic range, and specificity of qPCR to quantify library molecules that are suitable for the bridge PCR provide significant advantages over methods for total DNA quantification. However, these advantages are often offset by the time to result, requirement for inclusion of absolute DNA standards in every qPCR run and errors associated with dilution of libraries so that reportable results are within the linear dynamic range of the technology. Here we describe application of a new real-time quantitative PCR instrument, the Q from Quantabio, to simplify reliable library quantification with faster run times. The Q real-time PCR instrument exploits a novel magnetic induction technology to rapidly heat reactions held in a unique spinning aluminum rotor. This provides superior temperature uniformity of $\pm 0.05^\circ\text{C}$ that eliminates well position effects and cross-instrument variability associated with traditional peltier block-based real time cyclers. Robust optics with rapid data acquisition times are coupled to an intuitive data analysis interface driven by industry leading algorithms to provide highly accurate and reproducible qPCR results. To test the efficacy of this new qPCR instrument for quantification of NGS libraries, we measured libraries of various fragment sizes and GC contents over multiple trials using qPCR chemistries with the P5 and P7 primers, specific for Illumina adapter libraries under varying cycling conditions. The collective results clearly establish the suitability of the Q real-time PCR instrument for accurate quantification of NGS libraries. The clear benefits provided by the Q include: highly precise measurements across multiple trials; high efficiency amplifications under varied cycling conditions; exceptional quantitative sensitivity for distinguishing down to 1.2-fold differences; reliable results and performance from run times shorter than typical cycling protocols.

1800W

A quarter century of sequencing: Lessons learned and best practices for high quality genome sequencing at any scale. *M. Costello, Y. Farjoun, M. Fleharty, A. Hollinger, T. Desmet, J. Meldrim, S. Dodge, N. Lennon, S. Gabriel.* Broad Genomics, Broad Institute, Cambridge, MA.

In the mid 1990s, the Whitehead Institute for Genomic Research moved into an unassuming two-story building in Cambridge, MA that once housed a beer and popcorn distribution warehouse for nearby Fenway Park. This group would soon embark on an exciting new endeavour: the Human Genome Project. In November of 2003, the Whitehead Institute for Genome Research became the first cornerstone facility of the Broad Institute, and since that time we have sequenced hundreds of thousands of samples utilizing ever evolving technologies. Here, we provide our best practices and lessons learned from a quarter century of genome sequencing and analysis experience, with an emphasis on data quality, reduction in errors and variation, and sample integrity. Maintaining data quality at any scale can be taxing, particularly as samples from some biological sources and preservation methods may be inherently challenging. Here we highlight our experiences with both saliva and FFPE samples, two sample types with properties that can lead to decreased data quality. While saliva samples on average contain just 8-9% non-human DNA, this can range to greater than 50%, which we show in extreme cases could lead to low allele fraction false positives due to misalignment and describe a pre-sequencing assay we've developed to flag outlier samples. FFPE is notoriously hard to work with, and we describe our additional quality control steps to screen for and filter data containing FFPE-specific artifacts. Additionally, common genomics lab and analysis processes themselves can introduce errors and variation in downstream data, and we describe error profiles and suggest solutions for: (1) oxidation artifacts introduced during DNA shearing, (2) "index hopping" due to issues with Illumina patterned flow cell/exclusion amplification chemistry, and (3) false positives due to improper binning of quality scores for sequencer artifacts. Finally, knowing that the sequencing data we are generating is from the correct patient is extremely important, and we rely on a set of best practices to maintain sample chain of custody. This includes an orthogonal non-sequencing genotyping assay performed as samples enter the Broad to generate a "fingerprint" over carefully selected SNPs, as well as monitoring contamination in all samples using a metric sensitive to fractions of a percent of contamination.

1801T

CLIA whole genome sequencing and genotyping data at scale. *T. Howd, N. Fitzgerald, F. Kelley, M. DaSilva, T. Desmet, S. Ferreira, J. Meldrim, M. Harden, A. Macbeth, B. Woolf, K. Tibbetts, E. Banks, L. Gauthier, S. Friedman, L. Lichtenstein, W. Brodeur, M. Lebo, H. Rehm, N. Lennon, S. Gabriel.* The Broad Institute, Cambridge, MA.

As precision medicine continues to move from the realm of clinical utility studies towards standard of care, there is increasing demand for high-throughput and widely accessible molecular profiling delivered rapidly and securely. To meet this demand we have validated our germline whole genome sequencing (WGS) and genotyping processes for clinical use. We present here the major improvements made to our workflow to accomplish this goal, which include: Leveraging our existing research WGS and arrays workflows to produce CLIA validated data for diagnostic and clinical research use at scale; Optimizing laboratory processing and quality management to meet a deliverable of 95% of the genome at $\geq 20\times$ coverage or $>98\%$ call rates; Implementing and validating the Illumina Novaseq 6000, as well as upgrading our cloud analysis pipeline to maximize available capacity for both primary and secondary analysis of large WGS datasets. Broad has generated $>100,000$ WGS and $>1,000,000$ genotyping datasets in a research setting. By implementing and rigorously validating new technologies, streamlining user training and quality management, and integrating dynamic work design, we have created an environment that can be used to rapidly develop and implement new CLIA tests. We are continuing to learn from and improve upon the methodology used in the creation of our clinical WGS product to develop and launch a CLIA validated genotyping array workflow. Additionally, we are committed to providing the best analysis tools to serve the clinical community - we have integrated GATK SNV and InDel and calling with advanced single sample filtering to provide high quality variants that can be further annotated for pathogenicity. Our ability to generate high-quality, CLIA validated WGS and array data at scale serves the Broad's goals to improve human health by using genomics to advance our understanding of the biology and treatment of human disease, and to help lay the groundwork for a new generation of therapies.

1802F

Adapting long-read sequencing technologies for targeted and single cell applications. S. Iyer^{1,2}, S. Goodwin¹, W.R. McCombie¹. 1) Genome Center, Cold Spring Harbor Laboratory, Woodbury, NY; 2) Genetics Program, Stony Brook University, Stony Brook, NY.

Understanding the basis of complex diseases relies heavily on the ability to effectively study the genomic loci associated with these conditions. In the past, large-scale genomic studies have helped identify several disrupted genes and pathways associated with complex diseases, suggesting many risk loci. These risk variants - often copy number variants (CNVs), repeats, and structural variants (SV) - contribute to the highly complex regions of the genome that have not been successfully resolved yet. Long-read sequencing using the Oxford Nanopore Technology (ONT) allows for resolution of these complex genomic features by generating reads that are several (>10) kb long. Such reads have the potential to span difficult SV or repetitive regions with a single continuous read, giving a clearer idea of the variation and eliminating ambiguity in position or size. Currently, long-read sequencing methods are greatly limited by the DNA preparation step, with fragment lengths and yield often compromised by standard extraction methods. Moreover, the potential to generate reads at the megabase level relies greatly on library preps using large amounts (several µg) of DNA input. Our overarching goal is to be able to develop an approach that generates large contiguous reads from moderate to low input DNA masses (1-100 cell level). To this end, we use different extraction, amplification, fragmentation, and capture strategies to efficiently target and resolve complex regions of the genome that are often associated with disease conditions. We are currently employing and adapting CRISPR-based targeting techniques to enrich for specific regions of the genome directly from intact cells. Development of a targeted long-read sequencing strategy using low DNA input gives us the potential to study heterogeneous cell populations and work with rare samples. This approach would therefore give us the opportunity to characterize the landscape of genomic and transcriptomic variants in cancers, neuropsychiatric disorders, and other complex conditions across large populations of individuals suffering from these conditions.

1803W

Generating high-quality reference human genomes using PromethION nanopore sequencing. M. Jain¹, H. E. Olsen¹, K. H. Miga¹, J. Zook², R. E. Green¹, D. Haussler¹, M. Akeson¹, B. Paten¹. 1) Department of Biomolecular Engineering, University of California Santa Cruz, Santa Cruz, CA; 2) National Institute of Standards and Technology, Gaithersburg, MD.

The human reference genome assembly (currently GRCh38) is the bedrock of human genomics, however it represents the sequence of just one haplotype across much of its span. The modernization of the human reference to include a collection of diverse and highly accurate, haplotype-phased genome assemblies is now critical to future studies aimed at more comprehensively understanding the genetic and genomic basis of human health and disease. A comprehensive representation of common variants with higher than 1 percent allele frequency will require sequencing hundreds of such diploid genomes. This will require a consistent, maximally cost-effective and rapid protocol to dramatically increase the production of high accuracy, haplotype-phased assemblies. Current efforts for producing high-quality assemblies of a small number of individuals have mostly focused on expensive high-coverage, long-read sequencing and assembly protocols. While this work has been successful in delivering the necessary standards of quality, the overall cost and production time prohibit scaling to hundreds of individuals in a timely manner. However, new technologies and assembly methods allow for potentially substantial improvements in both cost and time without compromises in assembly quality. We will present our work on evaluating the PromethION sequencing platform from Oxford Nanopore Technologies. We are testing the platform's ability to produce long read, high-quality and high-coverage data in a fraction of time and cost necessary for contemporary long read sequencing methods. Our goal is to evaluate performance based on base accuracy, throughput/timing, and cost; goals that must be met to improve the scale of high-quality genome assembly. Recognizing that even 100kb reads are insufficient to scaffold through the most repetitive regions of the human genome, we will augment this sequencing with other technologies to facilitate scaffolding and haplotype phasing.

1804T

Sequencing by expansion: A novel technology enabling clinical DNA sequencing. *M. Kokoris, B. McRuer, M. Nabavi, J. Tabone, C. Machacek, M. Prindle, A. Jacobs, S. Vellucci, T. Lehmann, K. Collins, S. Keller, C. Cech, S. Zabezhinsky, R. Busam, J. Horsman.* Stratos Genomics Inc., Seattle, WA.

While the use of DNA data to inform clinical decisions is increasing at a rapid rate, there are still significant barriers to the widespread adoption of DNA sequencing. To address impediments to sequencing, Stratos Genomics has developed a novel, rapid and low-cost sequencing technology, utilizing a simple workflow that facilitates a wide variety of uses across both research and healthcare settings. Sequencing By Expansion (SBX) encodes DNA sequence into a highly-measurable surrogate polymer called an Xpandomer. To enable this technology, we developed expandable nucleoside triphosphates (X-NTPs) for template-dependent, polymerase-based Xpandomer synthesis. Much like PCR, Xpandomer synthesis uses the natural function of DNA replication, utilizing custom engineered polymerases to synthesize Xpandomers from targeted DNA templates. Once cleaved, the Xpandomer extends to 50X the original DNA length, encoding the sequence information in high signal-to-noise reporters that enable high-fidelity, single molecule sequencing in a low-cost nanopore instrument. Features unique to SBX solve many challenges presented by clinical DNA sequencing. Here we present SBX nanopore sequencing data showing accurate and specific detection of clinically relevant mutations. The simple Xpandomer workflow allows us to demonstrate sample to sequence in less than 2 hours. Further, SBX enables accurate homopolymer sequencing by utilizing a unique reporter as a molecular "clock" between each base, ensuring every base is called. SBX is specifically designed for versatility, which, in addition to meshing with current protocols, facilitates unique workflows such as direct sequencing of degraded, fragmented or rare DNA. SBX promises to expand the value of DNA data in the clinic. SBX's simple, flexible workflow and rapid time-to-data combined with the benefits of direct, single molecule nanopore sequencing are key to realizing the promise of clinical genomics, reducing the time required for personalized sequencing tests from weeks to hours.

1805F

Single-tube solutions to streamline DNA library preparation and improve sequencing results while lowering costs. *H. Liu, Y. Jin, M. Bolduc, D. Bays, E. Kolossovski, B. Komorous, D. Schuster.* QuantaBio, Beverly, MA.

Reliable, simple and rapid solutions for high quality next-generation sequencing (NGS) library preparation are in great demand as the technology has become widely adopted in research, clinical and many other fields. To offer streamlined and high-quality library preparation solutions for Illumina® sequencing platform, we have developed two high-quality library prep solutions, with and without enzymatic fragmentation, to streamline sequencing workflows on Illumina platforms for a variety of applications. The sparQ DNA Frag & Library Prep Kit supports easily tunable enzymatic fragmentation and end polishing in a single step, followed by high efficiency adapter ligation in the same tube without intervening purification steps. The streamlined workflow can be completed in under 2.5 hours with minimal hands-on time (30 min) accommodating DNA input amounts from 1 ng to 1 µg. The optimized chemistry leads to higher library yield (up to 2 folds), lower duplication rate, better sensitivity and better uniformity across broad range of GC content. This results in lower upfront library prep, bead purification, sequencing as well as processing costs. For sample types not requiring initial DNA fragmentation, the sparQ DNA Library Prep Kit provides a fast, single-tube library preparation solution ensuring great sensitivity (down to 250 pg) and efficiency, while maximizing library yields with minimal hands-on time. A high efficiency, low bias and high-fidelity PCR master mix was also developed to complete the workflow that requires library amplification. Our library solutions are ideal for whole genome sequencing, FFPE sample, and various target enrichment workflows. Leveraging our innovative chemistries, world-class enzyme purity and rigorously controlled production and ISO13485 quality system, both kits offer unmatched library solutions to address quality, speed, reproducibility and throughput while remaining a cost-effective option.

1806W**Nanobind magnetic disks for rapid high MW DNA extraction and library preparation in long-read sequencing and optical mapping applications.**

K.J. Liu¹, J.M. Burke¹, D. Kilburn¹, R. Fedak¹, V. Wallace¹, M. Kim¹, R. Workman², W. Timp², H. Ferrao³, K. Bjornson³, B. Galvin³, H. Sadowski⁴, C. Proskow⁴. 1) Circulomics Inc, Baltimore, MD; 2) Johns Hopkins University, Baltimore, MD; 3) Pacific Biosciences, Menlo Park, CA; 4) Bionano Genomics, San Diego, CA.

Long-read sequencing and optical mapping technologies from PacBio, Oxford Nanopore, and Bionano Genomics are rapidly becoming the de facto methods for de novo sequencing, phasing, and structural variant analysis. However, these capabilities are predicated on stringent requirements for high purity, high molecular weight (MW) DNA. The gold-standard method for high MW DNA extraction is plug lysis, a tedious two day method that can only process a handful of samples. Conventional spin columns and magnetic beads result in insufficient purity and fragment DNA resulting in compromised read lengths. We present methods for rapid high MW DNA extraction and library preparation based on our Nanobind magnetic disks. Each polymer disk is covered with a high density of micro- and nanostructured silica that protects DNA from damage and enables higher extraction yield and higher purity than conventional methods. High MW (300+ kb) and megabase (1+ Mb) DNA can be obtained from most sample types in less than 1 hour. High MW DNA extraction is readily automated using the Thermo Fisher KingFisher instrument. Two extraction chemistries have been developed. The Nanobind CBB Big DNA Kit has been validated for cultured cells, gram negative and gram positive bacteria, and whole blood. Sequencing on PacBio® Sequel® System and Oxford Nanopore GridION resulted in read length N50s of 26 – 49 kb with yields of 3 – 10 Gb depending on platform and sample type. Optical mapping of whole blood with Bionano Genomics Saphyr compared favorably to plug lysis with mapping metrics indicating comparable assembly performance and molecule length N50s greater than 250 kb. The Nanobind Plant Nuclei Big DNA Kit has been tested on samples including pine needles and maize. Sequencing on PacBio Sequel System and Oxford Nanopore GridION resulted in read length N50 up to 32 kb and yields up to 10 Gb. In addition to DNA extraction, Nanobind can be used for size selection purification with superior library yields and process time. Nanobind can achieve tunable size selection up to 10 kb with >80% recovery of high MW DNA and >99% removal of short DNA. Library prep can be completed in less than 4 hours with yields that are 2-5X higher than standard methods. Typical read length N50 of 22 kb – 30 kb and yields of 5 – 8 Gb are demonstrated on PacBio Sequel System and Oxford Nanopore GridION. In aggregate, these data demonstrate that Nanobind is a powerful platform for purification and preparation of high MW DNA.

1807T**Normalase – a novel library normalization tool for high-throughput NGS.**

D. Masser, V. Makarov, J. Lenhart. Swift Biosciences, Inc., Ann Arbor, MI.

Normalase is an enzymatic NGS library normalization procedure that produces libraries with a specified molar concentration. This eliminates the need for library quantification and concentration adjustment prior to library pooling, resulting in optimal clustering density and index balance. Normalase is bead-free and does not use a limited capacity immobilization step which saves time and improves performance. Instead, following PCR amplification, each library is incubated with Normalase reagents to produce a set of equimolar samples for simple pooling. Normalized libraries can then be directly sequenced without additional purification. Normalase allows for more than 10-fold variation in input quantity, while generating less than 10% variation in sample clustering within a pool. Sequence analysis of normalized libraries indicates that performance metrics such as coverage, complexity, base composition and coverage uniformity are not affected. At the same time, the Normalase procedure offers a significant reduction in time, number of steps, cost and the likelihood of errors in loading when preparing samples for high throughput sequencing without affecting biological data endpoints such as allele frequency. In addition, size-independent normalization by Normalase is beneficial for libraries produced by transposon-mediated or enzymatic fragmentation where a broad size distribution complicates molarity calculation. It can also be used for library normalization and pooling in hybridization-capture protocols.

1808F

A multi-omics approach to reappraising the protein-coding gene count in GENCODE. A. Frankish on behalf of the GENCODE Consortium. European Molecular Biology Laboratory, European Bioinformatics Institute, Cambridge, Cambridgeshire, United Kingdom.

Accurate and complete annotation of human genes supports both genome biology and clinical genomics, with missing and mis-annotated genes having the potential to significantly affect the validity of downstream analysis. While the question 'how many protein-coding genes are there in the human genome?' may seem like a legacy of the early 2000s, it remains unanswered – and indeed hotly debated – to this day. There are two reasons why current gene annotation catalogues may contain imprecise counts: firstly, as yet unannotated protein-coding genes could remain to be discovered; secondly, existing coding annotations may be incorrect. We produce the GENCODE reference gene annotation for human and mouse as part of the Ensembl project, and a key focus of our current effort is to provide more accurate annotation of coding sequences (CDS). Here, we will discuss the creation of a multi-faceted workflow that has allowed us to identify both prospective novel protein-coding loci and potentially dubious loci still retained in our current geneset, and its implementation in human. This workflow incorporates experimental data generated by both short and long transcriptomic methods, proteomics assays, evolutionary analyses based on PhyloCSF and comparative annotation, and a consideration of the spectrum of genetic variation in human populations. Application of this workflow has led to the addition of 144 protein-coding genes to the GENCODE human geneset, as well as additional coding annotations to over 300 existing loci. The majority of these annotations are truly novel and do not appear in any other geneset or protein sequence database. Conversely, survey of potentially dubious protein-coding genes has found nearly 200 loci in our annotation that when reviewed light of the experimental data and bioinformatic algorithms available today appear unlikely to produce genuine protein products. While these changes may appear modest in number, we shall demonstrate that they could have dramatic consequences on the results of user analyses. Most obviously, we are able to present new interpretations of dozens of variants linked to diseases or traits.

1809W

Quantitation and QC measurements for predicting downstream NGS success with FFPE and circulating cell-free DNA plasma samples. J. Mook, D. Wieczorek, S. Hermanson, C. Knox, D. Horejsh, E. Vincent, D. Storts, T. Schagat. Promega Corporation, 2800 Woods Hollow Rd, Madison, WI.

Introduction: Quantity and quality of DNA from formalin fixed, paraffin embedded (FFPE) tumor tissue samples is highly variable, with degradation and crosslinking due to the fixation process leading to issues with amplification and difficulty in downstream applications. An alternative to FFPE is circulating cell-free DNA (ccfDNA) from plasma or other biological fluids. Compared to genomic DNA, ccfDNA yields are typically low, with tumor cell present at significantly lower frequencies. Due to the inherent variability of FFPE and ccfDNA sample quality, DNA quantity is not in itself reliably predictive of downstream application success. Here we describe methods for predicting sequencing result quality utilizing the ProNex® DNA QC Assay. Methods: DNA was purified from matching tumor and normal FFPE tissue types as well as ccfDNA from plasma samples. DNA quantity was measured via single-target qPCR and used for NGS library construction with a 56 gene oncology panel. Discrepancies between quantity of DNA input into library preparation and expected library yield and sequencing coverage uniformity were noted. To investigate if downstream library yield and sequencing quality could be better predicted, samples were retroactively assayed with the ProNex® DNA QC Assay that includes three different amplicon sizes (75, 150, and 300bp). The quantitative differences between the increasingly larger amplicon sizes were calculated as ratios. FFPE samples with high 75/300bp ratios are indicative of highly degraded samples, and ccfDNA samples with low 150/300bp ratios are indicative of high molecular weight DNA contamination. Results: Retroactive testing with the ProNex® DNA QC Assay showed a strong correlation between high degradation ratios with low library yield and poor coverage uniformity. Samples with equal concentration and library input performed better, when degradation ratios were low. Conclusions: FFPE samples with high degradation ratios clearly correlated with poor coverage uniformity of sequencing, enabling users of the ProNex® DNA QC Assay to predict downstream performance. Use of a QC method allows researchers to qualify samples and make informed decisions about what downstream library method to use. Concentrating on less complex panels vs. highly multiplexed panels or whole exome sequencing for degraded samples can ensure getting the most useful information out of an individual sample, thus saving time, cost, and lost information on precious samples.

1810T

Automated FFPE purification using Maxwell HT DNA FFPE Isolation System. *L. S. Steffen, A. Zegers, E. Vincent.* Promega Corporation, Madison, WI.

FFPE tissues remain the most common sample type for solid tumor oncology research, including prospective and retrospective studies. While FFPE samples are ideal for storage and histological applications, the fixation process crosslinks and irreversibly damages nucleic acids, which provides challenges for achieving suitable performance in molecular techniques such as next-generation sequencing (NGS) and qPCR. Methods for purifying nucleic acids from FFPE samples have also historically been time-consuming, low-throughput manual methods using organic solvents. This study introduces a set of tools for genetic assessment of DNA from FFPE samples. The Maxwell HT FFPE DNA Isolation System enables high-throughput purification of DNA from FFPE tissues. This system can be implemented on a variety of automation platforms, is scalable, and reduces cost and hands-on time. When compared to other products, the Maxwell HT FFPE DNA Isolation System improves purification yields of amplifiable DNA as demonstrated using the ProNex DNA QC Assay for amplification-based quantification. Increased yields of amplifiable DNA improves the chance of success in downstream molecular assessment, including qPCR, ddPCR, and NGS applications.

1811F

No-amp targeted SMRT sequencing using a CRISPR-Cas9 enrichment method. *I. McLaughlin, Y. Tsai, B. Bowman, J. Ziegler.* Pacific Biosciences of California, Menlo Park, CA.

Targeted sequencing of genomic DNA requires an enrichment method to generate detectable amounts of sequencing products. Genomic regions with extreme composition bias and repetitive sequences can pose a significant enrichment challenge. Many genetic diseases caused by repeat element expansions are representative of these challenging enrichment targets. PCR amplification, used either alone or in combination with a hybridization capture method, is a common approach for target enrichment. While PCR amplification can be used successfully with genomic regions of moderate to high complexity, it is the low-complexity regions and regions containing repetitive elements sometimes of indeterminate lengths due to repeat expansions that can lead to poor or failed PCR enrichment. We have developed an enrichment method for targeted SMRT Sequencing on the PacBio Sequel System using the CRISPR-Cas9 system that requires no PCR amplification. Briefly, a preformed SMRTbell library containing the target region of interest is cleaved with Cas9 through direct interaction with a sequence-specific guide RNA. After ligation with new poly(A) hairpin adapters, the asymmetric SMRTbell templates are enriched by magnetic bead separation. This method, paired with SMRT Sequencing's long reads, high consensus accuracy, and uniform coverage, allows sequencing of genomic regions regardless of challenging sequence context that cannot be investigated with other technologies. The method is amenable to analyzing multiple samples and/or targets in a single reaction. In addition, this method also preserves epigenetic modifications allowing for the detection and characterization of DNA methylation which has been shown to be a key factor in the disease mechanism for some repeat expansion diseases. Here we present results of our latest No-Amp Targeted Sequencing procedure applied to the characterization of CAG triplet repeat expansions in the *HTT* gene responsible for Huntington's Disease. We demonstrate the ability to isolate and sequence hundreds of individual on-target molecules from a challenging enrichment target from multiple samples representing a range of repeat element expansion sizes in a single reaction.

1812W

Rapid library preparation using a bead-linked transposome for enrichment (eBLT). R.C. Tim¹, S. Gross¹, J. Zhao¹, A. Slatter², R. Kelley³, S. Snow¹, K. Shen¹, S. Deshmukh¹, A. Barr¹, A. Mai¹, M. You¹, J. Rodriguez¹, E. Vermaas¹, G. DeSantis¹, G. Schroth¹. 1) Core Applications Group, Illumina, Inc., San Diego, CA; 2) Advanced Platform Tech Dev, Illumina, Inc., San Diego, CA; 3) Bioinformatics, Illumina, Inc., San Diego, CA; 4) Product Development, Illumina, Inc., San Diego, CA.

Improving the efficiency and reducing the turnaround time for next-generation sequencing (NGS) library preparation is key for enabling faster sample-to-answer workflows. We have developed a novel and flexible self-normalizing, bead-linked transposome (eBLT) system for library preparation, that is compatible with targeted enrichment chemistry. Enzymatic tagmentation on-bead eliminates quantification of DNA, reduces sample extraction burden, creates self-normalizing libraries with very consistent fragment sizes and reduces both hands-on time and overall processing time. The fine-tuned and consistent fragment sizes achieved with eBLT reduces reliance on expensive and laborious methods for size selection. Using the eBLT technology, a simple single-day workflow for library preparation and targeted enrichment is possible. We describe the performance and application of a bead-linked transposome designed for enrichment applications (eBLT) for targeted NGS applications. The system is compatible with saliva, whole blood and FFPE treated samples. In most cases, library quantification and normalization can be eliminated. Library yields with < 20% CV were obtained using samples with an input range of as low as 10 ng and as high as 1000 ng. The eBLT system provides either short library sizes, well suited to enrichment applications or, long library sizes, suited to whole genome applications and enrichment applications that require phasing. We have also developed an array of unique dual (UD) indexes, which support both low-plexity and high-plexity pooling and substantially reduce index hopping missassignment. The eBLTs may also be applied to RNA for whole or targeted transcriptome applications.

1813T

Clinical metabolomics: A pivotal tool for the diagnosis and treatment of inherited metabolic disorders to enable precision medicine. N. Liu¹, R. Raina², L. Hubert¹, J. Alaimo¹, Y. Yang¹, J. Xiao¹, K. Pappen², V. Sutton¹, A. Kennedy², Q. Sun¹, S. Elsea¹. 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Metabolon Inc., Durham, NC; 3) Texas A&M University, Houston, TX.

The diagnosis of inborn errors of metabolism (IEMs) is challenging due to the complexity and variability of metabolic disorders, including clinical presentation and metabolic features. Clinical metabolomics, which is defined as the comprehensive analysis of small molecules in body fluids, is an emerging technology that holds promise of providing new insights into human diseases and serving as a primary diagnostic screening tool for IEMs. This method is capable of detection of >900 metabolites in a single test for the analysis of perturbations in biochemical pathways that would otherwise require multiple targeted tests. Consequently, the application of metabolomics leads to the discovery of novel biomarkers, monitoring of the treatment and improvement of interpretation of genetic variants. Metabolomic profiling of plasma was performed in 1302 individuals with diverse ethnicity and wide clinical spectrum with 90% having neurological phenotypes. Metabolites were analyzed by high resolution UPLC-MS/MS, and data were generated as Z-scores by comparing to a normal reference population. Of the total 1302 samples, metabolomic analysis contributed to identification and/or confirmation of the diagnosis in 91 cases by detecting the biochemical abnormalities that were consistent with the identified conditions, achieving a 7% diagnostic rate. Further, 21 cases were further confirmed by targeted biochemical tests, supporting the reliability of metabolomics. In total, we identified 33 known conditions including but not limited to aminoacidopathies, organic acidurias, fatty acid disorders, vitamin deficiencies, pentose pathway disorders, peroxisomal disorders, purine disorders, and neurotransmitter abnormalities, indicating that metabolomics can be utilized as an initial screening tool for diagnosis of known and novel or very rare IEMs. Using this method, we were also able to monitor therapies, detect disorders at the mild end of the phenotypic spectrum, identify novel biomarkers, and provide strong functional evidence in genetic variants interpretation. We here present the capability of clinical metabolomics in improving the diagnosis and treatment of IEMs, as well as the power of integrating clinical metabolomics with other molecular genetic data to provide a definitive diagnosis for metabolic disorders. The utilization of clinical metabolomic profiling facilitates the development of precision medicine approaches to personalized treatment, diagnosis, and prognosis.

1814F

New application of human-based methyl-capture sequencing for DNA methylome analysis in non-human primate. J.R. Lee¹, S.B. Yoon¹, Y. Son¹, D.H. Lee¹, S.I. Lee¹, K.J. Kim¹, J.W. Huh^{2,3}, J.S. Kim^{1,3}. 1) Primate Resource Center, Korea Research Institute of Bioscience and Biotechnology, Jeongseup 56216, Republic of Korea; 2) National Primate Research Center, Korea Research Institute of Bioscience and Biotechnology, Cheongju 28116, Republic of Korea; 3) Department of Functional Genomics, KRIBB School of Bioscience, Korea University of Science and Technology (UST), Daejeon 34113, Republic of Korea.

The non-human primate (NHP) have long been considered important animal models in biomedical research. The characterization of genomic or epigenomic variation in human disease models could advance the development of pathophysiological investigation, disease diagnosis, and research of clinical intervention. However, non-human primate-specific methods applicable to genomic or epigenomic analyses are still lacking. The recent development of methyl-capture sequencing (MC-seq) has a numerous advantage of cost-effectiveness, rapid sample processing time, broader genome coverage, and avoidance of the bias due to CpG-rich repeats. Here, we used a human probe-designed MC-seq method in cynomolgus monkey and African green monkey DNA samples to determine the applicability for these animal model. To effectively adapt the human probe-designed target region for methylome analysis in NHP, we redefined the MC-seq target region focusing on the sequence homology and variation in the promoter regions of the same genes between the NHP and human genomes. Methyl-capture efficiency was controlled by the sequence identity between the captured probes based on the human reference genome and the cynomolgus monkey and African green monkey genome sequences, respectively. Using reasonable guidelines, 65 and 56% of the human-based capture probes could be effectively mapped for DNA methylome profiling in the cynomolgus monkey and African green monkey genome, respectively, according to numeric global statistics. In particular, our method could cover up to 87 and 89% of the regulatory regions of the cynomolgus monkey and African green monkey genome, respectively. Therefore, adaptation of the established human-based MC-seq for NHPs can be a powerful tool for DNA methylation profiling, and help provide novel information about DNA methylation alteration patterns with direct clinical translation.

1815W

Combining barcode libraries with targeted gene expression for single-cell genetic analysis. A. Chenchik¹, M. Makhanov¹, P. Diehl¹, L. Kobzik², C. Frangou¹. 1) Cellecta, Mountain View, CA; 2) Harvard School of Public Health, Harvard Medical School, Boston, MA.

Labeling of target cells with lentiviral barcode libraries offers an effective approach for monitoring cell phenotype in time-course experiments *in vitro* and *in vivo* using single-cell molecular analysis. Cellecta's CellTracker libraries, with millions of different barcodes, enable rapid labeling of each cell in a large population with a unique sequence that is detectable in next-generation sequencing (NGS) RNA expression profiling assays such as RNA-seq. These barcodes then provide a key to identifying the sub-population of progeny cells derived from a single barcoded progenitor cell in a single-cell analysis data set. Furthermore, the progeny cells could be phenotypically characterized by expression profiles, such as stem cells, differentiated cells, activated or apoptotic cells, etc. Further, cell barcodes may be incorporated together with genetic effector libraries, such as CRISPR sgRNA libraries, to identify clonal phenotypic changes induced by specific genetic disruptions in progeny cells derived from the single progenitor cell. However, while barcoding provides an effective way to group cells based on clonal origin in heterogeneous cell populations, comprehensive single-cell expression profiling of large numbers of cells remains challenging. Generating genome-wide, quantitative RNA-seq data from thousands of cells is cost-prohibitive. With standard RNA-seq analysis of single cells, the depth of sequencing is limited and only highly expressed transcripts can be reliably measured. To address this limitation, quantitative single-cell gene expression profiling requires a targeted method that focuses on key genes involved in cell-specific biomarkers, signaling pathways or biological responses of interest. We will demonstrate how the combination of the DriverMap Targeted RNA Expression Profiling assay and CellTracker lentiviral barcode libraries provides a complete platform for highly sensitive single-cell expression profiling. Data will be presented showing how targeted DriverMap RNA expression profiling of single cells combined with cell barcoding could significantly improve phenotyping of distinct cell populations.

1816T

Demonstration of the successful extraction, library preparation, and sequencing of RNA from exosomes from plasma/serum samples frozen from the last decade. J. Garifallou¹, C. Vaccaro¹, L. Tian¹, A. McGarry², H. Hakonarson¹, R. Pellegrino¹. 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Cooper University Hospital at Rowan University.

Introduction: Human exosomes and extracellular vesicles contain nucleic acids such as mRNA, precursor miRNA, mature miRNA, noncoding RNA, and DNA that are released during cellular processes. Exosomes are involved in cell to cell signaling, immune response, and tumor progression. The RNA contained within these vesicles is protected by a lipid bilayer from degradation and lysis by RNases contained in the blood and therefore provides valuable, raw genetic information within the real-time context at the time of extraction. This genetic information has demonstrated value in gene expression analysis during disease development. Huntington's disease (HD) is an autosomal dominant disorder that is characterized by expansion CAG repeats within the HTT gene, with patients that have greater than 36 CAG repeats displaying the mutated gene. HD shows varying degrees of progression within patients (slow versus fast). In this study, we examined the exosomal RNA from plasma/serum taken at two different time points to assess the exosomal RNA for HD progression. **Methods:** RNA was extracted from plasma/serum using two separate kits, the Norgen Plasma/Serum Purification Mini Kit and the Norgen Plasma/Serum Exosome Purification and RNA Isolation Mini Kit. Following extraction, the RNA samples were quality assessed using the Nanodrop for RNA, the TapeStation RNA assay, the Qubit miRNA assay, and Qiagen miScript RT-qPCR assay. RNA quality was difficult to assess with these assays based on the small RNA samples that were extracted. However, the yield was sufficient to proceed with library preparation and sequencing. The samples were prepared using the Norgen Small RNA Library Prep Kit for Illumina. Sequencing was performed using the Illumina MiSeq instrument. **Results:** The sequencing produced 20 million reads for analysis. The sequencing did pass the standard QC metrics for the MiSeq, with cluster density of 1099K/mm² and the percent of reads greater than or equal to Q30 was 97.4%. **Conclusion:** This workflow was effective based on the successful sequencing of the libraries generated from long-term frozen plasma/serum samples. Sequencing data from exosomal RNA will continue to be valuable in assessing disease progression through real-time RNA transcriptome analysis as well as the treatment effectiveness in real-time. In addition to HD, exosome sequencing has potential utility in a wide-range of disease progression studies, including tumor development in cancers.

1817F

Sensitive and specific detection of low-frequency variants from degraded and cell-free DNA samples using a novel library preparation method. A.H. Potts¹, U.D. Chakravarty¹, Z. Zwirko¹, Y. Zheng¹, M. Light¹, K. Giorda¹, K. Lai¹, K. Bryan², S. Rose², Y. Bao¹, M. Jarosz¹, C. Chen¹. 1) Integrated DNA Technologies, Redwood City, CA; 2) Integrated DNA Technologies, Coralville, IA.

Next generation sequencing (NGS) has become a critical diagnostic tool that can be used to identify single nucleotide variants (SNV), indels, and copy number variations associated with diseases. However, clinically relevant samples, such as cell-free DNA (cfDNA) or FFPE DNA, present significant challenges for NGS analysis. DNA isolated from these samples is often of limited quantity and highly degraded, which makes both library preparation and bioinformatics analysis difficult. Existing library preparation methods require stringent size selection to eliminate adapter-dimers, which reduces library complexity and sensitivity. In addition, identification of low frequency mutations is hampered by the artifacts arising from poor sample quality, DNA extraction, library construction, PCR, and sequencing. We present a novel library preparation method that achieves a high level of conversion, eliminates adapter-dimers, and incorporates duplex unique molecular identifiers (UMIs) for *in silico* error correction. The method employs a 2-step ligation that does not require size-selection or adjustment of Ampure ratios during cleanups, which greatly reduces sample loss during library construction. In addition, maximal library conversion is achieved with a unique T4 DNA ligase and proprietary sequencing adapters that increase ligation efficiency and suppress chimera formation. The duplex UMIs in these adapters enables confirmation of the presence of mutations on both strands of input DNA. In this study, we tested the performance of our library preparation method using sheared genomic DNA, cfDNA, and FFPE DNA across a wide input range (1–250 ng). To mimic degraded DNA, we created libraries from mixtures of DNA isolated from NA12878 and NA24385 cell lines and sheared to an average length of 150 bp. We used 1–250 ng of input DNA with 0.25% mutant allele fractions (MAFs). To highlight its clinical applications, we also analyzed cfDNA (5–25ng) and FFPE DNA (25–100 ng) samples with 1% MAFs in each sample type. Libraries were enriched using a 75 kb xGen® Lockdown® Probes panel, followed by deep sequencing and SNV variant calling. Our results demonstrate that this new library preparation method yields a 50–100% increase in library complexity and substantially improves sensitivity for <1% variants compared to other commercially available methods. In addition, the inclusion of duplex UMI error correction virtually eliminates all false-positive calls, maintaining 100% specificity.

1818W

Xdrop: A new droplet-based technology for targeted long-read sequencing. A. Ameur¹, I. Höijer¹, E.B. Madsen², T. Kvist², M. Mikkelsen². 1) Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 2) Samplix, Herlev, Denmark.

Short-read next-generation sequencing (NGS) technologies are based on generation of millions of reads, typically 100-500 bp in length. As the sequence information is broken up in millions of short sequence fragments it is crucial that the reads subsequently are assembled correctly in order not to lose sequence information. This is a particularly challenging task in regions with repetitive sequence or with structural rearrangements. Long-read NGS technologies like Pacific Biosciences (PacBio) and Oxford Nanopore are capable of reading DNA molecules of tens to hundreds of kilobases in length, thereby practically solving the issue with assembling the sequence reads correctly. However, the long-read technologies come with an increased cost per sequencing unit as compared to short-reads. The cost can be reduced by enriching the DNA sample for the genomic regions of interest, and thereby increasing the number of reads covering those specific regions. Here we present the Xdrop technology, a novel microfluidic-based system that allows for targeted enrichment of long DNA fragments. The technology is based on isolation of long DNA fragments in millions of picoliter double emulsion droplets, which are fluorescently labelled and sorted on a cell sorting (FACS) instrument, resulting in an enriched population of long DNA molecules covering a target of interest. This approach is fundamentally different from other target enrichment methods. The enriched DNA can either be sequenced on a long-read sequencing platform, or alternatively provide long-fragment information from short-read sequencing. Only ~150 bases of sequence information is needed to perform enrichment of DNA fragments of up to 40kb in size, and the protocol requires as little as 1 ng of genomic DNA as input. Since all amplification reactions are performed in small droplets, the risk of forming chimeric molecules from a specific target is virtually eliminated. We have successfully used the Xdrop technology in combination with PacBio sequencing for a variety of applications, such as to determine integration sites of human papillomavirus 18 (HPV18) in the HeLa cell line, to obtain phasing information of co-existing mutations in TP53 gene, and to enrich for the Epstein Barr Virus (EBV) in a sample containing a 2200-fold excess of human genomic DNA. Our results demonstrate that the Xdrop technology is suitable for wide range of genomics experiments where long-read information is of importance.

1819T

Leveraging advances in uniformity and on-target rates to maximize coverage and reduce sequencing in first-pass designs of custom NGS target enrichment panels. R.I. Zeitoun, L. Arbiza, K.D. Butcher, S. Chen, E.H. Chilton, R.W. Gantt, C. Thompson. Twist Bioscience, San Francisco, CA.

Target enrichment has been an enabling approach to sequencing specific regions of the genome and calling variants while reducing sequencing cost per sample. The ability to create customizable panels has been a key driver for increased utilization of target enrichment within various workflows and applications. However, slow turn around times, challenges leading to context-specific biases, and lack of uniformity can strongly affect the quality and efficiency of results. These factors can be particularly evident in first-pass custom designs without optimization and can be especially evident for smaller panels and spike-ins where a lack of uniform enrichment and low on-target rates will quickly increase the cost per sample by requiring that samples are over-sequenced in order to cover all regions to a minimum sequencing depth. In order to create first-pass optimized custom panels, we focused on understanding the underlying variables that are critical for ensuring high on-target and uniform capture. Based on this understanding, we implemented innovative panel design rules and workflow chemistry ensuring high capture uniformity while providing fine control of off-target capture without compromising target coverage. To demonstrate the advantage of this design strategy and workflow, we designed two custom panels (40 kb and 800 kb) and an exome spike-in panel (120 kb) consisting of cancer targets.. This resulted in first-pass custom panel designs with >40% improved capture efficiency, ~2-5 fold improvements in off-target, and exceptional uniformity (fold 80 base penalty of 1.2), relative to naïve design and standard conditions. In addition, we show that the same design principle can be used to create very high quality spike-in panels to work in conjunction with our exome sequencing sequencing solution. In all, this study provides a clear demonstration of our approach to create high-performing custom target enrichment panels that effectively reduce per-sample costs with substantial gains in quality.

1820F

A fast and reproducible RNA-seq library prep with improved sensitivity and specificity across a broad range of RNA inputs. *Y. Jin, M. Bolduc, S. Li, H. Liu, E. Kolossovski, B. Komorous, D. Schuster.* QuantaBio, Beverly, MA.

RNA-Seq is the most popular method for whole transcriptome analysis and provides valuable information to quantify, profile, and uncover RNA. We have developed a rapid and streamlined directional RNA-Seq kit with increased sensitivity and specificity. The optimized chemistry and consolidated workflow greatly increases the library efficiency, reduces the hands time and overall cost to generate high quality RNA-Seq libraries. Together with the mRNA enrichment module, our RNA-Seq kit provides a fast (less than 5 hour), robust and highly reproducible solution with industry-leading sequencing performance for strand specificity, coverage uniformity, library complexity, and accuracy.

1821W

GRASS-C: Graph-based RNA-seq analysis in single cell level subgraph clustering. *H. Yang.* UCLA , Los Angeles, CA.

Single-cell RNA Sequencing (scRNA-Seq) has enabled quantification of gene expression in the resolution of a single cell. The characterizing of different cell populations in cellular heterogeneity can be potentially utilized for biological and clinical applications. In order to utilize the high resolution of scRNA-Seq, new methods must overcome the computational challenges arising from high dimensionality of the data and sparsity in gene expression from technical artifacts. Various statistical approaches have been applied to resolve these computational issues, but most methods are parameter-dependent and fail to fully characterize subpopulations. We present Graph-based RNA-Seq Analysis in Single-cell level Subgraph Clustering (GRASS-C), a clustering tool for identifying cell subpopulations in scRNA-Seq data. First, GRASS-C models the observed gene expression as a mixture of two Gaussian distributions, one representing the high expression signals, while the other indicating low expression and technical noises such as dropout. GRASS-C leverages a low-cut filter based on the Gaussian mixture estimate to keep important biological signals intact while minimizing the impact of noises in the data. We demonstrate that the low-cut filter can isolate biologically important signals from various biological and technical noises. Using cells as nodes and the similarity of gene expression between cells as edge weight, GRASS-C successfully identifies clusters with cells that share similar expression pattern with the Louvain algorithm for community identification. We show that GRASS-C successfully identifies different cell-types in six publicly available scRNA-Seq datasets comprised of different cell sizes, and experimental and biological backgrounds. Furthermore, GRASS-C is able to perform the clustering with higher accuracy than any other publicly available tools tested. Also, we demonstrate that the clustering information by GRASS-C can elucidate potential underlying biological insights through differential expression analysis on cell clusters.

1822T

Utilizing a PCR-based method to generate high confidence indel genotyping results as an orthogonal confirmation of NGS-identified variants.

K. Larkin¹, J. Wang², K. Beltz¹, M. Schnackel¹, K. Lai², W. Lee², S. Fiedler¹. 1) Research & Development, Integrated DNA Technologies, Coralville, IA; 2) Research & Development, Integrated DNA Technologies, Redwood City, CA.

Insertion and deletions (indels) are important sequence variants in human biology and disease, and are the second most common type of genomic variation only secondary to single nucleotide polymorphisms (SNPs). Reliable indel detection is therefore increasingly important in the era of personalized medicine. However, conventional methods such as next generation sequencing (NGS) have known limitations in detecting and calling indels. Because reads from large indel targets are often difficult to align to a reference genome, they may be removed or result in low-confidence genotype calls. Thus, an orthogonal method to confirm target genotypes is critical for obtaining high-confidence calls. Here, we compare the indel detection capabilities of two different methodologies. The first is an NGS-based hybridization capture (hybrid capture) method performed with xGen® Lockdown® Probes; the second is the PCR-based rhAmp® SNP Genotyping System. PCR-based genotyping methods offer a fast and cost-effective method to confirm NGS results, allowing for resolution of indel genotypes that is challenging with current NGS analysis methods. A combination of several elements makes the rhAmp SNP Genotyping System a particularly useful orthogonal strategy for high confidence indel detection and genotype calls: its allele-specific primer designs are highly successful with large indel targets; its dual-enzyme system provides high discrimination and sensitivity; and its universal reporter mix provides high signal-to-noise ratios. Our results show concordance in indel genotyping calls between the NGS hybrid capture method and the PCR-based rhAmp SNP Genotyping System across a panel of 46 human genomic DNA samples. We targeted indels ranging from 1 to 27 bases, allowing us to calculate concordance across a variety of indel sizes. Our results show that the PCR-based rhAmp SNP Genotyping System generates high-confidence calls for indels and successfully confirms NGS indel data.

1823F

Molecular barcode thresholding and unique dual sample indices reduce/eliminate index cross talk/hopping between NGS libraries to improve variant calling accuracy. *B. Mullinax¹, K. Chen¹, L. Niu², C. LeCocq², G. Amparo², C. Pabon², B. Arezi¹, H. Hogrefe¹.* 1) Agilent Technologies, La Jolla, CA; 2) Agilent Technologies, Santa Clara, CA.

The introduction of Illumina's ExAmp chemistry and patterned flow cell on the HiSeq 3000/4000/X Ten allow for the sequencing of multiplexed samples at high read depth and quality at a low cost per sample. While these advances provide advantages, they have also resulted in the misidentification of up to 5-10% of the sequencing reads in multiplexed samples. The misidentification is partly the result of "index hopping" believed to be due to the presence of unincorporated indexed primers in the pooled sample. Other potential sources of misidentification are cross-contamination of sample indices prior to use, index cross-contamination due to index hopping during post-capture amplification of multiple pooled libraries, and demultiplexing errors. The detection of variants at frequencies of 0.1-1.0% in heterogenous samples is often the goal of biomedical research and diagnostics. Index hopping or cross-contamination at a rate of up to 5-10% therefore has the potential to result in incorrect variant calls, especially false positives. As we will demonstrate, molecular barcode thresholding (i.e., removing all reads that have only a single molecular barcode) removes 90% of hopped reads thereby reducing the hopping rate to that of the HiSeq 2500 (0.3%) and significantly reducing the number of false positive calls on the HiSeq 4000. Thresholding leads to removing a significant number of sequencing reads, but with low DNA input and deep sequencing, panel sensitivity and specificity are comparable to the HiSeq 2500. Furthermore, the addition of a second sample index to be used in a unique combination with the first sample index, allows for the complete removal of all index hopped reads.

373W

Cell-type specific expression of circular RNAs in human pancreatic islets. S. Kaur¹, A.H. Mirza², F. Pociot^{1,3}. 1) Steno Diabetes Center Copenhagen, Gentofte 2820, Denmark; 2) Weill Cornell Medicine, Department of Pharmacology, New York, NY, USA; 3) Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen DK-1165, Denmark.

Background: Understanding distinct cell-type specific gene expression in human pancreatic islets is important for developing islet regeneration strategies and therapies to improve β -cell function in type 1 diabetes (T1D). While numerous transcriptome-wide studies on human islet cell populations have focused on protein-coding genes, their non-coding repertoire such as long non-coding RNA, including circular RNAs (circRNAs) remains mostly unexplored. Here, we explored transcriptional landscape of human α , β and exocrine cells and identified cell-type specific circRNAs. **Methods:** Total RNA-Seq dataset of FACS sorted α , β and exocrine cells from 3 human donors was retrieved from GEO (GSE50386). The raw fastq files were trimmed, cropped and adapters removed using Trimmomatic v.0.36. The resulting reads were aligned to human genome (GRCh38) using tophat-fusion and unaligned reads were further processed by CIRCexplorer2 (v.2.2). Intronic circRNA (ciRNA) candidates were discarded from the analysis and exonic circRNAs with ≥ 2 junctions spanning reads in ≥ 2 samples were considered expressed. **Results:** We identified 1526, 1007 and 456 circRNAs expressed in human α , β and exocrine cells, respectively. Only 219 circRNA candidates were commonly expressed in all three cell-types. The most highly and commonly expressed candidates were *MAN1A2* and *HIPK3* (≥ 100 junction spanning reads). Alternate circular isoforms were observed indicative of potential distinct functions in these cell-types. For example, *PTK2* had 12 and 10 circular isoforms in α and β cells, respectively, whereas only 2 isoforms in exocrine cells. We also observed cell-type specific expression of circRNAs. Functional annotation analysis revealed enrichment of β -cell-specific circRNAs in pathways related to regulation of apoptosis, DNA damage response and proteasome degradation. Non β -cell-specific circRNAs were mostly enriched for pathways related to diseases of signal transduction, SMAD2/3/4 transcriptional activity and insulin receptor signaling. **Conclusion:** Our study showed that circRNAs are highly abundant in both α and β -cells. We also showed that circHIPK3, previously implicated in β -cell dysfunction (Stoll et al. (2018)), is a non β -cell-specific circRNA and is globally expressed in all cell-types in human islets. We are currently investigating the potential roles of α , β -cell specific circRNA candidates in regulating β -cell function.

374T

Fast calling of copy number variants from population scale high-depth whole genome sequencing. A. Gilly¹, G. Png^{1,2}, D. Suveges^{1,3}, Y.C. Park^{1,2}, K. Walter¹, K. Kundu¹, I. Ntalla⁴, E. Tsafantakis⁵, M. Karaleftheri⁶, G. Dedousis⁷, E. Zeggini¹. 1) Human Genetics, Wellcome Sanger Institute, Hinxton, United Kingdom; 2) University of Cambridge, Cambridge, UK; 3) European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SH, UK; 4) William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK; 5) Anogia Medical Centre, Anogia, Greece; 6) Echinops Medical Centre, Echinops, Greece; 7) Department of Nutrition and Dietetics, School of Health Science and Education, Harokopio University of Athens, Greece.

Copy number variants (CNVs) are large deletions or duplications at least 50 to 200 base pairs long. They play an important role in multiple disorders, however accurate calling of CNVs is still a challenge today. Several large-scale high-depth whole-genome sequencing (WGS) efforts are underway, yet most current approaches to CNV detection use raw read alignments, which are computationally intensive to process. Here, we introduce UN-CNV, a regression tree-based CNV detection algorithm that produces population-wide callsets at unprecedented speed, using depth at variant calls from WGS. As a proof-of-principle, we applied UN-CNV to WGS ($>18\times$) from 6,898 samples across four European cohorts, and describe a rich, large variation landscape comprising 2,059 CNVs. 38.4% of detected events were previously reported in the Database of Genomic Variants, a dbSNP equivalent for CNVs. 18% of high-quality deletions excised entire genes, and we recapitulated known phenotype-altering events such as those affecting the *GSTM1* and *RHD* genes. We then tested for deletions associated with 275 protein levels in 1,457 individuals to assess the potential clinical impact of the detected CNVs. We describe the LD structure and copy number variation underlying an association between levels of the *CCL3* protein and a complex common structural variant affecting *CCL3L3*, a paralog of the *CCL3* gene (MAF=0.15, $p=3.6\times 10^{-12}$), expanding on previous evidence of copy-number variation at this locus. We also present a cis-association between a low-frequency *NOMO1* deletion and the protein product of this gene (MAF=0.02, $p=2.2\times 10^{-7}$) for which no cis- or trans- single nucleotide variant-driven pQTL has been documented to date. This work demonstrates that existing population-wide WGS callsets can be mined for CNVs with minimal computational overhead, delivering insight into a little-studied, yet potentially impactful class of genetic variant.

375F

Genome-scale sequencing to study the contribution of structural variation to human complex traits.

K. Walter¹, B. Howell¹, E. Gardner¹, J. Danesh^{1,2,3}, A. Butterworth^{2,3}, M.E. Hurles¹, N. Soranzo^{1,3,4,5}, The INTERVAL Study. 1) Human Genetics, Wellcome Sanger Institute, Hinxton, United Kingdom; 2) MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; 3) The National Institute for Health Research Blood and Transplant Unit (NIHR BTRU) in Donor Health and Genomics, University of Cambridge, Strangeways Research Laboratory, Worts Causeway, Cambridge, UK; 4) Department of Haematology, University of Cambridge, Cambridge Biomedical Campus, Long Road, Cambridge CB2 0PT, UK; 5) British Heart Foundation Centre of Excellence, Division of Cardiovascular Medicine, Addenbrookes Hospital, Hills Road, Cambridge, UK.

Structural variations (SV) are changes in the human genome usually larger than 50bp which include deletions, insertions, duplications and inversions. They are responsible for the majority of nucleotide variation among human genomes, and they are also implicated in various diseases and associated phenotypes such as cognitive disabilities to predispositions to obesity, cancer and other diseases. We are using whole genome sequence data (WGS) at 15X coverage from up to 12,000 samples in the INTERVAL study (<http://www.intervalstudy.org.uk>), which is a prospective cohort study of approximately 50,000 blood donors, who are 18 years and older, and who were consented and recruited from 25 National Health Service Blood and Transplant donor centers across England between 2012 and 2014. Participants are predominantly healthy individuals, they also completed online questionnaires containing basic lifestyle and health-related information, including self-reported height and weight, ethnicity, current smoking status, alcohol consumption, doctor-diagnosed anemia, use of medications (hormone replacement therapy, iron supplements) and menopausal status. The WGS data allow calling of SVs, including copy number variations (CNV) in coding as well as non-coding regions. After applying stringent quality control criteria and in-depth annotation, we evaluate the contribution of SVs and CNVs on blood cell phenotypes and a range of phenotypes related to inflammation and immunity. In particular, we assess the heritability of SVs that is not already explained by SNPs and INDELs, and we are investigating the burden of rare SVs on genes or exons, as well as their functional impact on the non-coding, regulatory part of the genome to study the genetic architecture of complex traits. The availability of WGS data at 15X coverage also allows detection of possible mosaicisms that could be linked to predisposition to disease.

376W

Defective gene splicing with mutant alleles of the ankyrin 1 gene isolated from patients with hereditary spherocytosis.

V. Alencar¹, A. Tondin¹, J. Klemm¹, J. Elliott¹, W.H. Cho², M. Carcao³, T. Lentz¹, A. Driver¹, G. Fischer², L. Drury², D. Caporale¹. 1) Biology, University of Wisconsin Stevens Point, Stevens Point, WI; 2) Prevention Genetics, 3800 South Business Park Avenue, Marshfield, WI; 3) Hospital for Sick Kids, Department of Hematology, 555 University Avenue, Toronto, Ontario.

Hereditary spherocytosis (HS) is a genetic disorder characterized by hemolytic anemia due to defective red blood cell morphology. The biconcave disc morphology of red blood cells enables them to traverse narrow capillaries. ANKYRIN1, encoded by the *ANK1* gene, tethers integral membrane proteins to the underlying spectrin-actin cytoskeleton. This interaction structurally supports the biconcave disc shape of the red blood cells. Loss-of-function variants in the *ANK1* gene lead to autosomal dominant forms of HS. We have identified a number of undocumented intronic variants in *ANK1* alleles from patients affected with HS. Computational analysis predicted these intron variants produce defects in gene splicing, resulting in formation of atypical mRNA and non-functional protein. To evaluate the effect of these intronic variants on splicing, gene fragments comprised of exon and *ANK1* intron sequences carrying each variant were cloned into the RHCglo minigene expression plasmid. Constructs carrying the wild-type allele sequences were also produced as controls. Plasmids were transfected into NIH/3T3 cells and total RNA was recovered at 72 hours. RT-PCR was performed to convert minigene transcripts to cDNA for sequencing. Constructs carrying wild-type allele sequences produced the minigene transcripts expected from wild-type splicing. Constructs carrying intronic variant sequences produced varied results, including altered and/or heterogeneous splice products. Many of the altered minigene splice products predicted by computational analysis were observed. These results were corroborated in replicate transfections. Extrapolation of observed minigene splicing to *ANK1* gene expression indicates many of these previously undocumented variants likely produce defective transcripts and altered ANKYRIN1 protein.

377T

TRBV polymorphism predicts adverse events during checkpoint blockade immunotherapy. *T. Looney¹, E. Linch², L. Miller², G. Lowman², F. Hyland¹.* 1) Thermo Fisher Scientific, South San Francisco, CA; 2) Thermo Fisher Scientific, Carlsbad CA.

Introduction Identifying predictive biomarkers for the emergence of immune mediated adverse events (IMAEs) during immunotherapy is a key objective of current immunotherapy research. Polymorphism within the TCRB variable gene (TRBV) has been implicated in autoimmune disease and may be mechanistically linked to IMAEs. Previous efforts to evaluate TRBV polymorphism have been hampered by the repetitive nature of the TCR locus and incomplete genome assembly. Here we employed a novel long-amplicon TCRB repertoire sequencing approach to evaluate the link between TRBV polymorphism and adverse events in 54 Caucasian research subjects receiving checkpoint blockade immunotherapy for cancer. We find that approximately one third of Caucasians possess a TRBV allele haplotype that appears protective against grade 3 or higher IMAEs. TRBV polymorphism may serve as a predictive biomarker of IMAEs to advance precision immunotherapy for cancer. **Methods** To measure TRBV polymorphism we used next generation sequencing of rearranged TCRB chains via RNA extracted from peripheral blood. We employed multiplex PCR to create amplicons spanning the three beta chain CDR regions to enable detection of polymorphism within the germline-encoded framework and CDR1 and 2 regions in addition to CDR3 profiling. TCRB receptor repertoires produced for each research subject were used to construct variable gene allele profiles to allow for inference of TRBV allele haplotypes. **Results** TRBV allele typing revealed the presence of distinct sets of co-inherited TRBV alleles, which we classified into major TRBV allele haplotype groups. We then evaluated the frequency of severe (grade 3 or higher) adverse events across haplotype groups. Strikingly, we find that the frequency of adverse events varies markedly between groups: one group, accounting for approximately one third of the sample set, appears protected from severe adverse events, while other groups showed moderate to elevated frequency of adverse events. **Conclusions** These data suggest that TRBV polymorphism may play a mechanistic or predictive role in autoimmune toxicity during checkpoint blockade immunotherapy. Current and future studies will further explore the utility of TRBV polymorphism as a predictive biomarker for IMAEs.

378F

Gene diversity of immunoglobulin variable domains among three ethnic groups in Nigeria: Implication for HIV vaccine development. *R.A. Ahmed¹, C. Scheepers², K.A. Adekoya¹.* 1) Medical Genetics Unit, Cell Biology and Genetics, University of Lagos, Yaba, Lagos State, Nigeria; 2) Centre for HIV and STIs, National Institute for Communicable Diseases, University of Witwatersrand, South Africa.

The quest of developing efficacious human immunodeficiency virus (HIV) vaccine is ongoing while crucial factors associated with past failure of vaccine efficacy trials are being critically considered. Identification of vaccine-compatible immunoglobulin or potent antibodies capable of eliciting vaccine-mediated protection via broadly neutralizing and effectors-activation responses is an essential factor of consideration. Several novel alleles of immunoglobulin gene are being reported recently from an ethnic group in Africa part of which contributed to functional antibodies against HIV infection. It is therefore hypothesized that significant novel and functional immunoglobulin gene variants are embedded in Africa population which are missing in the public databases. This may therefore have implications in the global development of a quality and efficient HIV vaccine. Nigeria, with the largest population in Africa and diverse ethnic group may harbor some of these novel and functional immunoglobulin gene variants. We started a study to sequence the entire rearranged antigen naive variable domain of heavy and light Immunoglobulin chain (VH, VK and VL) gene repertoire in 105 individuals from three major ethnic groups of Nigeria using Illumina Miseq. Data will be analysed using a new bioinformatics tool called IgDiscover that will enable the identification of novel and known germline gene segments. Novel alleles will further be assessed for functional HIV antigenic responses. Expected outcome of this study includes identification of large and diverse alleles from the heterogeneous study samples and novel alleles which will be vaccine compatible and capable of eliciting HIV broadly antigenic responses. This study will contribute to general knowledge of antibody especially towards utilization for a quality HIV vaccine development and other infectious diseases. It will also be of great importance to the Antibodyomics project which is aimed to provide holistic and cross-sectional information on human antibodies especially against HIV infection.

379W

Mosaic events in 181,022 Japanese individuals provide novel molecular insights into clonal hematopoiesis and hematopoietic malignancy. C. Terao^{1,2,3}, P. Loh⁴, Y. Momozawa⁵, M. Kubo⁵, Y. Kamatani¹. 1) Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 2) The Department of Applied Genetics, The School of Pharmaceutical Sciences, University of Shizuoka, Shizuoka, Japan; 3) Clinical Research Center, Shizuoka General Hospital, Shizuoka, Japan; 4) Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Boston, MA; 5) Laboratory for Genotyping Development, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan.

Clonal hematopoiesis produces DNA mosaicism in peripheral blood cells and can lead to hematopoietic malignancies. Previous studies have identified multiple inherited genetic variants associated with acquired *JAK2* mutation and myeloproliferative diseases. However, most mosaic events remain unexplained, and population differences in mosaicism have been largely unexplored. Here we investigated mosaicism in 181,022 individuals in BioBank Japan (BBJ) by developing a new method for processing signal intensities from DNA microarrays into clean B allele frequencies (BAF) and log R ratios (LRR) optimized for Illumina Infinium systems. We called mosaic events from this data using a recently-developed hidden Markov model-based method in combination with long-range phasing. We identified 59,491 mosaic events, an increase of 4-to-6 fold versus the number of events callable using 'raw' BAF and LRR and the number of events called in the largest previous study of mosaicism (151,202 UK Biobank participants). The genomic distribution of mosaic events in BioBank Japan was broadly similar to the distribution in UK Biobank, but we observed Japanese-specific focal deletions and co-occurrences of mosaic events. We also observed a marked reduction in the frequency of trisomy 12 events (a precursor of chronic lymphocytic leukemia) in BBJ, consistent with reduced CLL incidence in East Asians. We conducted genome-wide association studies to detect inherited variants which predispose carriers to mosaicism. We identified 10 cis associations ($p \leq 2.2 \times 10^{-10}$) including rare, highly-penetrant deleterious coding variants ($OR \geq 37$). We also found four trans associations ($p \leq 1.0 \times 10^{-11}$). The associations mapped to novel genes in the double strand break repair complex (*NBN* and *MRE11A*), a ploidy-related gene (*CTU2*), and hematopoietic malignancy-associated genes (*MPL*, *DLK1* and *NEDD-8*) in addition to *JAK2* and *TERT* which are known to associate with acquired *JAK2* mutation. The presence of mosaic events strongly associated with future development of hematopoietic malignancy, mortality from hematopoietic malignancy ($HR: 1.18$, $p = 1.3 \times 10^{-12}$), and overall mortality ($HR: 1.13$, $p = 3.7 \times 10^{-10}$) not due to hematopoietic malignancy. Furthermore, we found specific chromosomal mosaic events to be strong drivers of the associations (five events with $HR \geq 8.3$ and $p \leq 1.0 \times 10^{-9}$ and three events with $HR \geq 1.31$ and $p \leq 3.5 \times 10^{-4}$ for mortality from hematopoietic malignancy and overall mortality, respectively).

380T

USP8-wildtype pituitary corticotroph tumors harbor recurrent *TP53* somatic mutations and have a distinct aneuploid molecular subtype compared to *USP8*-mutated tumors. K. Ying¹, P. Javanmard², A. Roehnel², P. Taik¹, K. Cheesman², M.Y. Fink¹, A.S. Moe¹, E.E. Schadt^{1,4}, R. Sebra^{1,4}, K. Post⁵, R. Chen^{1,4}, E.B. Geer⁴, A.V. Uzilov^{1,4}. 1) Sema4 Genomics (a Mount Sinai venture), 333 Ludlow Street, Stamford CT, 06902; 2) Department of Medicine, Division of Endocrinology, Diabetes, and Bone Disease, Icahn School of Medicine, The Mount Sinai Hospital, 1 Gustave L Levy Place, Box 1055, New York, NY 10029, USA; 3) Division of Endocrinology, Department of Medicine, Multidisciplinary Pituitary and Skull Base Tumor Center, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, Box 419, New York, NY 10065, USA; 4) Department of Genetics and Genomic Sciences, Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, 1 Gustave L. Levy Place, New York, NY 10029, USA; 5) Department of Neurosurgery, Icahn School of Medicine at Mount Sinai, 1 Gustave L. Levy Place, New York, NY 10029, USA.

Pituitary corticotroph adenomas are rare tumors that can be associated with excess adrenocorticotrophic hormone (ACTH) and adrenal cortisol secretion, resulting in the endocrine condition Cushing's syndrome. Rarely, corticotroph tumors are 'silent' and do not result in elevated systemic ACTH and cortisol levels. Published whole-exome sequencing (WES) results are limited to only 23 such tumors across 3 studies, which revealed recurrent somatic mutations in ubiquitin specific peptidase 8 (*USP8*), the only driver gene known to date. Validation studies identified *USP8* driver mutations in 17-62% of ACTH tumors, with most studies reporting 31-36% prevalence; this leaves most corticotroph tumors without a genetic or molecular explanation of the driver mechanism. Moreover, while somatic copy number variant (sCNV) characterization of corticotroph tumors has been reported, it has not been integrated with molecular driver analysis. We carried out WES of corticotroph tumors and patient-matched normals from 18 patients that, compared to previous reports, are enriched for aggressive corticotroph tumor subtypes, including silent ACTH tumors, Crooke's cell adenomas, macroadenomas (>10 mm), and tumors with ki67 >3% or p53 positivity (14/18, 77.8%), including the first reported WES of an ACTH carcinoma. We performed integrative analysis of small somatic mutations (SNVs/indels) and large-scale genome rearrangements (sCNV and loss-of-heterozygosity, LOH). We found that *USP8*-wildtype tumors had a significantly higher proportion of the genome affected by sCNV and LOH, primarily via aneuploidy, than *USP8*-mutated patients, indicating different mutational mechanisms during tumorigenesis. We also found recurrent *TP53* mutations (in three *USP8*-wildtype tumors, 16.7%) at well-known hotspots. *USP8* and *TP53* somatic mutations had higher tumor allelic fractions (>20%) compared to somatic mutations in other candidate driver genes (<5%); comparison of allelic fractions with tumor purity estimates from WES suggests that these mutations are drivers occurring early in tumorigenesis and clonal expansion. We propose that corticotroph tumors may exhibit a *USP8*-mutated, genome-stable subtype versus a *USP8*-wildtype, genome-unstable subtype, the latter of which has a *TP53*-mutated subtype. These findings may inform our understanding of aggressive corticotroph tumor subtype pathogenesis, for which there are limited treatment options, with the aim to guide development of tumor-directed therapies.

381F

Changes in 3D genome architecture following chondrocyte differentiation as a model to understand skeletal disorders. I. Song, I.D. Bochkov, Y. Bae, E.L. Aiden, B.H. Lee. Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

Recent studies applying whole exome sequencing revealed that there are many aspects of skeletal dysplasia pathogenesis that cannot be explained by DNA coding regions. Accumulating evidences from genome-wide association studies and whole genome sequencing suggested that SNPs, copy number variations and structural variants in non-coding regions can play major roles in disease pathogenesis. Therefore, in this study, we aimed to identify the genome-wide 3D chromatin contact alterations associated with during chondrocyte differentiation and to explore their roles in skeletal disorders. We cultured ATDC5 cells using regular (undifferentiated) or chondrogenic (differentiated) medium for 21 days and performed loop-resolution *in situ* Hi-C on each cell population to obtain global 3D genome interaction profiles. We generated 8 replicate Hi-C libraries for each group and attained a total of 2 billion high-quality Hi-C contacts in each summed map. We do not observe dramatic changes following differentiation at the scale of topologically associated domains (TADs, median length = 175Kb in the control group, 170Kb in the differentiated group). However, we consistently observe a strengthening and reinforcement of loops throughout the genome in the differentiated cells. Using the Hi-C Computational Unbiased Peak Search (HiCCUPS) loop caller, we identified 10,020 and 13,145 loops in control and differentiated chondrocytes, respectively (median length = 300Kb and 305 Kb, respectively). To pinpoint cell stage-specific loops that may contribute to the regulation of the chondrogenesis, we performed a "diffHiCCUPS" differential loop analysis followed by manual curation of the identified DNA differential loops. We found 24 high-confidence control-specific and 146 high-confidence differentiation-specific loops. Currently, we are working on 1) elucidating the structural proteins that shape the chromatin during chondrocyte differentiation and 2) associating the differential loops with potential skeletal phenotypes, especially those known to have noncoding regulatory elements. This study is the first genome-wide profile of unbiased 3D chromatin contacts in the skeletal system. Additionally, this study will increase the understanding of how these interactions contribute to skeletal dysplasia pathogenesis as well as help to advance diagnosis and identify novel therapeutic strategies.

382W

The impact of mycobacterial antigen stimulation on isoform expression in whole blood. W. Correa-Macedo^{1,2}, M. Orlova^{1,3}, N. Van Thuc⁴, V. Hong Thai⁴, L. Abel^{5,6}, A. Alcaïs^{4,5,6,7}, L. Bruno Barreiro⁸, E. Schurr^{1,2,3}.

1) Program in Infectious Diseases and Immunity in Global Health, The Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada; 2) Department of Biochemistry, Faculty of Medicine, McGill University, Montreal, Quebec, Canada; 3) McGill International TB Centre, McGill University, Montreal, Quebec, Canada; 4) Hospital for Dermato-Venerology, Ho Chi Minh City, Vietnam; 5) Laboratory of Human Genetics of Infectious Diseases, Necker Branch, Institut National de la Santé et de la Recherche Médicale U1163, Paris, France; 6) University Paris Descartes, Imagine Institute, Paris, France; 7) St. Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, Rockefeller University, New York, New York, United States of America; 8) Department of Genetics, Centre Hospitalier Universitaire Sainte-Justine Research Center, Montreal, Quebec, Canada.

Background Leprosy type-1 reactions (T1Rs) are a major contributor to nerve damage and disability in leprosy patients. Early recognition of patients with T1R is a key element in efforts to prevent nerve damage in leprosy but useful biomarkers for advanced diagnosis are lacking. In the present study, we asked if the transcript isoform response of whole blood to *Mycobacterium leprae* antigens could be used as biomarker to identify patients that are at increased risk of developing T1R. **Methods** Whole blood aliquots at times of diagnosis from 41 leprosy patients were stimulated for 26 hrs with PBS or PBS + *M. leprae* sonicate. Of these patients 10 developed T1R over a 3-year follow-up. Stranded 75-bp paired-end reads were aligned, transcript quantification was performed with RSEM-v1.3.0, differential isoform expression test with limma and differential isoform usage analysis with DRIMSeq-v1.6.0. Gene ontology (GO) analysis, was carried out with PANTHER-13.1. **Results** For T1R-free subjects we identified 5,378 up and 1,286 down-regulated isoforms while for T1R-destined group 4,512 isoforms were up and 952 down-regulated. Groups had high overlap of differentially expressed (DE) isoforms, both at the gene and isoform levels and with respect to fold change direction, which yielded no DEI between groups. GO testing for up-regulated isoforms implicated GO terms representing processes such as phagocytosis, antigen processing, presenting and clearing of bacterial antigens. Due to the strong overlap in DE isoforms between both groups, no significant GO term was found specifically for either group. For DIU analysis we identified 15,565 isoforms being differentially used (DU). Intriguingly only a fraction of isoforms (T1R-destined: 1,509, T1R-free: 1,638) had differences in proportion changes above 10% when contrasted with baseline conditions. Genes with isoform usage difference above 10% also presented a strong overlap between tested groups, however 36 isoforms (20 genes) were significant after testing for interaction with 9 genes presenting a prevalent isoform switch. GO analysis failed to detect significant terms for any DU isoforms lists, however inspection of non-significant GO terms revealed that they were subsets of gene ontology identified for DE isoforms. **Conclusion** Our work failed to detect whole blood RNA isoforms that could serve as predictive biomarkers for increased T1R risk.

383T

TREM2 expression differs by cell type and genetic content. *R. Akhter¹, Y. Shao¹, E. Dyne², M. Khrestian¹, G. D'Aleo¹, S. Formica¹, J. Leverenz³, L. Bekris¹.* 1) Cleveland Clinic Lerner Research Institute, Cleveland, OH; 2) School of Biomedical Sciences, Kent State University, Kent, Ohio; 3) Cleveland Clinic Lou Ruvo Center for Brain Health, Cleveland Clinic, Cleveland, Ohio.

Genetic variation in triggering receptor expressed on myeloid cells 2 (TREM2) is associated with Alzheimer's disease (AD) and plays a role in neuroinflammation in AD mouse models. In the human brain, TREM2 mRNA is differentially expressed in multiple brain regions and human cell types. Little is known about how the expression of TREM2 is differentially regulated. In this study it was hypothesized that trans-acting factors present in certain cell types modulate TREM2 regulatory element activity. Therefore, TREM2 protein levels, cleaved soluble TREM2 (sTREM2), or TREM2 promoter activity will differ according to tissue or cell type. The objectives were to demonstrate that: 1) Cerebrospinal fluid (CSF) and plasma sTREM2 levels are elevated in AD. 2) Cell line, hippocampus and cerebellum TREM2 protein expression varies by tissue type. 3) Genetic variants in the TREM2 promoter and the 3' UTR influence expression in certain human cell lines. Human DNA, plasma and CSF samples were obtained from the Cleveland Clinic Center for Brain Health Aging and Neurodegeneration Biobank. Post-mortem brain samples were obtained from the University of Washington Alzheimer's Disease Research Center. CSF and plasma sTREM2 were evaluated using a custom capture sandwich immunoassay. Brain TREM2 was analyzed by Western blot. PCR amplification of genomic DNA was performed using TREM2 primers located outside of the promoter or 3'UTR region. TREM2 PCR products were cloned into a Luciferase reporter constructs. A transient transfection using these reporter constructs were performed in human cell lines. CSF sTREM2, but not plasma, levels were marginally higher in carriers of the AD GWAS rs7748513 risk allele suggesting that TREM2 genetic variation might influence gene expression depending on tissue type. Post-mortem brain and multiple human cell lines also showed differential expression by tissue or cell type. TREM2 promoter + 3'UTR reporter levels showed differential expression depending on genetic content of the promoter in human neuronal SHSY5Y cell lines. TREM2 promoter reporter inhibition in the presence of the TREM2 3'UTR was cell type specific. This inhibition was lost in the presence of genetic variants located in the TREM2 promoter region suggesting that SHSY5Y cells contain inhibitory factors that target this region. This finding is important because it will allow us to identify factors that regulate TREM2 expression and therefore potentially modulate neuroinflammation in AD.

384F

Copy number and structural variants in the Autism Speaks MSSNG Whole Genome Sequencing Resource. *S. Walker¹, R.K.C. Yuen¹, D. Merico^{1,2}, M. Bookman³, J. Howe⁴, B. Bhiruvahindrapuram¹, R.V. Patel¹, J. Whitney⁴, B. Trost¹, E. Higginbotham^{1,5}, N. Deflaux³, J. Bingham³, Z. Wang¹, G. Pellecchia¹, C.R. Marshall¹, N. Hoang¹, S.L. Pereira¹, T. Paton¹, W. Van Etten⁴, M. Szego⁶, B. Fernandez⁶, L. Zwaigenbaum⁷, E. Anagnostou⁸, P. Szatmari¹, W. Roberts⁹, D. Glazer⁹, T.W. Frazier⁹, S.W. Scherer^{1,5}.*

1) The Hospital for Sick Children, Toronto, ON, Canada; 2) Deep Genomics Inc, Toronto, ON, Canada; 3) Google, Mountain View, CA, USA; 4) BioTeam Inc, Middleton, MA, USA; 5) University of Toronto, Toronto, ON, Canada; 6) Memorial University of Newfoundland and Eastern Health, St. John's, NF, Canada; 7) Department of Pediatrics, University of Alberta, Edmonton, AB, Canada; 8) Holland Bloorview Kids Rehabilitation Hospital, University of Toronto, Toronto, ON, Canada; 9) Autism Speaks, Princeton, NJ, USA.

Due to high levels of heterogeneity, furthering the understanding of the genetic basis of Autism Spectrum Disorder (ASD) requires detailed characterisation of both genetic and phenotypic data from individuals across the clinical range of the disorder. The Autism Speaks MSSNG resource is a database of whole genome sequencing and phenotypic data for individuals from families affected with ASD. All data are available to ASD researchers via an open (but controlled) data access process and can be accessed in a cloud-based environment or through a web-based portal. Currently there are 7,231 samples available, all sequenced at high depth on Illumina platforms or by Complete Genomics. A further 1,710 in the process of being sequenced with a total of 10,000 sequenced samples anticipated by the end of 2018. From our earlier analysis, 61 genes were implicated as contributing to ASD risk based on *de novo* or X chromosome single nucleotide variants (SNVs) and small indels, with 18 of these genes identified for the first time from the MSSNG data. Subsequently, we have focused on the identification and characterisation of copy number variants (CNVs) and structural variants (SVs). We have detected CNVs >1 kb in size using our established pipeline combining ERDS and CNVnator, or from Complete Genomics data, and we are now adding SVs detected using algorithms Manta, LUMPY and DELLY. Of the 3,427 affected subjects currently available, 6.6% carry an ASD-risk rare CNV, falling into one of four categories: chromosome abnormalities (0.8%), large CNVs >3Mb (0.9%), CNVs corresponding to known genomic disorders (3%) and deletions at known ASD-susceptibility genes or loci (1.9%). A further 1.7% of subjects have rare duplications impacting ASD-risk genes, and we are combining CNV and SV data to resolve the structure of these duplications to facilitate their interpretation. Of 36 of these variants for which the structure has been analysed, 28 are tandem duplications and 8 are more complex in nature, including inversions, translocations and others with deletions or triplications embedded within the duplication. We are also analysing all SV data to include small CNVs (<1 kb), insertions and balanced rearrangements. Additional updates expected for the MSSNG resource include improved phenotype querying capabilities, re-analysis of data with hg38, inclusion of epigenetic data and availability of induced pluripotent stem cell lines from selected subjects.

385W

Haplotype resolved analysis and improved genome coverage with

Linked-Reads. S.R. Williams¹, P. Marks¹, A.M. Barrio¹, A. Fehr¹, I. Fiddes¹, J. Herschleb¹, N. Keivanfar¹, S. Maheshwari¹, A. Royall¹, P. Shah¹, N. Rahman², D. MacArthur^{3,4}, D.M. Church¹. 1) 10x Genomics, Pleasanton, CA; 2) The Institute of Cancer Research, Division of Genetics & Epidemiology, London, UK; 3) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 4) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA.

Traditional Short-read approaches struggle to access repetitive regions of the genome and fail to resolve the human genome into its component haplotypes. We describe an approach that identifies a more complete spectrum of genomic variation by retaining long range information while harnessing the power and cost-effectiveness of short-reads. From ~1ng of DNA, we produce barcoded short-read libraries in partition. Novel informatic approaches allow for the barcoded short-reads to be associated with their original long molecules producing a new datatype known as 'Linked-Reads'. Linked-Read whole genome sequencing (lrWGS) expands the amount of the genome that can be accessed and analyzed. We prepared 14 libraries for 3 individuals derived from the 1000 Genomes sample set and assessed high quality exonic mapping for each replicate. Using lrWGS enabled significantly more high-quality read coverage at coding regions ($\geq 5 \times$ coverage; MAPQ ≥ 30) than PCR-Free TruSeq. On average, lrWGS also provided an additional 38Mb of high-quality total bases containing medically relevant genes such as *SMN1*, *STRC*, and *NAIP*. The same trend of increased high-quality coverage is evident between Linked-Read Whole Exome (lrWES) preparations vs. traditional short-read whole exome preparations. lrWGS and lrWES provide the ability to reconstruct long haplotypes over the vast majority of the human genome which is essential to improved variant calling and interpretation. In five samples with known clinically-relevant compound heterozygous variants we were able to call all the known variants, phase them with respect to each other, and demonstrate that they are present in trans by 60x lrWES without the need for parental information. Being able to do long range haplotype reconstruction is also important for the identification of exon scale duplications or deletions. We show the advantage of this approach in notoriously difficult regions such as, *PMS2*, *BRCA1*, and *BRCA2* wherein all known small dups/dels that were phased were also called without the need for ultra-deep sequencing or known normal reference panels, when we could reconstruct haplotypes. The data presented here show that Linked-Reads provide a scalable approach or comprehensive genome analysis that is not possible using short-reads alone. We demonstrate the ability of both lrWGS and lrWES to accurately call small variation, to phase and call clinically-relevant variation, and to detect single exon deletions and single exon duplications.

386T

Analysis of rare coding variants in ALS *C9orf72* hexanucleotide repeat expansion carriers.

A. Chubick¹, A.P.S. Ori², T. Wu², R.H. Baloh^{3,4}, C. Lomen-Hoerth⁵, M. Wiedau⁶, N. Zaitlen⁷, L.M. Olde Loohuis², R.A. Ophoff^{1,2}. 1) Department of Human Genetics, University of California Los Angeles, Los Angeles, CA; 2) Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California Los Angeles, Los Angeles, CA; 3) Department of Biomedical Science, Cedars-Sinai Medical Center, Los Angeles, CA; 4) Department of Neurology, Cedar-Sinai Medical Center, Los Angeles, CA; 5) Department of Neurology, University of California San Francisco, San Francisco, CA; 6) Department of Neurology, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA; 7) Department of Medicine, University of California San Francisco, San Francisco, CA.

Amyotrophic lateral sclerosis (ALS) is a late-onset neurodegenerative disease that primarily affects motor neurons and often results in death from respiratory failure 2-5 years after diagnosis. The hexanucleotide repeat expansion (HRE) in chromosome 9 open reading frame 72 (*C9orf72*) is the most common genetic cause of ALS, present in roughly 6% of patients with no obvious genetically inherited component and explains one-third of all patients with a positive family history of the disease. Since an oligogenic basis of ALS has been reported, we hypothesized that this may also be the case for ALS patients with the *C9orf72* HRE. We therefore performed a whole exome sequencing (WES) burden analysis in $n=96$ of these patients and examined enrichment of rare damaging protein-coding variants in different gene sets: (i) known ALS genes ($n=30$), (ii) DNA helicase genes as precursors for sporadic development of the HRE ($n=30$), (iii) differentially expressed genes in motor neurons of ALS patients ($n=220$), (iv) ALS implicated genes identified by GWAS ($n=387$), and (v) loss-of-function (LoF) intolerant genes ($n=3230$). We used the non-Finnish European population of the Exome Aggregation Consortium (NFE ExAC; $N_{\text{subject}}=33,370$) as controls. To compare against NFE ExAC, variants in protein-coding regions were filtered by allele frequency ($AF < 0.01$ and $NFE\_ExAC\_AF < 0.001$), variant type (variants predicted to be damaging by Polyphen2) and normalized by the number of non-damaging protein-coding variants. We observed no evidence for enrichment or depletion of rare damaging protein-coding variants in any of the ALS-linked gene sets, providing no further evidence for an oligogenic inheritance of ALS among *C9orf72* HRE carriers. Significance was calculated using 1,000 permutations, which randomly selected genes for each gene set to generate the distribution of rare damaging protein-coding variants in cases compared to NFE ExAC. Interestingly, we did observe a significant depletion of rare damaging protein-coding variants in the LoF intolerant gene set in cases compared to NFE ExAC (permutation 1,000 times; $P = 0.001$). However, we observe a similar depletion in another WES data set of ALS patients without the *C9orf72* HRE ($n=60$), suggesting that this result may be due to a technical artifact. While the aim of our study was to learn more about the genetic architecture of ALS patients with the *C9orf72* HRE, it also demonstrates both the potential and challenge of utilizing ExAC as a control.

387F

Chemical chaperone screening for Pelizaeus-Merzbacher disease. *T. Kougai¹, S. Koizume², E. Jimbo¹, T. Yamagata¹, K. Inoue³, H. Osaka¹.* 1) Pediatrics, Jichi Medical University, Tochigi, Japan; 2) Kanagawa Prefectural Cancer Center, Yokohama, Japan; 3) National Center of Neurology and Psychiatry, Kodaira, Japan.

Background and Objective: Pelizaeus-Merzbacher disease (PMD) is caused by the mutation or overexpression of proteolipid protein 1 (PLP1). PLP1 is abundantly expressed in mature oligodendrocytes and located mainly in the cell membrane. When PLP1 is mutated due to an amino acid substitution, the protein is unable to fold properly and is subsequently degraded, resulting in a failure to localize. The objective of this study was to screen the drug for PMD from change of the amount of protein level and intracellular localization. **Methods:** Among the various substitutions of PLP1, we selected PLP1^{A242V} because of the marked difference in the amount and intracellular localization of this protein compared with the wild-type protein and the availability of naturally occurring mutant mice. We established a stable cell line expressing PLP1^{A242V} fused with turbo GFP (tGFP) in MO3.13 cells, which are derived from oligodendrocytes. As the first screening, we analyzed the whole-cell intensity of PLP1^{A242V} after adding chemical libraries using an In Cell Analyzer® (GE Healthcare UK, Ltd., England) and selected agents that showed a tGFP intensity more than 1.5 times that of the control. As the second screening, we analyzed the cell membrane intensity of PLP1^{A242V} by images taken using a confocal microscope and selected the agents that significantly increased the tGFP intensity at the cell membrane compared with the control. Finally, we analyzed the change in the endoplasmic reticulum (ER) stress response of XBP1 based on the luciferase activity and the change in the gene expression using a micro array analysis after adding candidate chemicals. **Results:** After the first screening, nine agents were selected from a triplicate analysis. After the second screening, two agents significantly increased the tGFP intensity at the cell membrane compared with the control. In addition, both reduced the ER stress response. We then showed that one chemical reversed the gene expression changes by PLP1^{A242V} in a micro array analysis. **Conclusion:** We successfully identified a chemical that increase the protein level and proper localization of PLP1. This chemical is a promising candidate, as it normalizes the gene expression pattern and decreases the ER stress response. .

388W

High-throughput repeat expansion sequencing on PromethION: An ABCA7 VNTR case study. *A. De Roeck^{1,2}, W. De Coster^{1,2}, T. De Pooter³, S. D'Hert³, J. Van Dongen^{1,2}, M. Strazisar³, C. Van Broeckhoven^{1,2}, K. Sleegers^{1,2}.* 1) Neurodegenerative Brain Diseases group, Center for Molecular Neurology, VIB - University of Antwerp, Antwerp, Belgium; 2) Institute Born-Bunge, University of Antwerp, Antwerp, Belgium; 3) Neuromics Support Facility, Center for Molecular Neurology, VIB - University of Antwerp, Antwerp, Belgium.

Tandem repeats (TRs) can cause disease through the number of repetitions, genetic variants in the TR units, and nucleotide modifications. For many TRs, however, these features are very difficult - if not impossible - to assess, and require low-throughput and labor intensive assays. One example is a VNTR in *ABCA7*. We recently discovered that expanded *ABCA7* VNTR alleles strongly increase risk of Alzheimer's disease. Currently, this GC-rich TR can only be studied through Southern blotting, which hampers further molecular characterization and clinical applications. Here, we investigated the potential of long-read whole genome sequencing to overcome these challenges, using the high-throughput Oxford Nanopore Technologies PromethION platform. We optimized wet-lab and bioinformatic protocols, and we developed an algorithm to study the size, sequence, and nucleotide modifications of TRs. Since the release of PromethION in early 2018, we were able to sequence eleven human genomes originating from dementia patients and healthy elderly controls. All individuals were Southern blotted for the *ABCA7* VNTR, with VNTR allele sizes ranging from 1kb to 10kb. We obtained an average output of 71 gigabases (Gb) per flowcell, and a maximum yield of 98 Gb (30 X genome coverage; at this time the highest reported yield in the field). *Albacore*, *Guppy*, and *Scrappie events* basecallers were capable of processing all VNTR alleles. The VNTR strand (either cytosine-rich or guanine-rich) affected the estimated TR length, but when corrected for this bias, estimates were comparable to Southern blotting. *Scrappie raw* produced the most accurate sequence and length estimates, but failed to detect expanded VNTR alleles. The available tools, however, could not be used to detect differences between TR units due to a high basecalling error rate, strand bias, and the need for non-trivial TR reference genomes. To circumvent these issues and enable the analysis of (modified) nucleotide composition, we developed a novel algorithm based on dynamic time warping to identify and classify repetitive patterns in raw PromethION sequencing data. Each *ABCA7* VNTR allele presented a unique combination of alternative TR units, which was confirmed by multiple sequencing reads per allele. In this study, we demonstrate the feasibility of long-read human genome sequencing on PromethION, and subsequent in depth analysis of the *ABCA7* VNTR. We hope that this work will lead to high-throughput whole genome analysis of TRs.

389T

Targeted enrichment and long-read sequencing to genotype medically-relevant genes that cannot be assayed using short-read technologies.

A. Wallace¹, K. Varley², S. Gibson³, J. Schiffman⁴, A. Quinlan¹. 1) Department of Human Genetics, School of Medicine, University of Utah, Salt Lake City, UT; 2) Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah, Salt Lake City, UT; 3) Department of Neurology, School of Medicine, University of Utah, Salt Lake City, UT; 4) Department of Pediatrics, Huntsman Cancer Institute, University of Utah, Salt Lake City, UT.

While repeat-expansion polymorphisms are known to underlie several developmental and neurological disorders, technological limitations have inhibited the ability to resolve accurate genotypes for these variants. A pertinent example is the (CCCCGG)_n repeat expansion in *C9orf72* that segregates with up to 40% of familial amyotrophic lateral sclerosis (ALS) cases. There is a clear relationship between this repeat expansion and neurodegeneration. However, widely-used short-read sequencing is ill-suited to characterize the *C9orf72* repeat expansion due to polymerase slippage over low complexity sequences and the inherently limited mapping/haplotype resolution of redundant short sequences. Consequently, researchers and physicians must rely on crude variant characterization methods (e.g. Southern blots), which limit direct genotype-phenotype associations and create challenges for the diagnosis of disease in patients with atypical clinical presentation. In contrast, ultra-long (e.g., 10-150kb) sequencing reads generated by Oxford Nanopore Technologies (ONT) enable direct measurement of loci containing complex structures without being subject to amplification or alignment bias. We are therefore developing a molecular and computational framework for targeted sequencing and quantification of pathogenic repeat expansions by combining cas9 and amplification-free sequence capture methods with Nanopore long-read sequencing technology. First, extracted high molecular weight DNA is digested with cas9 nickase and exonuclease resistant adapters are ligated to the resulting sticky ends. Non-target DNA is degraded and the remaining sample is sequenced using the ONT MinION sequencer. Then, base-called reads are aligned to the target locus and phased into haplotypes by the presence of flanking heterozygous variants. Haplotype-specific repeat counts are then estimated using Tandem Repeat Finder software. We will present our application of this molecular and computational framework to 30 ALS patients with *C9orf72* repeat expansions to demonstrate its generalizability as a new genomic medicine diagnostic that is faster, cheaper, and more accurate than currently available tests. This rapid, adaptable method can be used illuminate other structurally complex "blind spots" in the human genome, which include over 600 genes, >30 of which have known clinical relevance, and perhaps more intriguingly and generally, regions of the genome that remain unstudied as sources of disease heritability.

390F

The 15q11.2 BP1–BP2 microduplications are enriched in patients with anorexia nervosa.

X. Chang, H. Qu, Y. Liu, J. Glessner, P. Sleiman, H. Hakonarson. Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA.

Anorexia nervosa (AN) is a complex psychiatric disorder with significant morbidity and mortality. Multiple lines of evidence suggest a strong genetic component in susceptibility to AN. For example, family studies indicated that first-degree relatives of patients with AN have an increased risk of the disease. Twin studies have estimated the heritability of AN at 50-60%. Recent genome-wide association studies (GWAS) also identified a significant locus of AN at 12q13.3. Given the critical role of copy number variations (CNV) in psychiatric disorders, we conducted a genome-wide CNV study of 1033 AN cases and 7529 controls. In addition to the previously detected 1.5 Mbp deletion at 13q12, we uncovered that the 15q11.2 BP1–BP2 microduplications are enriched in individuals with AN (AN cases: 1.26% (13/1033); Controls: 0.32% (24/7529); Fisher's Exact Test *P*-value = 0.00023). There are five breakpoints BP1–BP5 located in the proximal long arm of chromosome 15, which mediate multiple pathogenic deletions and duplications via non-allelic homologous recombination, such as Prader-Willi syndrome and Angelman syndrome (BP1-BP3 or BP2-BP3). The BP1–BP2 region spans approximately 500 kb and contains four evolutionarily conserved genes (*NIPA1*, *NIPA2*, *CYFIP1*, and *TUBGCP5*). Both BP1-BP2 deletions and duplications are emerging as a recognized syndrome. Affected individuals have been reported with developmental, motor, and speech delays, and neurological and/or behavioral issues, specifically autism, autistic spectrum disorder and attention deficit hyperactivity disorder. Taken together, our results suggest that the 15q11.2 BP1–BP2 microduplications may confer risk to the development of AN.

391W

Characterization of obesity in individuals with 16p11.2 deletions using electronic health record derived phenotypes in the DiscovEHR cohort. A.E. Hare-Harris^{1,2}, S.M. Myers², K.P. Flannery¹, K.J. Haddock², A.L. Morehouse¹, E. Maxwell¹, L. Habegger³, J. Overton³, J. Reid³, D.H. Ledbetter², C.L. Martin². 1) Biological and Allied Health Sciences, Bloomsburg University, Bloomsburg, PA; 2) Autism & Developmental Medicine Institute, Geisinger Health System, Lewisburg, PA; 3) Regeneron Genetics Center, Tarrytown, NY.

Deletions of 16p11.2 BP4-5 (16pdel) are associated with neurodevelopmental and psychiatric disorders (NPD), including depression and anxiety. In addition to NPD, obesity that begins in adolescence and persists into adulthood has also been observed in these individuals. Although many studies have characterized the NPD phenotype associated with 16pdel, few have focused on the obesity phenotype, particularly in adult cohorts. In this study, we used a genotype-first approach to identify 35 individuals with a 16pdel (mean age = 48 years) from ~60,000 patient-participants in the Geisinger-Regeneron Genetics Center (RGC) DiscovEHR project, which includes exome and electronic health record (EHR) data. We used EHR-derived phenotypes to assess the prevalence of NPD and obesity in these individuals. NPD diagnoses were more common in the 16pdel group (57%) than in those without a pathogenic recurrent CNV (NoCNV) (48%). Mean body mass index (BMI) among the 16pdel group was 51.3 (range 27.3-85.8), and was also greater than the mean BMI of 35.8 in the NoCNV group (n=48,064; $p<.001$). We investigated whether this increased BMI could be influenced by increased use of weight gain-associated psychotropic medications (WGP). We assessed the relationship between BMI and WGP in the 16pdel and NoCNV groups. Overall, 66% of the 16pdel group and 63% of the NoCNV group had 1+ EHR-documented WGP. Despite the increased NPD prevalence, the 16pdel group did not have a higher number of WGP (5.9 vs. 6.0) or longer duration of WGP use (36.1 vs. 32.7 months) than the NoCNV group ($p=ns$). A greater number of WGP and duration of WGP use were predictive of higher BMI in the NoCNV group ($p<.001$) but not in the 16pdel group ($p=ns$). We next compared the 16pdel group to individuals with other recurrent pathogenic CNVs associated with NPD (CNV; n=315). The 16pdel group had higher mean BMI than the other CNV group (38.2; $p<.001$), but the prevalence of NPD diagnoses was the same in both groups (57%; $p=ns$). Overall, the 16pdel group had fewer WGP than the CNV group (5.9 vs. 15.8; $p<.001$). Unlike the 16pdel group, CNV individuals with at least one WGP (55%) had higher BMIs than those without a WGP (36.5 vs. 39.6; $p=0.015$). We observed similar results when considering non-psychotropic medications that may cause weight gain as a side effect. These findings suggest that the biological effects of the 16pdel have a greater influence on the increase in BMI observed in these individuals than the effect of WGP.

392T

Non-coding structural variation in whole genome sequencing data impacts attention-deficit hyperactivity disorder (ADHD) and reveals population differences between African American and Caucasian ADHD Children. Y. Liu¹, X. Chang¹, J. Glessner¹, L. Tian¹, Y. Guo¹, D. Li¹, H. Qiu¹, K. Nguyen¹, H. Hakonarson^{1,2,3}, P. Sleiman^{1,3}. 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Department of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA; 3) Division of Human Genetics, Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

Attention-deficit hyperactivity disorder (ADHD) is a common, highly heritable syndrome with a poorly understood molecular mechanism. Previous studies have suggested that structural variants play an important role in ADHD but these were mainly focused on coding regions and in populations of European ancestry. In this study, we generated deep (>30x) whole genome sequencing data on 205 ADHD cases, including 116 African Americans (AA) and 89 Americans of European ancestry (CEU) and 506 matched controls. Structural variants (SVs) were called from the WGS data in all subjects using MANTA. SVs were classified as coding or non-coding if they encompassed exonic or intronic/intergenic regions respectively. We tested for enrichment of both classes of SVs in ADHD vs controls in AAs and CEU and subsequently determined if the impacted genes were overrepresented for pathways or gene ontology terms. Coding SVs showed a marginal enrichment in ADHD cases compared to controls (adjusted p value = 0.012). Whereas, non-coding SVs were significantly enriched in ADHD, impacting genes involved in neuronal functions, including nervous system development ($p 1.1 \times 10^{-11}$) neuroactive ligand-receptor interaction ($p 1.9 \times 10^{-5}$), axon guidance ($p 1.4 \times 10^{-3}$) and the glutamatergic synapse ($p 0.047$). Comparing SVs between ancestries showed no significant differences in the distribution of coding SVs between AA and CEU. However, we observed significant differences in non-coding SVs impacting the same ADHD associated pathways. Taken together, these results suggest that non-coding SVs may play important roles in ADHD and that genomic structural differences exist between AA and CEU ADHD cases. We conclude that WGS is a powerful discovery tool for studying the molecular mechanisms of ADHD and associated drug targets.

393F

Modeling regional rates of structural mutation and selection in the human genome. P. Tandon, O. Shanta, D. Antaki, M. Kleiber, O. Hong, J. Sebat. University of California, San Diego, San Diego, CA.

Genome rearrangement is a driving force in human evolution and a contributor to complex traits and disease. Structural variants (SVs), which include deletions and duplications, exhibit significant clustering in the genome, reflecting regional differences in mutation rate and natural selection. However, factors that shape patterns of SV mutation are poorly understood. We developed a model to predict SV mutation rates based on DNA sequence features. We then applied our SV mutation model to population genetic data to identify functional elements that are under significant selection. The SV mutation model was developed using the phase-3 SV call set of the 1000 genomes project. We first determined the correlation of genetic features (including GC content and repetitive elements) with the frequency of SV breakpoints. A regression model was constructed to determine the major effects driving frequencies of multiple classes of SVs measured in windows of various scale (1KB-1MB bp). We then used the model to predict SV mutation probabilities throughout the genome. The model was validated via demonstration of similar mutation patterns on multiple cohorts including an independent sample of 49,604 deletions and 5,619 duplications from our genetic studies of autism (PMID:29674594). Our model has better statistical fit (deletion $r^2=0.15$, duplication $r^2=0.08$) compared to previous efforts (deletion $r^2=0.09$, duplication $r^2=0.02$) due to the inclusion of a greater variety of DNA features and modeling of effects at multiple scales. At scales of 1 KB and 100 KB, estimated regional mutation rates varied from the mean by up to 9-fold and 3-fold respectively. Patterns differ by SV length. Major drivers of smaller (<10KB) SVs include short interspersed nuclear elements (SINEs) and short tandem repeats (STRs). Larger SVs (>100KB) are driven by segmental duplications and long terminal repeats (LTRs). Mutation rates of small SVs are negatively correlated with evolutionary conservation, but trends can vary based on SV length, suggesting that genes and other functional elements are predisposed to large scale rearrangements. We demonstrate that promoters and enhancers that are ubiquitously expressed in tissues from the Epigenome Roadmap Project are under significant functional constraint. Our results provide a more detailed view of SV mutation rates and help to delineate categories of non-coding functional elements that are under strong selection.

394W

Detecting inversion polymorphisms at population scale with linked read sequencing. C.W. Whelan^{1,2,3}, R.E. Handsaker^{1,2,3}, S.A. Gagliano⁵, L.J. Scott⁶, O.V. Evgrafov⁶, C. Armoskus⁶, J.A. Knowles⁶, M. Cippichio⁴, S. Gabriel⁴, M. Boehnke², M.T. Pato⁶, C.N. Pato⁶, S.A. McCarroll^{1,2,3}, *The InPSYght Consortium*. 1) Stanley Center for Psychiatric Research, Broad Institute, Cambridge, MA; 2) Medical and Population Genetics Program, Broad Institute, Cambridge, MA; 3) Department of Genetics, Harvard Medical School, Boston, MA; 4) Genomics Platform, Broad Institute, Cambridge MA; 5) School of Public Health, University of Michigan, Ann Arbor, MI; 6) SUNY Downstate Medical Center College of Medicine, Brooklyn, NY.

Inversions of genomic segments have been one of the most difficult forms of genetic variation to detect accurately: alignment errors, non-inversion sequence variation, and inverted duplications often present the same signatures in short-read sequencing data as real inversion alleles, resulting in high false discovery rates. Long-read sequencing or optical mapping technologies better detect inversions, but their cost has made it difficult to survey variation beyond a few individuals. Linked-read sequencing technology – in which droplet-based barcoding of DNA sequences from the same high-molecular weight molecules allows reads from the same genomic segment to be analyzed together – has the potential to enable analysis at the scale of cohorts and populations. Here we present an analysis of segmental inversions in a schizophrenia case/control cohort of 545 individuals sequenced with the 10x Genomics Chromium linked-read technology. Our analysis cohort includes similar numbers of African-American and European-ancestry samples, allowing for the study of variant frequency in more than one ancestral background. After manually curating preliminary calls produced by the Long Ranger analysis software from 10x Genomics, we identified 40 high-confidence inversion variants (30-500 kilobases in length), of which 6 were flanked by large deletions or other inversions and the rest had simpler architectures. We observed a range of allele frequencies: from 11 singletons to 5 events visible in nearly every sample. The latter represent likely errors in the reference assembly. 14 of the variants disrupted annotated transcripts. 7 of the more common variants exhibited significant differences in allele frequency between African-American and European-ancestry subcohorts. We discuss the challenges of identifying inversions from linked-read data and provide comparisons of published computational methods including GROC-SVs (Spies et al., 2017) and NAIBR (Elyanow et al., 2018), as well as our own algorithms under development, on benchmark samples and several samples with technical replicates. Discovering structural variants with linked-read sequencing, and using the long-range phasing information the technology provides to impute their presence into larger cohorts for which only short-read or SNP array data are available, can enable the study of additional classes of segregating genomic variation at population scale.

395T

Prenatal and postnatal molecular characterization by whole genome sequencing (WGS) of *de novo* structural variations (SVs). P. Callier^{1,2}, N. Marle^{1,2}, AL. Mosca-Boidron^{1,2}, A. Vitobello^{2,3}, D. Callegarin^{1,2}, S. Verdez^{1,2}, J. Puechberty⁴, J. Thevenon^{2,3}, E. Tisserand⁵, AM. Guerrot⁶, P. Chambon⁶, G. Joly-Hélas⁶, A. Guichet⁷, B. Keren⁸, L. Tosca⁹, S. Brisset⁶, L. Van Maldergem¹⁰, P. Khau Van Kien¹¹, M. Doco Fenyzy¹², E. Pipiras¹³, C. Beneteau¹⁴, JF. Deleuze¹⁵, O. Robert¹⁶, A. Bolland¹⁶, M. Poulleau¹, T. Jouan², C. Poe², T. Christel^{2,3}, L. Faivre^{2,3}, Y. Duffourd². 1) Laboratoire de Génétique Chromosomique et Moléculaire, DIJON, DIJON, France; 2) Inserm UMR 1231 GAD, Université de Bourgogne Franche-Comté, Dijon, France; 3) Centre de Génétique et Centre de Référence Anomalies du Développement et Syndromes Malformatifs de l'Interrégion Est, FHU TRANSLAD, Centre Hospitalier Universitaire Dijon Bourgogne, 21079 Dijon, France; 4) Service de Génétique Médicale, CHU Montpellier, Montpellier, France; 5) Service de Génétique, CHU Rouen; 6) Laboratoire de Cytogénétique, CHU Rouen; 7) Service de Génétique, CHU Angers; 8) Département de Génétique, Hôpital Pitié-Salpêtrière, Paris; 9) Service d'Histologie Embryologie Cytogénétique, Clamart; 10) Centre de Génétique Humaine, Besançon; 11) Centre de Génétique Médicale et Cytogénétique, Nîmes; 12) Service de Génétique Médicale, Reims; 13) Laboratoire de Cytogénétique, AP-HP, Hôpital Robert Debré, Paris, France; 14) Service de Génétique Médicale, Nantes; 15) Centre National de Recherche en Génétique Humaine (CNRGH), Evry, France.

In the pre- and postnatal period, structural variations (SVs) such as supernumerary chromosomal markers (sSMCs), apparently balanced translocations and inversions *de novo* are evidenced by karyotype. The presence of these SVs raises, especially during pregnancy, a difficult issue of prognosis due to the risk of intellectual disability. The advent of new technologies such as Whole Genome Sequencing (WGS) makes it possible to consider the detection and characterization of SVs. The goal is to use the WGS to identify and characterize SVs, and secondarily to improve our pipeline. A WGS was performed in 19 patients with *de novo* SVs found in postnatal for 11 patients and in prenatal for 8 patients. Bioinformatic analysis of the raw data was performed with Lumpy algorithms for translocations and inversions and ControlFreeC for CNVs. The analyzes show that 7/8 translocations are identified, one of which interrupts the *GRIN2B* gene at the breakpoint compatible with the phenotype of the patient. Undetected translocation involves pericentromeric breakpoints. The 7 complex chromosomal rearrangements are identified with, sometimes, the discovery of more than 10 breakpoints, in favor of a chromothripsis type mechanism. Of the 4 sSMCs, 2 with euchromatin are detected by WGS and 2 with heterochromatin are not detected. The use of the WGS is therefore particularly effective in characterizing the breakpoints of the SVs (16/19) in both prenatal and postnatal period. The project allowed us to identify SVs and improve our WGS pipeline.

396F

A multi-genic model accounts for the pathogenicity of variably expressive CNVs. M. Jensen¹, S. Girirajan^{1,2}. 1) Bioinformatics and Genomics Program, Pennsylvania State University, University Park, PA; 2) Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA.

Classical studies of pathogenic copy-number variants (CNVs) have relied on atypical deletions and translocations for identifying specific causative genes in CNV syndromes, such as *RAI1* in Smith-Magenis syndrome and *NSD1* in Sotos syndrome. A lack of genomic events to delineate critical intervals for causative genes and variable expressivity of phenotypes in another class of CNVs, including duplications and deletions at chromosomes 1q21, 15q11.2, and 16p11.2, have made causative gene discovery challenging. Here, we analyzed genomic and phenotypic data from 37,625 individuals with neurodevelopmental disorders and functional data on 402 genes in four model organisms to identify causative genes within 18 disease-associated CNVs. We first analyzed microarray and clinical data of 28,499 affected individuals, and found that each CNV contributed to a range of 18 neurodevelopmental phenotypes at varying degrees of penetrance (average of 29%) and expressivity. We next searched for causative genes in these CNV regions by comparing exome sequencing data of 9,126 affected individuals in three separate cohorts to unaffected individuals. Although these cohorts were well-powered (>0.8) to detect an enrichment of mutations in individual genes, we found that none of the 113 genes in these CNV regions were enriched for rare pathogenic mutations among affected individuals (FDR>0.05), even though 26% (29/113) of the CNV genes were highly intolerant to variation (RVIS <20th percentile). We further examined the functional effects of each CNV gene by systematically analyzing 27 categories of phenotypes from mouse, zebrafish, *Xenopus* and *Drosophila* models curated from model organism databases, and found that 65% (73/113) of the CNV genes presented neurodevelopmental phenotypes in at least one model organism. In fact, CNV genes were enriched for developmental phenotypes as well as for genes under evolutionary constraint (RVIS<20th percentile or pLI >0.9) compared to the entire genome (p<0.05). Additionally, animal models of individual CNV genes did not fully recapitulate the phenotypes of models for the entire CNV. Overall, our data show that a single causative gene model is not sufficient to account for the incomplete penetrance and variable expressivity of these CNVs. We instead propose a multi-genic model for CNV pathogenicity, where multiple causative genes within CNVs interact in an additive or synergistic manner and are in turn modulated by genes outside of the CNV region.

397W

Novel evidence of high resolution and high field diffusion mri reveal a primate LGN-LGN connection. *J.R. Korenberg¹, L. Dai¹, D. Pritchett², A. Van-Hoek².* 1) Pediatrics, Medical Genetics, University of Utah, Salt Lake City, UT; 2) Neurology/Korenberg Research, University of Utah, Salt Lake City, UT.

Establishing a primate brain connectivity and linking it function and human diseases remain an elusive goal. Diffusion MRI, which quantifies the magnitude and anisotropy of water diffusion in brain tissues, offers unparalleled opportunity to link the macroconnectome to histological-based microconnectome at synaptic resolution, providing opportunity to confirm the known connection and propose novel connection. In this report, we show evidence using high resolution diffusion spectrum imaging (DSI) in primate brain to support the existence of a novel brain connection between the left and right lateral geniculate nucleus (LGN). We obtained high resolution diffusion MRI data on *ex vivo* brain from *Macaca fascicularis*: MRI 7T, resolution 0.5 mm isotropic, 515 diffusion volumes up to b-value (aka diffusion sensitivity) of 40,000 s/mm² with scan time ~100 hrs. Tractography results show direct connection between the left and right LGN. This finding is supported by previous functional and neuro-anatomic data. First, Gitlin and Lowenthal (1969), reviewing 150y of literature including and post Gudden, there is no evidence from either Gudden or others, for the existence of a commissure connecting the two medial geniculate (MGN) in humans or even rabbits. Rather, the "medial root of the optic tracts" likely is what has been seen connecting the optic chiasma-related to optic tract bulges that go to the ventral medial aspect of the LGN. Second, although a number of very well done papers referred to above, identify other tracts in the region, no reports find or confirm the predicted Commissure of Gudden. It is interesting that Dougherty's review (2018) set out the possible modulatory circuitry responsible for binocular modulation, in the LGN, something they refer to as "an apparent paradox", and note that this may be explained not only by their proposals but also "through mechanisms yet to be uncovered". We propose that simplest explanation for their data may be found in the multiple paths of evidence for the LGN-LGN connection supported by our new finding. In conclusion, by combining the previous data in literature and our new tractography results, we hypothesize that there is direct LGN-LGN connection, and this finding will provide a flush of new interest and funding for the field to identify the mechanisms, the molecules responsible, the embryo/fetogenesis and pruning, and the consequences for disease malfunction and correction.

398T

Genetic modulation of RNA splicing with a CRISPR-guided cytidine deaminase. *X. Chang.* Shanghai Institute for Biological Sciences, Shanghai, Shanghai, China.

RNA splicing is a critical mechanism to modify transcriptome, and its dysregulation is the underlying cause of many human diseases. It remains challenging, however, to genetically modulate a splicing event in its native context. Here, we demonstrate that a CRISPR-guided cytidine deaminase (*i.e.* Targeted-AID mediated mutagenesis (TAM)) can efficiently modulate various forms of mRNA splicing. By converting invariant guanines to adenines at either 5' or 3' splice sites (SS), TAM induces exon skipping, activation of alternative SS, switching between mutually exclusive exons or targeted intron retention. Conversely, TAM promotes downstream exon inclusion by mutating cytidines into thymines at the polypyrimidine tract. Applying this approach, we genetically restored the open reading frame and dystrophin function of a mutant *DMD* gene in patient-derived induced pluripotent stem cells (iPSCs). Thus, the CRISPR-guided cytidine deaminase provides a versatile genetic platform to modulate RNA splicing and intervene human diseases associated with aberrant splicing..

399F

PAX6 deep intronic single nucleotide variants as a cause of congenital aniridia. A.V. Marakhonov^{1,2}, A.Yu. Filatova¹, T.A. Vasilyeva¹, R.A. Zinchenko^{1,3}, M.Yu. Skoblov^{1,2}. 1) Research Center for Medical Genetics, Moscow, Russian Federation; 2) Moscow Institute of Physics and Technology (State University), Dolgoprudny, Russian Federation; 3) Pirogov Russian National Research Medical University, Moscow, Russian Federation.

Aniridia is a rare autosomal dominant panocular disorder caused by either mutations in the *PAX6* gene or chromosomal 11p13 rearrangements. Molecular analysis (Sanger sequencing and MLPA) of a large cohort of aniridia patients from Russia (91 unrelated families) conducted earlier revealed a significant proportion of *PAX6* mutations affecting splicing (14 mutations including 5 deep intronic sequence variants). Five intronic SNVs located outside constitutive splicing site dinucleotides are chosen for functional analysis. DNA samples for analysis were obtained from patients with congenital aniridia, in which 5 SNVs near and at some distance from the canonical splice sites (4–14 nucleotides) were identified. Intronic SNVs pathogenic status was established according to ACMG/AMP recommendations for interpretation of sequence variants. Human Splicing Finder and IntSplice on-line tools predict them to disrupt *PAX6* pre-mRNA splicing. To determine the effects of each sequence variant on *PAX6* pre-mRNA splicing we used an *in vitro* minigene assay. To create minigene plasmids, three different genomic regions were amplified for cloning into splicing reporter vectors depending on the location of the investigated splice sites. Then, the constructs were transfected into HEK293T and A375 human cell line, and splicing products were identified with RT-PCR using plasmid-specific primers. *In vitro* analysis showed all five intronic variants located deeper than splicing site constitutive dinucleotides lead to alterations of splicing pattern: either exon skipping or exon elongation. These SNVs should result in open reading frame shifting, premature termination codon formation and subsequent aberrant mRNA degradation by nonsense-mediated decay. They produce a null allele and haploinsufficiency of the *PAX6* function. So investigated variants of unknown clinical significance were reclassified as loss-of-function mutations. Our results emphasized the necessity of functional analysis and advanced search for *PAX6* mutations. This work is supported by RFBR grant 17-04-00475.

400W

Microcephaly associated to 5pter deletion in patients with Cri du chat syndrome. S.N. Chehimi¹, E.A. Zanardo¹, J.R.M. Ceroni², F.A.R. Madia¹, A.T. Dias¹, L.L. Vieira¹, G.M. Novo-Filho¹, M.M. Montenegro¹, A.M. Nascimento¹, T.V.M.M. Costa¹, Y.G. Oliveira¹, A.F. Brasil¹, J.G. Damasceno¹, C.A. Kim², L.D. Kulikowski¹. 1) Departamento de Patologia, Faculdade de Medicina FMUSP, Sao Paulo, Brazil; 2) Unidade de Genetica, Departamento de Pediatria, Instituto da Crianca, Hospital das Clinicas HCFMUSP, Faculdade de Medicina, Universidade de Sao Paulo, Sao Paulo, SP, BR.

Deletions in the short arm of the chromosome 5 (5p-) result in a genetic disorder associated to a wide spectrum of phenotypic features that include high-pitched cry, developmental delay, hypertelorism, broad nasal bridge and microcephaly. The delineation of the chromosomal breakpoints that involve the 5p deleted region enable a more detailed recognition of the missed genomic content, allowing the selection of critical regions associated with the phenotype. Using Illumina Infinium CytoSNP-850K we evaluated 14 patients with previously detected deletions in 5p in order to investigate a region potentially associated with microcephaly. This experiment provided us copy number variation (CNV) regions in all the chromosomes, including chromosome 5. We were able to detect 5p terminal deletions in 13 patients, with the initial breakpoint in 25,328 basepairs (bp), and one patient with interstitial deletion starting in 4,788,892 bp. The patient with the interstitial deletion had p=10 for head circumference while the others had p<3. CNVs in other chromosomes were analyzed and these were found to have no phenotypic effects concerning to microcephaly. All the patients presented other phenotypic features unrelated to the presence of microcephaly. Since microcephaly is characterized by a head circumference below the third percentile, our results indicate that the region associated with the presence of microcephaly is narrowed to a 4,7 Mb region between 25,328 bp and 4,788,892 bp. That observation explains the absence of this specific features in the patient with interstitial deletion. Multiple critical regions have been reported for microcephaly in patients with 5p- syndrome but it was always implied with other features making it difficult to delineate a genomic region corresponding to this particular phenotype. The findings in this study highlights the importance of using cytogenomic techniques to detect regions that contribute to the phenotype and therefore provide personalized clinical care for these patients.

401T

Investigation of the altered gene expression in cases with trisomy 18. *I. Albizua, M.J. Gambello, S.L. Sherman, S.T. Warren.* Human Genetics, Emory University School of Medicine, Atlanta, GA.

Trisomy 18, also called Edward syndrome, occurs in 1/6000 live births and is associated with significant morbidity and mortality. The extra copy of chromosome 18 causes the altered expression of many genes and leads to substantial brain, heart and kidney malformations as well as other medical problems. Due to the low rate of survival and the massive genetic imbalance (an extra copy of an entire chromosome representing 300-400 genes), very little research has been aimed at understanding the molecular consequences of trisomy 18. To fill this gap in knowledge, we have initiated a study to show that the extra copy of the chromosome 18 provokes genome wide expression dysregulation. Starting with fibroblasts samples from two live births with trisomy 18 and two age and sex matched controls, we used RNAseq (in triplicate) to examine genome-wide expression differences. As a first approach to data analysis, we have performed a mixed effects model with case/control status as the outcome and using sex and individual ID as co-variables. We adjusted for sex due to the observed higher loss of male fetuses that translates into a reduced male-to-female ratio at birth for trisomy 18. Our preliminary analysis has yielded promising results. We have identified genes as well as metabolic pathways significantly different between cases and controls. Among the most significant ones we find pathways involved in heart development and limb morphogenesis. For example, *TBX5*, a gene involved in heart development, had much lower expression levels in cases compared with controls. Altered expression of genes involved in limb development included *ALX3*, *ALX4*, *EN-1* and *FMN1*. Additionally, we found *PIEZO2* to be dysregulated, a gene known for its association with distal arthrogryposis and a feature of trisomy 18. We are in the process of using qPCR to confirm these results obtained through RNAseq. Our preliminary results show, for the first time, candidate genes that begin to explain some of the phenotypes observed in Edward syndrome. Our planned studies will confirm our current results and further investigate the alterations in gene expression due to the extra chromosome 18.

402F

Integration of next-generation mapping and sequencing technologies for identification of pathogenic structural variants. *E. Vilain^{1,2}, S. Bhattacharya¹, J. LoTempio^{1,3}, M. Almalvez¹, B. Dufault¹, R. Wang⁴, E. Douine⁴, E. Délot^{1,2}, S. Nelson⁴, H. Barseghyan¹, UCLA UDN Clinical Site, Undiagnosed Diseases Network.* 1) Center for Genetic Medicine Research, Children's National Health System, Children's Research Institute, Washington, DC 20010, USA; 2) Genomics and Precision Medicine, George Washington University Washington, DC 20052, USA; 3) Institute for Biomedical Science, George Washington University Washington, DC 20052, USA; 4) Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, California 90095, USA.

Next-generation sequencing (NGS) provides clinical diagnosis for ~30% of undiagnosed cases. However, due to innate methodological limitations, NGS fails to identify large structural variants (SVs), such as deletions, insertions, inversions, translocations and copy number variants, especially in repetitive regions of the genome. To overcome these limitations, we performed next-generation mapping (NGM), which captures a pattern of fluorescent labels in native-state, megabase-size DNA molecules in nanochannel array for de novo genome assembly and SV calling. We recently reported on NGM's clinical utility and capacity to identify large pathogenic SVs in a series of 10 Duchenne muscular dystrophy cases. We successfully identified all major SV types (deletions, insertions and balanced inversions) ranging from 13 kbp to 5.1 Mbp impacting the DMD locus. We have now performed NGM on an additional 100 patient trios with a variety of phenotypes, including autism, developmental delay, and disorders of sex development. On average, we identified ~2000 insertions, ~1000 deletions, ~250 inversions and 0-1 translocations per individual, showing that, while less common than single nucleotide variants, SVs account for a substantial fraction of genetic variation. The Database of Genomic Variants, internal dataset frequency and data from 150 healthy individuals were used to filter out common variants. In addition, we identified 2-5 rare de novo SVs in each proband. One of the defining features of NGM is its capability to generate de novo genome assembly with contiguous maps spanning entire chromosome arms. However, SV breakpoint resolution is limited by the intervals between fluorescent dye labeling motifs (6-10 kb intervals in the human genome). Using the available whole genome sequencing data for many of these cases to realign short reads around the SV breakpoints predicted by NGM decreased breakpoint uncertainty from 3-5 kbp to a few bases. Additionally, RNA sequencing was performed on whole blood for identification of variants affecting gene expression and splicing. Here, we show the ability of NGM to detect large pathogenic structural variants otherwise missed by sequencing or chromosomal microarrays. We developed methods to integrate NGS and NGM data to provide the ability to survey variants ranging from single nucleotide to large SVs and link them with changes in gene expression. Integration of this triad of technologies is poised to identify novel disease causing variants.

403W

Identification of genetic modifiers by next-generation sequencing (NGS) of candidate genes in a 22q11.2 deletion syndrome patient. A.G. Dantas¹, N. Nunes¹, L.C. Santos¹, C.A. Kim², D.C.Q. Soares², V.A. Meloni¹, S.I. Belanger¹, G.G. Carvalheira¹, M.I. Melaragno¹. 1) Genetics Division, Universidade Federal de São Paulo, São Paulo, Brazil; 2) Genetics Unit, Instituto da Criança, Universidade de São Paulo, São Paulo, Brazil.

The 22q11.2 deletion syndrome (22q11.2DS) results from hemizygous deletions of chromosome 22, being the 3 Mb the most frequent one. It is a multi-system syndrome with a remarkable variability in the severity among individuals. Approximately 180 clinical manifestations have been described, including congenital cardiac malformations. Even though most 22q11.2DS patients present similar size deletions, the phenotype is highly variable among individuals, suggesting a role of single nucleotide variants (SNV) in genes mapped outside the deleted region. Here, we report on a female patient with 22q11.2DS in which we sequenced candidate genes to the phenotype of the syndrome, mapped on the remaining allele of 22q11.2, and also on chromosomes 1, 8, and 11. We found a missense SNV on *ZFPM2* (Zinc Finger Protein, FOG Family Member 2) (chr8:106813942 G>A), predicted as deleterious through the *in silico* tools *SIFT* and *PolyPhen2*. This variant was annotated on the dbSNP (rs187043152), where it was also classified as pathogenic. In the same patient, we found another SNV on the 3'UTR of *MAPK1* (Mitogen-Activated Protein Kinase 1) (chr22:22117065 G>A), predicted as deleterious by the tools *FATHMM* and *Mutation Taster* and already annotated on the dbSNP (rs58437134). *Mutation Taster* inferred that the variant leads to signal's change of mRNA PolyA tail of *MAPK1*. The function of *ZFPM2* was previously associated with heart disease. Its protein act as a cofactor of *GATA4* (GATA Binding Protein 4) during embryonic heart development. The *MAPK1* gene is mapped distal to the deleted region, and it is one of the factors that may contribute for craniofacial and cardiac outflow tract defects observed in patients with distal 22q11 deletions, due to deficiencies in neural crest *MAPK1* signaling (Breckpot et al, 2012). Interestingly, our patient does not present any cardiac defects, such as the patient with 22q11.2DS reported by Chung et al (2015), who also presents a variant on *ZFPM2* and no cardiac defect. It has been suggested that a loss-of-function variant on one allele of *ZFPM2* may normalize expression of *GATA4* or other downstream genes, suppressing the cardiac phenotype observed on the patient. Thereby, the pathogenic variant found in the *ZFPM2* gene in our patient could have a role in the modulation of the cardiac phenotype, even with the presence of the 3 Mb deletion and the variant on *MAPK1*. **Financial support: FAPESP, Brazil.**

404T

Single-cell analysis of somatic mutations in aging of human liver. K. Brazhnik¹, S. Sun¹, A.Y. Maslov¹, O. Mohammed², M. Kinkhabwala³, A.W. Wolkoff⁴, J. Vijg⁴. 1) Department of Genetics, Albert Einstein College of Medicine (AECOM), Bronx, NY, USA; 2) Division of Transplant Surgery, Montefiore Medical Center, AECOM, Bronx, NY, USA; 3) Marion Bessin Liver Research Center, Division of Hepatology, AECOM and Montefiore Medical Center, Bronx, NY, USA; 4) Department of Anatomy and Structural Biology, AECOM, Bronx, NY, USA.

Liver is an essential organ with a high regenerative capacity that plays a central role in homeostasis and detoxification. It is associated with various age-related pathological processes, such as fatty liver disease, cirrhosis, hepatitis, infections and cancer, including primary liver cancer and metastatic invasion from other organs. Many functional changes in liver have been observed with aging, including a severe reduction in its metabolic capacity. Due to the intense metabolic activity liver is a highly genotoxic tissue with many sources of spontaneous mutagenesis. Hence, somatic mutagenesis in liver cells may be an important causal factor in human aging and disease. Indeed, genome instability, i.e. genome structural variants (SVs; also called large genomic rearrangements), base substitutions and small indels, copy number variation and retrotransposition, is a hallmark of aging and likely to be a major causal contributor to age-related diseases, including but not limited to cancer. However, *de novo* mutations cannot be assessed in bulk DNA from the whole tissue due to their dispersed nature and low abundance. Each individual cell has its own unique mutations, which cannot be detected by analyzing cells collectively. Nevertheless, mutation loads may rise to levels in each cell that affect their function. We used our advanced next-generation sequencing (NGS)-based approach to analyze the mutational landscape across the genome in individual human liver cells in relation to age, liver health status, and differentiation stage, i.e., stem or mature cells. We found that somatic mutations in human liver (1) dramatically accumulate with age; (2) accumulate faster in diseased liver as compared to age-matched normal liver; (3) occur at lower frequency and spectrum in liver stem cells than in differentiated hepatocytes; (4) display signatures that reflect age level, disease status and stem versus differentiated cells. Potentially, these mutations affect the functional genome, e.g., occur in transcribed, liver-specific genes and, hence, may contribute to liver functional failure and aging. **References:** 1. Dong, X. *et al.* Accurate identification of single-nucleotide variants in whole-genome-amplified single cells. *Nat Methods* **14**, 491-493, doi:10.1038/nmeth.4227 (2017).

405F

Genome-wide analysis of stem cell mutational signatures in healthy tissues. M. Eriksson¹, H. Helgadottir¹, I. Franco¹, A. Moggio², P. Vrtacnik¹, M. Larsson³, N. Norgren⁴, A. Johansson⁵, P. Lundin⁶, P. Stenvinkel⁷, L. Wennberg⁸, T. Lundgren⁸. 1) Biosciences and Nutrition, Karolinska Institutet, Huddinge, Sweden; 2) Dept of Medicine Huddinge, Karolinska Institutet; 3) Science for Life Laboratory, Dept of Physics, Chemistry and Biology, Linköping University; 4) Dept of Molecular Biology, National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Umeå University; 5) Science for Life Laboratory, Dept of Cell and Molecular Biology, Uppsala University; 6) Science for Life Laboratory, Dept of Biochemistry and Biophysics, Stockholm University, and Department of Biosciences and Nutrition, Karolinska Institutet; 7) Dept for Clinical Sciences, Intervention and Technology; Division of Renal Medicine, Karolinska University Hospital, Huddinge; 8) Dept for Clinical Sciences, Intervention and Technology; Division of Transplantation Surgery, Karolinska University Hospital, Huddinge, Sweden.

Aging is associated with reduced health and a significant part of the population suffers from age-related diseases or diseases mostly seen in the elderly. Somatic mutation burden increases during a lifetime as a result of accumulating errors, occurring either during cell division or because of environment-induced DNA damage. While somatic mutations are well studied in the context of cancer, less is known about the consequences of mosaicism for aging-related degeneration of generally healthy tissues. In the last years, we have developed a protocol that analyze somatic mutation load in cell clones from single-cell plated stem cells from various tissues. Our results show that somatic mutations (SNVs and indels) accumulate with aging in skeletal muscle, kidney, skin, subcutaneous and visceral fat progenitor cells. The rate of somatic mutation accumulation is highest in a subset of kidney progenitors, with 60 SNVs per year, compared to other progenitors that range 10-30 SNVs per year. The process of somatic mutation accumulation is not random and show a protection of functional regions, including exons. In certain progenitors this protection is reduced with aging resulting in the accumulation of mutations in exons. Somatic mutation accumulation follows specific mutational signatures, including signatures 1, 5 and 8 previously associated with certain forms of cancer. Clustering of samples on tri-nucleotide context identified three major sample groups, less dependent of age and tissue. This suggest that the majority of progenitors are subjected to the same mutational events, where specific cell-type and/or its location within tissues are less important determiners of somatic mutagenesis.

406W

Multi-tissue discovery of postzygotic variation in healthy humans. S. Vattathil^{1,2}, R. de Vito^{2,3}, B. Engelhardt^{2,3}, J.M. Akey^{1,2}. 1) Dept. of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ; 2) Lewis Sigler Institute, Princeton University, Princeton, NJ; 3) Dept. of Computer Science, Princeton University, Princeton, NJ.

Somatic mutations are known to drive many forms of cancer and contribute to an increasingly large catalog of human diseases. Yet, theoretical and empirical studies suggest that the burden of somatic mutation in humans is high and imply substantial genomic robustness to post-zygotic variation. Therefore, understanding the patterns and pressures of somatic variation in normal tissues is critical for estimating the impact of variants observed in disease. Specific questions of interest include the frequency with which cancer-linked mutations occur without precipitating a cancer phenotype, the distribution of observable mosaic mutations over developmental time and how mutational timing affects spatial distribution and mutation frequency, and the variation of mutational signatures across tissues. In an effort to address these questions, as part of the enhanced Genes, Tissues, and Expression (eGTEx) Project, we sought to identify somatic single-nucleotide variants from high-coverage (>150x) exome sequencing of 267 tissue samples collected through rapid autopsy from 14 donors. Samples were collected from a wide range of tissue types including skin, gastrointestinal organs, muscle, kidney, liver, and brain. Each donor was represented by 11 to 32 tissue samples, with a median of 18 tissue samples per donor. We designed a novel somatic variant calling framework to take advantage of the unique structure of our data set, and jointly called variants across tissues. The multi-tissue sampling structure also allowed us to infer the developmental timing of mutations by classifying mutations as tissue-specific or multi-tissue, and to compare somatic mutation profiles across tissue types. Finally, we leveraged RNA sequence data available for our study samples to both confirm somatic mutations and assess their regulatory consequences. We expect that our observations will further illuminate the landscape and determinants of human somatic variability, deepen our understanding of human development and mutational processes and mechanisms, and improve interpretation of mutations observed in cancer samples.

407T

Genetic basis for APOBEC3H alternative splicing, L1 exonization, and interaction with HIV-1 protease in sub-Saharan African populations. D. Ebrahimi¹, C. Richards¹, M. Carpenter^{1,2}, R. Harris^{1,2}. 1) University of Minnesota, Minneapolis, MN; 2) Howard Hughes Medical Institute.

Human APOBEC3H (A3H) is a single-stranded DNA cytosine deaminase and RNA binding protein with roles for innate antiviral immunity and in cancer mutagenesis. Seven haplotypes (I-VII) and four splice variants haplotype (SV154/182/183/200) have been described. A3H-II predominates in sub-Saharan Africa and is a potent viral restriction factor. In contrast, A3H-I is common outside Africa and has weak antiviral activity. Other A3H haplotypes encode defective proteins and/or are less common. Here we report the discovery of 6 additional A3H haplotypes (VIII-XIII) and, most strikingly, an intronic deletion (Δ ctc) embedded within the A3H-II genotype. This deletion induces the exonization of pieces of an antisense LINE1 element leading to alternative mRNA splicing and ultimately, production of 200 amino acid A3H enzyme (SV200). Our data indicate that A3H SV200 enzyme is present in 80% of sub-Saharan African populations and is at least 4-fold more HIV-1 restrictive than the other shorter splice variants. Furthermore, virus infectivity experiments revealed an unexpected role for HIV-1 protease in cleaving SV200 isoform into a variant with less restriction activity inside virus particles. Our further genomic and transcriptomic analyses revealed that Δ ctc and SV200 expression are not present in non-human primates but Δ ctc is present in the Neanderthal and Denisovan genomes. Taken together, these data suggest that A3H-mediated immunity against pathogens is more diverse and complex than previously thought, and HIV-1 uses its cognate protease, in addition to vif, as an additional layer of protection against restriction by A3H in sub-Saharan Africa.

408F

The commitment complex *in vivo* over long intron genes. G. Lev Maor, Y. Leader, G. Ast. Tel Aviv University, Tel Aviv, Israel.

The first and critical step in mRNA splicing is identification of the two splice sites delimiting an intron and bring them into proximity – termed the commitment complex. Within the commitment complex U1 small nuclear ribonucleoprotein particle (snRNP) binds the 5' splice site and form a protein bridge with U2AF65 (also called U2AF2) that binds the 3' end of that intron. Although the commitment complex was study extensively *in vitro*, how the same pairing occurs *in vivo* for splice sites separated by long introns is still obscure. Our laboratory proposed a new model, suggesting a dynamic pairing between the two splice sites involving interplay of RNA polymerase II (RNAPII), splicing factors and chromatin organization. We propose that U1 snRNP and U2AF65, associate with the CTD of RNAP II when that travels from one end of an intron downstream. We propose that U1 binds to 5' splice site while being associates to the CTD of RNAPII. When the 3' end of the intron is synthesize, U2AF65 binds the 3' splice site and the commitment complex is formed. Our research aim to uncover paramount mechanistic aspects of co-transcriptional splicing by experimental means; by identify and characterize co-transcriptional associations between splicing factors and pre-mRNA that are RNAPII-dependent; and by revealing the fashion by which the splicing reaction and splicing complex assembly occur in relation to RNAPII processivity. This will be performed using novel methodologies, including CRISPR interference-based and single cell techniques. We aim to reveal, for the first time how exons are selected by the splicing machinery co-transcriptionally *in vivo*, even when flanked by vastly longer introns.

409W

Comparison of chitinolytic properties of mouse Chit1 and AMCase with *Serratia marcescens* chitinase B. M. Kimura, T. Umeyama, S. Wakita, K. Okawa, M. Sakaguchi, F. Oyama. Department of Chemistry and Life Science, Kogakuin University, Hachioji, Tokyo, Tokyo, Japan.

Chitinases hydrolyze chitin, a polymer of β -1, 4-linked *N*-acetyl-D-glucosamine (GlcNAc). Humans and mouse express two active chitinases, chitotriosidase (Chit1) and acidic mammalian chitinase (AMCase). Chit1 is synthesized by activated human macrophages and neutrophils. Chit1 is increased in individuals with Gaucher disease, atherosclerosis disease and Alzheimer's disease. AMCase is increased in individuals with mouse models of asthma. However, the contribution of mammalian chitinases to the pathogenesis of these diseases and the physiological condition remains to be determined. *Serratia marcescens* express high levels of chitinases, Chitinase A (*Serratia* ChiA), Chitinase B (*Serratia* ChiB) and Chitinase C (*Serratia* ChiC). *Serratia* ChiB is a well-studied bacterial chitinase and structurally similar to the mammalian chitinases. Previously, we successfully expressed mouse AMCase and Chit1 as recombinant proteins (Protein A-AMCase-V5-His and Protein A-Chit1-V5-His, respectively) in *Escherichia coli*. In this study, we compared mouse mammalian chitinases with *Serratia* ChiB. AMCase had the highest activity at around pH 2.0 among their enzymes. Chit1 and AMCase show about 1/2 lower chitinolytic activity than *Serratia* ChiB under acidic to neutral conditions. Our results indicate that mammalian chitinases possess chitinolytic activity comparable to *Serratia* ChiB.

410T

Evaluations of enzymatic properties for human and mouse chitotriosidase. T. Watanabe, M. Kimura, S. Wakita, K. Sekine, M. Sakaguchi, F. Oyama. Department of Chemistry and Life Science, Kogakuin University, Hachioji, Tokyo, Japan.

Chitin, a linear polymer of β -1, 4-linked *N*-acetyl-D-glucosamine (GlcNAc), is an integral component of the exoskeletons of crustaceans and insects, the microfilarial sheaths of parasites and the fungal cell walls. Although mammals do not produce chitin, two active chitinases, chitotriosidase (Chit1) and acidic mammalian chitinase (AMCase), have been identified in mouse and human. Chit1 was the first mammalian chitinase to be purified and cloned. Chit1 have attracted considerable attention due to their increased expression in individuals with different pathological conditions including Gaucher disease, chronic obstructive pulmonary disease (COPD), Alzheimer's disease and diabetes mellitus. Thus, Chit1 may play important roles in many pathophysiological conditions. However, the contribution of Chit1 to the pathophysiology of these diseases remains to be determined. Mouse Chit1 shows 75% identity and 78% similarity to human Chit1 based on their primary structure alignment. In this study, we compared the enzymatic functions between human and mouse Chit1 under the same conditions. To characterize their properties, we examined the chitinolytic activity using 4-nitrophenyl *N,N'*-diacetyl- β -D-chitobioside as a substrate under various pH conditions. Human Chit1 was active in more wider pH range than mouse Chit1. Furthermore, when the products from the substrate were analyzed using fluorophore-assisted carbohydrate electrophoresis (FACE) method, human Chit1 exhibited more transglycosylation activity than mouse enzyme under tested condition. Our results suggest that there are remarkable differences between human and mouse Chit1 proteins.

411F

Time and space dimensions of gene dosage imbalance of aneuploidies revealed by single cell transcriptomes. *F.A. Santoni¹, G. Stamoulis², M. Garieri², A. Letourneau², M. Guipponi⁴, N. Panousis², E. Falconnet², P. Ribaux², C. Borel², S.E. Antonarakis^{2,3,4}.* 1) Service of Endocrinology, Diabetes and Metabolism, University Hospital of Lausanne - CHUV, Lausanne, Switzerland; 2) Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland; 3) iGE3 Institute of Genetics and Genomics of Geneva, University of Geneva, Geneva, Switzerland; 4) Geneva University Hospitals, Service of Genetic Medicine, Geneva, Switzerland.

The mechanisms underlying cellular and organismal phenotypes due to copy number alterations (CNA) are not fully understood. Aneuploidy is a major source of gene dosage imbalance due to CNA and viable human trisomies are model disorders of altered gene expression. To understand the cellular impact of gene dosage imbalance, we studied gene and allele specific expression (ASE) of 9668 single-cell fibroblasts in trisomies T21, T18, T13 and T8. To limit the bias of interindividual noise, all comparisons between euploid and trisomic single-cells were performed on an isogenic setting for all trisomies studied. Initially we examined 928 single cells with deep RNA-Seq. For T21 we used fibroblasts from one pair of monozygotic twins discordant for T21 and from mosaic T21. For T18, T13 and T8 we analyzed single cells from mosaic individuals. Single-cell analyses revealed inconsistencies concerning the overexpression of some genes observed in differential trisomic vs euploid bulk RNAseq while this imbalance was not detectable in trisomic vs. euploid single cells. Moreover, ASE profiling of all single cells uncovered a substantial monoallelic pattern of expression in the trisomic fraction of the genome. By classifying genes according to the level of mono and bi-allelic transcription, we have observed that, for genes with monoallelic and low-to-average expression, the altered gene dosage is mainly due to the higher fraction of cells simultaneously expressing these genes in the trisomic samples. These results were confirmed in a further experiment of 8740 single fibroblasts from the monozygotic twins discordant for T21 samples. We conclude that gene dosage imbalance is of bidimensional nature: over time (simultaneous expression of all alleles resulting in increased accumulation of RNA of copy altered genes in each single cell) as previously stated, and over space (increased fraction of cells simultaneously expressing copy altered genes). These results strongly suggest that each class of genes contributes to the phenotypic variability of trisomies according to its temporal and spatial behavior and propose an improved model to understand the effects of copy number alterations.

412W

Chitinase 3-like-1 with amino acid substitutions at the active site remains inactive. *N. Kishigami, K. Okawa, M. Sakaguchi, F. Oyama.* Department of Chemistry and Life Science, Kogakuin University, Hachioji, Tokyo, Japan.

Chitinase-like proteins (CLPs) are structurally homologous to chitinases but lack the ability to degrade chitin. Based on sequence similarities, CLPs belong to the family 18 of the glycosyl hydrolases. The conserved sequence involved in catalysis in family 18 of the chitinases is DXXDXDXE. It is generally assumed that the lack of chitinase activity in CLPs is due to the mutation of crucial residues within the conserved catalytic sequence. Chitinase 3-like-1 (Chi3l1), one of the CLPs, is expressed in mouse. Its levels are increased in individuals with asthma, chronic obstructive pulmonary disease (COPD), inflammatory bowel disease, cancer, Alzheimer's disease and different types of malignant tumors. Presently, the exact physiologic and biological functions of Chi3l1 remain to be determined. Despite Chi3l1 has a relatively high sequence identity to chitotriosidase (Chit1), it lacks chitinase activity. In this study, we attempted to activate Chi3l1 by introducing two substitutions of amino acid (A139D and L141E) in the active center of the wild type Chi3l1. The Chi3l1 with amino acid substitutions at the active site remains inactive. This result suggests that lack of chitinase activity in Chi3l1 is not due to merely the simple amino acid substitution in the conserved catalytic motif.

413T

Quantification of chitinases mRNA levels by qPCR in crab-eating monkey tissues: Comparison with mouse and human. *M. Uehara, E. Tabata, M. Ohno, M. Sakaguchi, F. Oyama.* Department of Chemistry and Life Science, Kogakuin University, Hachioji, Tokyo, Japan.

Chitin is a polymer of *N*-acetyl-D-glucosamine. Although mammals do not contain endogenous chitin, mammals express two active chitinases, acidic mammalian chitinase (AMCase) and chitotriosidase (CHIT1). Because the level of expression of these chitinases is increased in many inflammatory conditions including Gaucher disease and mouse models of asthma, both chitinases may play important roles in the pathophysiological conditions. Crab-eating monkey is one of the most important nonhuman primate animal models in basic and applied biomedical researches. We performed gene expression analysis of two chitinases in normal crab-eating monkey tissues by quantitative real-time PCR (qPCR) system to quantify the chitinases and reference genes on the same scale. Levels of AMCase and CHIT1 mRNAs were highest in the stomach and the lung, respectively, among tested other tissues. Next, we performed comparative gene expression analysis of mouse, monkey, and human. Gene expression of AMCase and CHIT1 varies between tested species. The differences of mRNA expression between the species in the stomach tissues were basically reflecting the levels of the chitinolytic activities. These results indicate that difference of their gene expression level between mammalian species and requiring special attention in handling data in chitinase-related studies in particular organisms.

414F

RegulomeDB: A resource for the human regulome. *Y. Luo¹, S. Dong², T. Dreszer¹, N. Ouyang², B. Hitz¹, J. Cherry¹, A. Boyle^{2,3}.* 1) ENCODE Data Coordination Center, Department of Genetics, School of Medicine, Stanford University, Stanford, CA; 2) Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI; 3) Department of Human Genetics, University of Michigan, Ann Arbor, MI.

The rapid growth of human genome sequencing has identified a large number of genetic variants present in the human genome relevant to human disease and other phenotypes. The majority of sequence variation is found within the noncoding regions of the genome where it is presumed to affect gene regulatory elements. We have developed a unique web-accessible database, RegulomeDB, which provides integrated knowledge of existing information describing regulatory elements that lie within noncoding regions. The RegulomeDB v1.1 (<http://www.regulomedb.org/>) was originally built from ENCODE and other community data, and has proven to be a useful tool for noncoding variant annotation. Various genomic information, including eQTL, transcription factor binding, chromatin modification and DNA methylation have been integrated into a common database and displayed at nucleotide resolution. Researchers are able to compare variants of their interest to the wealth of information in RegulomeDB, and rapidly gain insight into possible mechanism of gene regulation. In order to keep up with emerging genome sequencing projects and provide an up to date regulome annotation, we have integrated RegulomeDB into the ENCODE portal, as RegulomeDB v2.0. We are also investigating improved annotation with new data types and integration of novel, statistical scoring methods. Overall, we expect this unique resource to have wide impact in the biomedical community.

415W

Structural and functional network abnormalities in the human brain of XYY syndrome. S. Liu, L. Clasen, F. Lalonde, J. Blumenthal, A. Raznahan. Developmental Neurogenomics Unit, National Institutes of Mental Health, Bethesda, MD.

XYY syndrome (47, XYY) is one of the frequent sex chromosome aneuploidy disorders (1/1000 live male births) characterized by duplication of Y-Chromosome in males. It has been associated with cognitive deficits and previous studies identified abnormal gray matter volume (GMV) changes in isolated brain regions using the plain T1 anatomical magnetic resonance imaging (MRI). Here we ask whether XYY syndrome further alters the structural and functional networks in the human brain, which could be measured using advanced MRI techniques. In the current study, we recruited 65 XYY participants (age [mean/SD]: 13.0/5.7 years) and 62 age-matched healthy male participants (13.2/5.1 years) as controls, to our knowledge, which represents the largest neuroimaging study of XYY to date. Genetic diagnosis of XYY was confirmed by karyotype. All MR images were collected on a 3T GE whole-body scanner (Discovery MR750). For each participant, a high-resolution T1 weighted anatomical image, a set of 10-minute resting-state functional MRI (rs-fMRI) images, and a group of 60-direction diffusion tensor imaging (DTI) images were obtained. Standard pipelines were used to analyze the data and two-sample t-tests were used to compare XYY and controls. The significance level of family wise error (FWE<0.05) was used to correct for multiple comparisons across voxels. We reproduced previous T1 findings and identified more regions showing group differences, which is likely due to the higher degrees-of-freedom in this study. In brief, compared to controls, XYY participants showed significant GMV loss in bilateral insula and anterior cingulate cortices, cerebellum and subcortical structures including the caudate, putamen and thalamus, as well as gain in the posterior cingulate and bilateral prefrontal and inferior parietal cortices. Further DTI and fMRI analyses of structural and functional connectivity confirmed the convergent and divergent findings in these highlighted regions. Taken together, abnormal changes in these areas and their connections, which have been linked to executive control and other high-order cognitive functions, may entail the increased risk of behavioral deficits in XYY syndrome.

416T

Simultaneous analysis of long-range chromatin interactions and DNA methylation by bulk and single cell Methyl-HiC. G. Li¹, Y. Liu^{3,4}, M. Kellis^{2,4}, B. Ren^{1,2}. 1) Ludwig Institute for Cancer Research, La Jolla, CA; 2) University of California, San Diego, School of Medicine, Department of Cellular and Molecular Medicine, La Jolla, CA; 3) Broad Institute of MIT and Harvard, Cambridge, MA; 4) Massachusetts Institute of Technology, Computer Science and Artificial Intelligence Laboratory, Cambridge, MA.

DNA methylation plays an important role in mammalian gene regulation. Although genome-wide bisulfite sequencing can measure DNA methylation levels at a single nucleotide resolution, it is difficult to measure the methylation status of two cytosines in single DNA molecule separated by a large genomic distance due to short DNA fragment sizes below kilobases scale. Here, we developed a new approach to study relationships between methylated cytosines across large genomic distances genome-wide. This approach, named Methyl-HiC, combines *in situ* Hi-C with bisulfite sequencing to simultaneously capture chromosome conformation and methylation status of cytosines on the same DNA molecules. Methyl-HiC generates consistent information to both *in situ* Hi-C and WGBS data from the same cell line when treated separately. We can detect allelic-specific methylation over a large genomic distance by incorporating allelic nucleotides and CpG methylation on the same DNA molecules. By analyzing the CpG methylation from the single molecule, we found that DNA methylation status is generally concordant between two cytosines that are spatially close. The levels of concordance vary depending on the chromatin states of the interacting DNA. We showed that this long-range DNA methylation concordance is further enhanced in cells lacking the Ten-eleven Translocation (TET) methylcytosine dioxygenases. We further developed the MethylHiC method to single cell level, which showed comparable quality to previous single cell HiC and single cell DNA methylome. Chromosome conformation difference can be observed in subpopulations divided by DNA methylation signature. Our method here paves the road to evaluate the direct long-range effect of genetic and epigenetic alterations in different pathological conditions in genome-wide and single cell level.

417F

Deep characterization of the contribution of short tandem repeats (STRs) to gene expression across tissues. *S. Feupe, Fotsing, M.G. Gymrek.* UC San Diego, La Jolla, CA.

The identification of STRs contributing to the variation of gene expression in relevant tissue can be helpful in the identification of disease causing gene or other phenotype causing genes. Thousands of genetic variants contributing to gene expression variation have been identified in the past years through large scale studies and traits focused studies. However, these efforts have largely focused on SNPs, ignoring more complex variants such as repetitive regions and structural variants. This is due both to bioinformatic limitations in genotyping complex variants from short reads and low coverage sequencing data, and the challenges of performing traditional association tests at highly multiallelic loci. STRs have been implicated in dozens of Mendelian diseases such as Huntington's disease and Fragile X Syndrome and have been shown to affect human phenotypes through a variety of mechanisms. Recent evidence suggest an important role of STRs in gene regulation. For instance, a genome-wide survey of the effects of STRs on gene expression identified more than 2,000 STRs whose lengths were linearly associated with expression of nearby genes (eSTRs). These STRs explained 10-15% of gene expression heritability due to cis variants, suggesting an important role for STRs in complex traits. That study along with others have faced several limitations: Short read lengths and low coverage sequencing resulted in low quality genotypes, reducing power to detect true associations - Restriction to single tissue or cell type, limiting biological insight and generalizability of the results in different contexts. Here we look at influence of STRs on gene expression across multiple tissue types. We used HipSTR, a new and accurate STR profiling algorithm to genotype STRs in 650 high coverage (30x) whole genome sequencing dataset from the Genotype-Tissue Expression (GTEx) project. We performed association tests relating STR lengths and expression levels of nearby genes, and identified more than 9000 genes with eSTRs (eGenes) across seventeen tissues at 10% FDR. 38.7% of eGenes were shared by two or more tissues. Finally, using the extensive GTEx resource we fine-map the contribution of different variants classes to gene expression, identify putative causal regulatory variants overlapping GWAS loci, and characterize tissue-specific features of predicted causal eSTRs. Overall, our analysis provides a valuable resource for future efforts to integrate STRs into complex trait analyses.

418W

Gene expression analysis of mammalian chitinases in common marmoset (*Callithrix jacchus*) tissue. *E. Tabata^{1,2}, A. Kashimura¹, M. Uehara¹, S. Wakita¹, M. Sakaguchi¹, Y. Sugahara¹, T. Yurimoto³, E. Sasaki³, F. Oyama¹.* 1) Department of Chemistry and Life Science, Kogakuin University, 2665-1 Nakano, Hachioji, Tokyo 192-0015, Japan; 2) Research Fellow of Japan Society for the Promotion of Science (DC1), Koujimachi, Chiyoda-ku, Tokyo 102-0083, Japan; 3) Central Institute for Experimental Animals, Tonomachi, Kawasaki, Kanagawa, 210-0821, Japan.

Chitin, a polymer of *N*-acetyl-D-glucosamine (GlcNAc), functions as a major structural component in crustaceans, insects and fungi and is the second most abundant polysaccharide in the nature. Although mammals do not produce chitin and its synthase, genes encoding chitotriosidase (CHIT1) and acidic chitinase (CHIA) and their translation products have been found in both human and mouse. CHIT1 levels are elevated in the plasma of Gaucher disease and the bronchoalveolar lavage fluid of smokers with chronic obstructive pulmonary disease (COPD). CHIA expression and activity is upregulated in allergic airway responses in mouse models of asthma. However, the expression levels of chitinases in the other mammals have remained unknown. Common marmoset is an insectivorous New World monkey and attractive as a biomedical model for human diseases due to its handling size and genetic similarities. Here, we quantified expression levels of the chitinases and reference genes in common marmoset tissues on the same scale using a quantitative reverse transcriptase-coupled PCR. The relatively high levels of CHIT1 mRNA were detected in the lung, liver and spleen. Clear tissue-specific pattern was observed in CHIA mRNAs expression. CHIA mRNA was predominantly detected in the stomach and the other tissues were similar to or lower than that of CHIT1. Our study suggests that CHIA functions as a digestive enzyme that breaks down chitin in common marmoset stomach tissues.

419T

Composite *de novo* Armenian human genome assembly and haplotyping via optical mapping and ultra-long read sequencing. H. Barseghyan¹, S. Bhattacharya¹, J. LoTempio^{1,2}, A. Almalvez¹, B. Dufault¹, E. Vilain^{1,3}. 1) Center for Genetic Medicine Research, Children's National Health System, Children's Research Institute, Washington, DC 20010, USA; 2) Institute for Biomedical Science, George Washington University Washington, DC 20052, USA; 3) Genomics and Precision Medicine, George Washington University Washington, DC 20052, USA.

Commercialization of short-read sequencing (SRS) in the past decade allowed characterization of millions of small genetic variations across thousands of human genomes for genetic diagnosis and disease susceptibility associations. However, due to its innate methodological limitations, SRS misses on identification of large structural variants (SVs), such as deletions, insertions inversions, translocations and copy number variants, which account for the largest human-to-human genetic variability and have been shown to be implicated in disease. Hence, there is a need for large-scale hybrid technology approach for *de novo* human genome assembly and variant calling. Advances in genomic technologies now allow researchers to survey the diploid architecture of human genomes across different ethnic groups. Here, we report the *de novo* assembly and phasing of the Armenian human genome with the precise localization of epigenetic markers. The hybrid genome assembly and phasing was performed by incorporating optical mapping and long-read nanopore sequencing (LRS). Optical mapping molecule N50 values were >200 kbp whereas the LRS N50 values were > 100 kbp. Coupling of both technologies allowed for an assembly of highly contiguous haplotype specific scaffolds covering entire chromosome arm lengths. Direct comparison of the assembly with the human genome reference GRCh38 identified thousands of SVs, many of which were previously unseen. Epigenetic modifications were also detected from LRS analysis and localized across the scaffolds generating the epigenetic profile of the Armenian genome. This is the first report of its kind that includes both *de novo* genome assembly of an isolated ethnic group (Armenia) as well as localization of epigenetic markers. This work serves as a precursor of precision medicine as the identified Armenian-specific structural variants may have a clinical relevance in diseases such as Familial Mediterranean Fever, which has a higher frequency in the Armenian population.

420F

Use of nanopore sequence to validate and refine genome assembly.

D. Mohr, A.F. Scott. Institute of Genetic Medicine, Johns Hopkins University School of Medicine, 1812 Ashland St., Baltimore MD 21205.

De novo genome assembly is becoming increasingly common as technologies and analysis tools improve and become more cost-effective. One limitation that remains is that long runs of repetitive sequence are difficult to resolve and can lead to gaps within scaffolds or breaks between scaffolds. We previously used linked read sequencing and optical mapping to create a genome for an endangered species. Overall sequence contiguity was excellent with an N50 of 61.7 Mb and 99% of the predicted genome occurring in 104 scaffolds. However, we observed some regions of disagreement between DNA lengths estimated from the optical maps and the actual sequence assembled with Supernova. These occurred, as expected, at regions of repetitive DNA where they manifested as read pile-ups. Because nanopore sequencing is agnostic to repetitive motifs we made long DNA libraries using Rapid chemistry (RAD004) and ran these on Oxford Nanopore R9.4.1 flow cells on a GridION instrument using MinKNOW 1.11. Reads were aligned to our existing scaffolds using minimap2, with ~91% identity. We obtained reads with an N50 of over 30kb and up to 235kb, >90% of which aligned extremely well with our published genome, thereby validating the linked-read approach used for assembly. Several regions were identified where the assembly improved with the use of long nanopore reads. We also have tested new approaches for the isolation of high molecular weight DNA from lymphocytes that produced long DNA suitable for nanopore sequencing. We conclude that nanopore sequencing is a valuable adjunct to more conventional NGS approaches and as throughput increases and accuracy improves will become an important tool in genome assembly.

421W

Validation of copy number variations on TA repeat regions using MinION.

T. Takahashi, C. Nakashima, F. Kitaoka, N. Amano, N. Takasu, M. Nomura.
Kyoto University, Kyoto, Japan.

We are developing a stock of clinical-grade induced pluripotent stem (iPS) cells. For their clinical sequencing, we have used illumina HiSeq to detect genomic mutations such as single nucleotide variations, insertions, deletions and copy number variations (CNVs). On occasion, we have detected low sequence depths on TA repeat regions, which may indicate deletion-like CNVs. However, because many of these regions have no clear breakpoints, we consider them to be false positives. To test this hypothesis, we conducted amplicon sequencing of TA repeat regions of human genome NA24385 using Oxford Nanopore MinION, which can handle much longer reads than HiSeq. We selected target TA repeat regions based on the RepeatMasker database. We sorted the TA repeat regions and chose the top 100 regions by length, but excluded 24 regions that proved difficult for sequencing. We designed 46 primer pairs for the remaining 76 regions and also for 4 known CNVs of NA24385. We performed long PCR to amplify 40 random candidate regions and 4 known CNVs. We sequenced three of the CNVs to compare deletion-like CNV on TA repeat regions and confirmed clear patterns of deletions in depth and breakpoint. Though these known CNVs are hetero-deletions, two of them appeared as homo-deletions when using MinION. The reason is that the short DNA molecules between primer pairs outside the CNVs were increased dominantly by the PCR reaction of the amplicon generation and that MinION preferentially sequenced the shorter amplicons. In fact, we got complete reads that support the existence of intact allele. Next, we sequenced 9 candidate TA repeat regions. Among them, 8 had no clear deletion patterns of CNVs, although they did show gradual changes in read depths and no breakpoints. We confirmed all 8 were false positives, but could make no conclusions about the remaining one. Our results suggest that amplicon sequencing with MinION is recommended to validate ambiguous CNV candidate regions found using HiSeq.

422T

A comprehensive map of cis-regulatory elements and 3D structure of the zebrafish genome. *T. Liu¹, H. Yang¹, Y. Luan^{1,2}, Y. Wang¹, G. Gerhard³, D. Balciunas⁴, K. Chen¹, F. Yue¹.* 1) Penn State College of Medicine, Hershey, PA; 2) Key Laboratory of Agricultural Animal Genetics, Breeding 17 and Reproduction of Ministry of Education and Key Laboratory of Swine Genetics and Breeding of Ministry of Agriculture, Huazhong Agricultural University, Wuhan 430070, China; 3) Dept of Biology, Temple University; 4) Dept of Medical Genetics and Molecular Biochemistry, Temple University.

Zebrafish has been widely used for the study of human diseases, as ~70% of the protein-coding genes are conserved between the two species. Moreover, zebrafish embryos are transparent and thus can serve as an ideal model for genetic studies in animal development. Surprisingly, the functional annotation of zebrafish genome itself has been severely lagging when compared with other model systems such as mouse and drosophila. Here we took a similar approach adopted by the ENCODE and Roadmap Epigenomics projects, and performed RNA-Seq and ChIP-Seq for H3K27Ac and H3K4me3 histone marks to generate a comprehensive map of transcriptomes and regulatory elements in a variety of zebrafish tissues, including brain, heart, liver, skeletal muscle and kidney. We predicted over 100,000 cis-regulatory elements in the zebrafish genome, the most comprehensive functional annotation effort in zebrafish so far to our knowledge. By comparing the data generated by the ENCODE and Roadmap Epigenomics projects, we also defined a set of functionally conserved and species-specific regulatory sequences among zebrafish, mouse and human. To further explore the three-dimension (3D) structure of zebrafish genome, we performed chromosome conformation capture (Hi-C) experiment using zebrafish adult brain and muscle. Our Hi-C matrix shows that topologically-associating domains (TADs) are also conserved among zebrafish, mouse and human. Muscle specific enhancer and promoter interactions predicted from ChIP-seq data are further validated by muscle Hi-C matrix. In summary, we generated a great genomics/epigenomics resource for the functional annotation and the 3D structure of the zebrafish genome and further expanded the value of zebrafish as a model of human disease.

423F

Progression of the molecular immune response in patients infected with Andes orthohantavirus (ANDV). G.Esteves. Ribeiro¹, P. Vial¹, A. Cuiza¹, M. Ferrez², M. Calvo³, J.Luis. Perez⁴, L. Ferreira⁵, M. Rioseco⁶, J. Graff⁷, L. Leon⁸, C. Vial¹. 1) Instituto de Ciencia e Innovación en Medicina, Universidad del Desarrollo, Santiago, Santiago, Chile; 2) Departamento de Enfermedades Infecciosas e Inmunología Pediátricas; 3) Hospital Base de Valdivia, Facultad de Medicina, Universidad Austral de Chile; 4) Hospital Base San José de Osorno, Facultad de Medicina; 5) Hospital Clínico Regional Guillermo Grant Benavente; 6) Hospital Puerto Montt Dr. Eduardo Schütz Schroeder; 7) Clínica Alemana, Facultad de Medicina, Universidad del Desarrollo; 8) Instituto de Ciencias Biomédicas, Ciencias de la Salud, Universidad Autónoma de Chile.

Hantavirus are important human pathogens that cause a severe zoonotic disease called hantavirus cardiopulmonary syndrome (HCPS). HCPS presents a fatality rate of ~30% in Chile. Clinical manifestations may present as a mild condition with moderate respiratory failure or progress quickly to a severe condition with shock that can be fatal. This progression towards HCPS is a complex multifactorial process that involves mechanisms directly induced by the virus or by the host response, but it is not known exactly what are the host response factors associated with the progression towards HCPS. Our goal is to better understand the role of the host's immune responses in this progression towards HCPS, by longitudinally characterize the transcriptional profile of the peripheral blood mononucleated cells (PBMCs) isolated from ANDV infected patients, and healthy subjects as controls. Total RNA was extracted from PBMCs isolated from peripheral blood of healthy controls and patients infected with ANDV on days 1, 3, 5 post hospitalization, during the acute phase of the disease, and on day 60, during the convalescent phase. Isolated RNA was sequenced at the Broad Institute (Boston, USA) using Illumina's HiSeq platform (50M reads aligned in pairs). We analyzed 10 patients and four healthy controls. Transcriptome analyzes were performed with STAR, Kallisto, Deseq2 and ShinyGO. The results show different patterns of gene expression during the progression of HCPS. Evaluation of gene ontology (GO) biological processes enrichment, showed a difference between samples from the acute phase and convalescent phase or healthy controls. We found that during the acute phase (days 1 and 3), there was an enrichment of GO processes related to cell activation, humoral immune response and inflammatory response. Then they decrease on day 5, and continues to decrease on day 60 to levels similar to healthy controls, suggesting the full recovery of these patients. Future directions of this study include the analysis of differential expression by gene, to help us understand the mechanisms involved in the pathogenesis of HCPS.

424W

De novo mutations shared by siblings. H.J. Jónsson¹, P.S. Sulem¹, G.A.A. Arnadóttir¹, G.P. Pálsson¹, H.P.E Eggertsson¹, S.K. Kristmundsdóttir¹, F.Z. Zink¹, B.K. Kehr¹, K.E.H. Hjorleifsson¹, B.O.J. Jónsson¹, I.J. Jónsdóttir¹, S.E.M. Marelsson², S.A.G. Gudjonsson¹, A.G. Gylfason¹, A.J. Jonasdóttir¹, A.J. Jonasdóttir¹, S.N.S. Stacey¹, O.T.M. Magnusson¹, U.T. Thorsteinsdóttir^{1,3}, G.M. Masson¹, A.K. Kong^{1,4}, B.V.H. Halldorsson^{1,5}, A.H. Helgason^{1,6}, D.F.G. Gudbjartsson^{1,4}, K.S. Stefansson^{1,3}. 1) deCODE genetics / Amgen Inc., Sturlugata 8, Reykjavik, Iceland; 2) Department of Pediatrics, Landspítali University Hospital, Reykjavik, Iceland; 3) Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik, Iceland; 4) School of Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland; 5) School of Science and Engineering, Reykjavik University, Reykjavik, Iceland; 6) Department of Anthropology, University of Iceland, Reykjavik, Iceland.

Somatic and germ cells are derived from a single zygote. De novo mutations (DNMs) that occur early may result in mosaicism and can be transmitted to more than one offspring. Rare diseases of childhood are often caused by DNMs and therefore early DNMs can cause recurrence of disease in siblings. We searched for early DNMs by scanning 1,007 sibling pairs from 251 families for DNMs shared by siblings (ssDNMs). To avoid excluding early DNMs, we used absence of a mutation in the blood of a haplotype-sharing sibling to detect DNMs in addition to absence of the DNM from the blood of the parents. We estimated the probability of DNM recurrence based on presence in the parental blood, sharing by other siblings, parent-of-origin, mutation type, and genomic position. We identified 878 DNMs shared by siblings (ssDNMs) at 448 genomic sites and found that majority of ssDNMs is present in parental blood (57.2%). The probability of a DNM being shared goes down by 2.27% per year for paternal DNMs and 1.78% for maternal DNMs. Shared paternal DNMs are more likely to be T>C mutations than maternal ones, but less likely to be C>T mutations. Depending on DNM properties, the probability of recurrence in a younger sibling ranges from 0.011% to 28.5%. The data available on a DNM can be used to assess the recurrence probability. We have launched an online DNM recurrence probability calculator, to explore our results in the context of genetic counselling.

425T

Associations of *CHRNA3*, *CYP2B6* and *UGT2B28* gene variants and urine levels of nicotine metabolites. G. Borrego-Soto¹, A. Rojas-Martinez¹, Y.X. Perez-Paramo², S.K. Santuario-Facio¹, R. Ortiz-Lopez¹. 1) Tecnológico de Monterrey, Monterrey, Mexico; 2) Washington State University, US.

Tobacco consumption is a risk factor for 6 of the top-eight causes of death. The metabolism of nicotine, tobacco's addictive substance, may be subject of genetic variability and differences in addiction susceptibility and detoxification. Nicotine addiction is associated to variants in acetylcholine receptors in the basal ganglia while nicotine detoxification involves two phases: in phase I, 70 - 80% of nicotine is metabolized to cotinine and this is subsequently metabolized to trans-3'-hydroxycotinine (3HC) by *CYP2A6* and *CYP2B6* (90 and 10%, respectively). In the phase II, nicotine, cotinine, and 3HC undergo detoxification by UDP glucuronosyltransferases (*UGT*) generating products that are eliminated in the urine. These enzymes are highly polymorphic. This study aimed at analyzing the levels of nicotine and its metabolites in the urine of Mexican smokers (1-5 cigars/day) to determine genotype-phenotype associations for variants involved in nicotine addiction and metabolism. 95 smokers (18 - 35 y.o.) were genotyped and their urine nicotine metabolites were measured by UPLC-MS-MS. Associations between nicotine metabolites and genetic variants were analyzed with the Mann-Whitney test (significance level $p \leq 0.05$). The CC genotype of rs2869547 (*CHRNA3*) was associated with high concentrations of cotinine-Gluc and low levels of HPBA. The GG genotype of rs2279343 (*CYP2B6*) was associated with low concentrations of free nicotine. Finally, the GG genotype of rs6828191 (*UGT2B28*) was associated with low levels of 3HC-Gluc. No associations were found with *CYP2A6* variants. This study suggests three genetic consequences for nicotine addiction and metabolism in Mexicans: The *CHRNA3* variant is possibly involved in tobacco addiction and the *CYP2B6* and *UGT2B28* variants maybe related to increased nicotine toxicity. These variants have been scarcely evaluated regarding their role in tobacco pharmacogenomics.

426F

Transcriptome response of human skeletal muscle to divergent exercise stimuli. M. Naymik¹, J. Dickerson², A. D'Lugos², A. Siniard¹, A. Wolfe¹, D. Curtis², M. Huentelman¹, C. Carroll^{1,4}. 1) Translational Genomics Research Institute, Phoenix, AZ; 2) School of Nutrition and Health Promotion, Healthy Lifestyles Research Center, 9 Exercise Science and Health Promotion, Arizona State University, Phoenix, AZ; 3) Midwestern University, Glendale, AZ; 4) Department of Health and 12 Kinesiology, Purdue University, West Lafayette, IN.

Aerobic (AE) and resistance exercise (RE) elicit unique adaptations in skeletal muscle that have distinct implications for health and performance. The purpose of this study was to identify the unique transcriptome response of skeletal muscle to acute AE and RE. In a counter-balanced, cross-over design, six healthy, recreationally active young men (27±3y) completed acute AE (40 min cycling, ~70% maximal HR) and RE (8 sets, 10 reps, ~65% 1RM), separated by ~1wk. Muscle biopsies (vastus lateralis) were obtained before and at 1 and 4h post exercise. Whole-transcriptome RNA sequencing (HiSeq2500, Illumina) was performed on cDNA synthesized from skeletal muscle RNA. Sequencing data were analyzed using HTSeq and differential gene expression was identified using DESeq2 (adjusted P value (FDR) < 0.05, > 1.5-fold change from preexercise). RE resulted in a greater number of differentially expressed genes at 1h (67 vs. 48) and 4h (523 vs. 221) compared to AE. We identified 348 genes that were differentially expressed only following RE, whereas 48 genes were differentially expressed only following AE. Gene clustering indicated AE targeted functions related to zinc interaction, angiogenesis, and ubiquitination, whereas RE targeted functions related to transcription regulation, cytokine activity, cell adhesion, kinase activity and the PI3K/Akt pathway. *ESRRG* and *TNFSRF12A* were identified as potential targets related to the specific response of skeletal muscle to AE and RE, respectively. These data describe the early post exercise transcriptome response of skeletal muscle to acute AE and RE and further highlight that different forms of exercise stimulate unique molecular activity in skeletal muscle.

427W

Functional annotation of eSNPs by CRISPR-Cas9 mediated genome editing. R. Tian¹, Y. Pan², T. Etheridge², G. Bao², C. Lee², G. Gibson¹. 1) Center for Integrative Genomics, Georgia Institute of Technology, Atlanta, GA; 2) Department of Bioengineering, Rice University, Houston, TX.

Precision identification of causal variants that are responsible for disease risk is one of the major challenges in interpretation of the molecular basis of genome-wide association signals. Since the majority of disease risk signals are located in the non-coding regions of loci, they are thought to influence phenotypes by altering gene expression. We have identified a number of autoimmune-disease associated loci for which at least two independent eQTL influence transcript abundance in whole blood and monocytes, and are testing CRISPR-Cas9 mediated site-specific gene editing strategies for scanning across the credible intervals of each eQTL, to define the most likely causal variant. The first strategy utilizes qRT-PCR to quantify target gene expression in multiple single cell clones with characterized microdeletions induced by up to 5 different guide RNAs for each eQTL. The major limitation of this approach is heterogeneity of gene expression across clones and batches. Bulk RNA sequencing of 7 single cell clones of HL60 and a derivative cell line, HL60/S4, found that PC1 (57% of the variance explained) captured differences between the two cell lines, while PC2 (10%) captured differences between single cell clones within the lines. Following chemically induced differentiation of these clones toward macrophage and neutrophil identity, the between-clone variability increased further, largely because different single cell clones grown in different batches have distinct growth and differentiation rates. Consequently, power to detect a 50% reduction in transcription with 3 biological replicates of each guide RNA is only around 30%, implying that this careful but labor-intensive approach is inefficient. Subsequently, we are now evaluating the performance of single cell RNASeq following parallel lentiviral transfection of cells with distinct guide RNAs. Simulations under the assumption of negative binomial/ Poisson distribution models suggest 80% power at experiment-wide $p < 0.05$ to detect 20% expression reduction. Results of scans across multiple eQTL credible intervals will be presented and discussed with respect to the potential of this type of approach for fine mapping causal eSNPs.

428T

A *de novo* CNV deletion of the *FOXF1* enhancer in 16q24.1 unmasks a rare noncoding SNP that likely prevented a lethal ACDMPV phenotype. P. Szafranski¹, V. Poisson², D. Bérubé², L. Oligny², J.L. Michaud², P. Stankiewicz¹. 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) CHU Ste-Justine, Montreal, Canada.

Haploinsufficiency of the *FOXF1* gene is causative for a neonatal lethal lung developmental disorder Alveolar capillary dysplasia with misalignment of pulmonary veins (ACDMPV). In addition to over 50 pathogenic point mutations, we have identified 22 CNV deletions involving *FOXF1* and 22 overlapping CNV deletions mapping upstream to *FOXF1*, defining an ~ 60 kb transcriptional enhancer region located ~ 272 kb to *FOXF1*. In all but one case, the deletions arose *de novo* on maternal chromosome 16q24.1, suggesting genomic imprinting at this locus. This enhancer region harbors fetal lung-expressed long non-coding RNA (lncRNA) genes and multiple binding sites for transcriptional regulators. Recently, we have narrowed it to a 15 kb core interval featuring binding sites for the lung-expressed transcription factors GATA3, ESR1, and YY1 (PMID: 19500772, 23034409, 27071622, 27822317). Here, we report a 6-year-old French-Canadian girl with pulmonary hypertension in whom we have identified a *de novo* ~ 2.5 Mb deletion on maternal chromosome 16q23.3q24.1 involving the entire enhancer region and leaving *FOXF1* intact. Lung biopsy showed no features of ACDMPV. DNA sequence comparison of the patient enhancer on the remaining paternal allele with the reference genome revealed 59 SNPs. Interestingly, analysis of their status in 13 patients with histopathologically-verified ACDMPV and heterozygous deletions of the entire *FOXF1* enhancer revealed the presence of G to A conversion (rs150502618) absent in the ACDMPV patients. This SNP maps to the ChIP-seq-determined overlapping binding sites for chromatin structure regulators: CTCF and an essential component of cohesin, RAD21. We propose that the SNP acts as a hypermorphic modifier of *FOXF1* expression by affecting the insulator function of CTCF or CTCF-cohesin-mediated chromatin folding. Changes in enhancer architecture and/or its interactions with the *FOXF1* promoter might increase *FOXF1* expression compensating for the deletion of its maternal copy, and preventing the development of lethal ACDMPV features in the patient.

429F

SEMP: Predicting the effects of single nucleotide polymorphisms on transcription factor binding affinity. S.S. Nishizaki¹, A.P. Boyle^{1,2}. 1) Human Genetics Department, University of Michigan, Ann Arbor, MI., United States; 2) Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI., United States.

One of the most surprising results to emerge from genome-wide association studies (GWAS) is 95% of all disease associated single nucleotide polymorphisms (SNPs) identified by this method reside in non-coding regions of the genome. Despite this finding, non-coding SNPs remain hugely understudied, due in part to the uncertain functional consequences of such mutations. However, a large proportion of these SNPs reside within regulatory regions of the genome, such as transcription factor binding sites (TFBSs). TFBSs only cover 8.1% of the genome, yet they contain 31% of GWAS SNPs. SNPs in these binding sites may alter the binding affinity of transcription factors, leading to changes in downstream gene expression, and ultimately human disease. Here, we propose a novel screening tool that estimates transcription factor (TF) binding affinity through observation of SNPs in TFBSs genome-wide. This tool utilizes publically available data from the ENCODE database to generate SNP effect matrixes (SEMs). We hypothesize that SEMs will allow us to predict disease causing SNPs in TFBSs genome wide.

430W

In vivo enhancer analysis with zebrafish. T. Kimura, I. Inoue. National Institute of Genetics, Mishima, Japan.

Genome-wide association studies have discovered tremendously large number of associations between single nucleotide polymorphisms (SNPs) and human diseases. However, there are little knowledges about functional impact of causal polymorphisms on disease onset. Indeed, we do not know the causal or functional SNP related with the disease. Informative analyses have shown that large part of the diseases associated SNPs were located on regulatory regions marked by deoxyribonuclease I hypersensitive sites, in other word, open chromatin regions. This suggested that the disease-associated SNPs perturbed transcription factor recognition sequence and altered enhancer activity. On the other hand, biological experiments have showed that enhancer activities were highly cell type specific. In order to analyze an enhancer activity of all cell types, we performed the in vivo enhancer analysis using zebrafish. As fertilized egg of zebrafish makes all types of cells, we can easily test enhancer activity throughout all types of cells by microinjection of reporter construct. We made a Tol2 transposon-based vector which harbored cfos minimal promoter and EGFP gene. We amplified regions containing glaucoma-associated SNPs in 9p21 by PCR and made reporter constructs by cloning these regions into upstream of the promoter. One of these constructs containing a rs2157719 showed EGFP signals in eye of F0 fish. Another construct containing a rs6475604 showed EGFP signal in eye and neuron. As we expected, enhancer signal was observed mostly in eye and enhancer activity differs depend on the functional SNP. Most of the enhancer studies have been performed in cell lines or primary cells, however, in vivo enhancer activity needs to be measured, particularly allele specific enhancer activity needs to be studied in vivo using zebrafish system.

431T

A pedigree-based estimate of the human germline retrotransposition rate. J. Feusier¹, W.S. Watkins¹, J. Thomas¹, A. Farrell^{1,2}, L. Baird¹, M. Leppert¹, L.B. Jorde¹. 1) Human Genetics, University of Utah, Salt Lake City, UT; 2) USTAR Center for Genetic Discovery, Salt Lake City, UT.

Non-LTR retrotransposons are active and continue to contribute new structural variation in the human genome. However, the current germline retrotransposition rates have been estimated from either phylogenetic or disease studies. We report the first direct estimate of the retrotransposition rates in a large series of 3-generation human pedigrees. We performed whole genome sequencing at 32x average coverage on blood-derived DNA from 603 CEPH individuals, comprising 33 multi-generation pedigrees. The families were joint-called using Mobile Element Locator Tool (MELT) and RUFUS to identify *de novo* Alu, LINE1, and SVA elements. We also looked for but did not detect any *de novo* HERV-K or pseudogene retrotransposition events. The three-generation families allowed us to estimate a false-negative detection rate of 5% using MELT. All candidate *de novo* mobile element insertions (MEI) were validated using PCR and Sanger sequencing. PCR validation demonstrates a false-positive rate of 0% after evaluation of candidate loci with the Integrative Genomics Viewer (IGV). Flanking SNPs were typed to determine parental chromosome of origin. *De novo* MEIs were transmitted at a 3-fold higher rate on the paternal chromosome than the maternal, with no apparent parental age bias. All *de novo* elements present in generation 2 were transmitted to generation 3, and there is one identified case of maternal germline mosaicism. The retrotransposition rate estimate for Alu elements, one in 41, is roughly half the rate estimated using phylogenetic analyses, a difference similar to that observed previously for single nucleotide variants. The SVA and LINE1 retrotransposition rates are both approximately one in 70-100 births, which is higher than phylogenetic estimates, and significantly higher for the SVA elements. We observed at least one case of endonuclease-independent LINE1 retrotransposition, which acts as a retrotransposition-mediated DNA-repair mechanism.

432F

Global protein-protein and protein-DNA interactions of DNA topoisomerases. L. Uuskula-Reimand^{1,2}, P. Samavarchi-Tehrani³, H. Hou^{1,4}, K. Abe⁵, R. Ostadsharif Memar¹, S. Akhtar Alvi¹, J. Reimand^{5,6}, A.C. Gingras^{3,4}, M.D. Wilson^{1,4}. 1) Genetics and Genome Biology Program, SickKids Research Institute, Toronto, Ontario, Canada; 2) Department of Gene Technology, Tallinn University of Technology, Tallinn, Estonia; 3) Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, ON, Canada; 4) Department of Molecular Genetics, University of Toronto, ON, Canada; 5) Ontario Institute for Cancer Research, Toronto, ON, Canada; 6) Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada.

Type II DNA topoisomerases (TOP2) regulate DNA topology by generating transient double stranded breaks during replication and transcription. Vertebrates possess two TOP2 proteins, TOP2A and TOP2B. Unlike TOP2A, which is essential for cell proliferation, ubiquitously expressed TOP2B facilitates rapid post-mitotic gene activation and functions at the later stages of development and differentiation. We previously applied proteomics (BioID), chromatin immunoprecipitation sequencing (ChIP-seq), and high-throughput chromosome conformation capture (Hi-C) approaches to identify novel proximal TOP2B protein interactions and characterize the genomic landscape of TOP2B binding. These experiments revealed close interactions between TOP2B, the cohesin complex, and CTCF, including a distinct spatially constrained DNA occupancy at the borders of topologically associating domains. However, the extent to which TOP2B's interactions are distinct from, or shared with, TOP2A and other topoisomerases such as the type I topoisomerase TOP1 is not known. Here we extend our TOP2B protein-protein interactome by performing BioID of TOP2B interacting proteins CTCF, the cohesin complex subunit RAD21, and TOP2A. For comparison, we also profiled TOP1 which is known to play a critical role in transcription and resolving DNA supercoils near CTCF sites. In addition to finding biologically relevant and distinct protein-protein interactions for each of the above proteins, several protein-protein interactions shared by TOP2B and RAD21 or TOP2B and CTCF were detected. For example, in contrast to TOP2A and TOP1, the cohesin complex proteins and loading factors showed robust protein proximal interactions with TOP2B. Our extended topoisomerase interactome reveals known and novel proteins involving of cellular processes such as transcription regulation, cell cycle, DNA repair, and chromatin organization among the interacting proteins, highlighting the interplay of gene regulatory and DNA architectural processes.

433W

Structural complexity at 3q29 locus: Contributions of genomic architecture to genomic disorders. T. Mosley¹, M. Zwick¹, J. Mulle^{1,2}. 1) Department of Human Genetics, Emory School of Medicine, Atlanta, GA, USA; 2) Department of Epidemiology, Emory Rollins School of Public Health, Atlanta, GA, USA.

Introduction: 3q29 deletion syndrome is a rare genomic disorder that occurs in approximately 1 in 30,000 births. It is caused by a recurrent, typically *de novo* deletion on the third chromosome spanning 1.6 Mb and 21 genes. It is associated with increased risks for anxiety and intellectual disability, a 30-fold increase in risk for autism, and a 40-fold increase in risk for schizophrenia. The 3q29 region is flanked by three inversely and directly oriented low-copy repeats (LCRs), suggesting the region undergoes non-homologous allelic recombination (NAHR) to form the 3q29 deletion, as well as other structural variants. Inversions, specifically, have the potential to predispose the region to structural rearrangement. Thus, individuals who carry inversions in the 3q29 region may have increased risk for forming the deletion. **Methods:** As part of the Emory 3q29 Project (genome.emory.edu/3q29), we performed Bionano® Saphyr™ whole-genome mapping on six 3q29 trios and Bionano® provided mapping data for 155 control individuals. We searched for additional structural variants, especially inversions, at the 3q29 locus and assessed the frequency of the variants in the control population versus the 3q29 population. **Results:** We observed a 1.9-Mb inversion in 2/155 (0.01%) control samples screened. Additionally, we uncovered a second 350-kb inversion adjacent to the 3q29 deletion interval in 27/157 (17%) control samples. Analyses are ongoing for the 3q29 trios; so far, we have identified the 350-kb inversion in the father of one 3q29 trio. Additional data from the collected six 3q29 trios will be presented. **Conclusion:** These data suggest an underlying structural complexity at the 3q29 locus that may influence the formation of the 3q29 deletion. There may be population-level risk factors, such as inversions or more complex structural variants that predispose the 3q29 region for structural rearrangement. **Implications:** The genomic architecture encompassing a region at risk for structural rearrangement is a critical contributing factor underlying the formation of structural variants. Ultimately, understanding the genomic architecture and risk factors contributing to structural instability in the 3q29 region can be extrapolated to other loci associated with genomic disorders. Identifying the specific factors contributing to the structural mutagenesis at the 3q29 locus (and other loci) will provide a guide for standards to identify individuals at risk for genomic rearrangement.

434T

Measuring the rate of spontaneous structural variation through whole-genome sequencing of three generation human pedigrees. J.R. Belyeu¹, R.M. Layer^{1,2}, A.R. Quinlan^{1,2,3}. 1) Department of Human Genetics, University of Utah; 2) USTAR Center for Genetic Discovery, University of Utah; 3) Department of Biomedical Informatics, University of Utah.

Accurate measurement of the rate of *de novo* structural variation (SV) in human genomes is difficult due to the historic lack of genome sequences from nuclear families, the complexity of structural variant calling, and the apparently lower mutation rate as compared to spontaneous single-nucleotide mutation. SV callsets are also dominated by false positives, which are difficult to remove without employing insufficiently specific filtering techniques, which can decrease detection of real variants. These complications have led to disagreements in previous measurements of the spontaneous SV rate, with recent estimates of the number of events per genome ranging from ~1/6 to ~1/12. To address this problem, we have measured *de novo* SV mutations in whole genome sequences from 33 large, three-generation CEPH pedigrees, comprising an average of ~9 children per F2 generation (603 individuals in total). Multi-generation pedigrees provide an ideal experimental design for directly validating putative *de novo* SVs, as true mutations in the F1 generation should be transmitted to at least one child in the F2 generation. Using this approach, we manually curated the resulting set of candidate SVs using our tool SV-plaudit, by plotting the evidence supporting the variant calls in P0, F1, and F2 generations. This allowed us to remove additional false positive variants and finalize a set of highly accurate *de novo* SV calls. We further identified candidate *de novo* SVs in the F2 generation, and again manually curated variant calls. Using the large number of F2 individuals, we will report a revised estimate of the rate of *de novo* SV mutation, as well as the prevalence of *de novo* SVs arising from germline mosaicism. Finally, we used results from both sets of curated SV calls to create a set of reproducible filters for the removal of false-positive SV calls from additional SV datasets. We applied these filters to >3000 family quartets from the Simons Foundation Autism Research Initiative (SFARI) to explore the *de novo* SV rate in a larger cohort, assess how the rate changes with parental age, and compare the rate in affected vs unaffected siblings. We anticipate that a better understanding of the *de novo* SV rate will improve our understanding the mutation of the human genome and that these filters will become a commonly used tool to identify true *de novo* SVs in many family datasets.

435F

Comprehensive structural variant identification and genomic characterization using 720 deeply sequenced whole genomes. D. Jakubosky^{1,2}, M. D'Antonio³, C. DeBoever⁴, M.J. Bonder⁵, C. Smail⁶, W.W. Greenwald⁴, H. Matsui⁶, A. D'Antonio-Chronowska³, S.B. Montgomery^{7,8}, O. Stegle^{9,10}, E.N. Smith¹¹, K.A. Frazer^{9,11}, *i2QTL Consortium*. 1) Biomedical Sciences Graduate Program, University of California, San Diego, La Jolla, CA 92093, USA; 2) Department of Biomedical Informatics, University of California, San Diego, La Jolla, CA 92093, USA; 3) Institute of Genomic Medicine, University of California San Diego, 9500 Gilman Dr, La Jolla, CA 92093, USA; 4) Bioinformatics and Systems Biology Graduate Program, University of California San Diego, La Jolla, CA 92093, USA; 5) European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Cambridge, UK; 6) Departments of Pathology and Biomedical Data Science, Stanford University School of Medicine, Stanford, CA 94305, USA; 7) Department of Pathology, Stanford University, Stanford, CA 94305, USA; 8) Department of Genetics, Stanford University, Stanford, CA 94305, USA; 9) European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, UK; 10) European Molecular Biology Laboratory, Genome Biology Unit, Heidelberg, Germany; 11) Department of Pediatrics, University of California San Diego, La Jolla, CA 92093, USA.

Structural variants (SVs) and short tandem repeats (STRs) encompass at least 11% of the human genome and are important sources of genetic and phenotypic diversity; however, their functional impact remains poorly understood. Previous efforts have focused on specific non-SNV subtypes, contained few samples, or utilized low coverage sequencing data. To thoroughly characterize the functional landscape of non-single nucleotide variants (non-SNV), high-resolution mapping of all known SV and STR classes is needed. Here we assembled a set of 720 blood, fibroblast, and induced pluripotent stem cell samples (478 individuals) with deep whole genome sequencing (WGS) data (mean 42x) and: 1) profiled SV and STR variation using 5 algorithms, 2) applied a systematic filtering strategy to enable the identification of variants with highly reproducible genotype calls, and 3) characterized the genomic properties of different variant classes, thereby producing one of the most complete and well characterized maps of structural genetic variation to date. We obtained WGS data from participants of the iPSCORE and HipSci studies as part of the i2QTL Consortium and called SVs and STRs using HipSTR, Genome STRIP, MELT, LUMPY, and CNVnator. To obtain high quality variants, we filtered to achieve an overall replication rate of >90% for calls in genetically duplicated samples (25 MZ twin pairs and 152 fibroblasts with matched iPSC). Filtered calls included over 200,000 STR variable regions and 66,471 SVs, of which ~15.4% were mobile element insertions (MEI), 4.3% reference MEI, 29.5% deletions (DEL), 5.5% duplications (DUP), 3.3% multi allelic copy number variants (mCNV), and 41.3% unspecified breakends (BND). These calls were highly concordant (FDR 5%) with data from orthogonal genotyping arrays. Moreover, we performed enrichment analyses of variant classes in functional regions of the genome and observed that DUPs and mCNVs were more likely to overlap genes than DELs and MEIs, and that DELs were more likely to occur in intergenic regions. Additionally, DELs, DUPs, and mCNVs occurred less often in active chromatin compared to inactive chromatin. Finally, rare variants of all classes were more likely to occur in conserved regions (PhyloP) than common variants. Overall, this study establishes the largest set of high quality SV and STR genotype calls in samples with paired gene expression data that can be used to interrogate the functional impact of non-SNV variation in the human genome.

436W

The impact of structural variation on human gene expression. A.J. Scott¹, C. Chiang¹, L. Ganel¹, A. Wilfert², D.F. Conrad^{2,3}, I.M. Hall^{1,4}. 1) McDonnell Genome Institute, Washington University School of Medicine, St. Louis, MO; 2) Department of Genetics, Washington University School of Medicine, St. Louis, MO; 3) Department of Pathology & Immunology, Washington University School of Medicine, St. Louis, MO; 4) Department of Medicine, Washington University School of Medicine, St. Louis, MO.

Structural variants (SVs) are an important but poorly understood source of human genome variation. The Genotype-Tissue Expression (GTEx) project provides an unprecedented opportunity to address the contribution of SV to traits, disease, and gene regulation due to the generation of deep (50X) WGS data and multi-tissue RNA-seq data. Here we present results from ongoing analysis of 613 individuals from the GTEx cohort. We used comprehensive detection methods to map 64,568 high confidence SVs, including copy number variants, inversions, reference and non-reference mobile element insertions, and novel adjacencies of indeterminate type. We performed *cis* expression quantitative trait loci (eQTLs) analyses in 48 tissues via joint analysis of common (>1% MAF) SVs, SNVs, and short indels. Based on haplotype structure and heritability partitioning, we estimate that 2.7-5.3% of eQTLs are caused by common SVs, a relatively modest contribution that nonetheless represents a 17- to 33-fold enrichment relative to their abundance in the genome. We assessed the contribution of different classes of regulatory elements at SVs predicted to causally alter gene expression and observed an enrichment in features such as enhancers and promoters. By evaluating linkage disequilibrium between SVs, SNVs, and indels, we nominate 104 SVs as plausible causal variants at published genome-wide association study (GWAS) loci from UK Biobank, including smoking status, sleep duration, and basal cell carcinoma. In contrast to the modest contribution of common SVs, we observe a remarkable abundance of rare SVs associated with highly aberrant expression of nearby genes. These predicted causal rare SVs are nominally enriched in several regulatory elements including genic enhancers and CTCF binding sites, which may be useful for improving rare variant interpretation methods. Our results indicate that both common and rare SVs play an important role in causing gene expression differences in the human population and that comprehensive SV analyses will improve the power of WGS-based association studies.

437T

Germline copy number variants are associated with methylation variation in the genome. S.R. Setlur¹, X. Shi², J.Y. Chen¹, S. Radhakrishnan¹, J. Wen², B.A. Lam², R.E. Mills³, B.E. Stranger⁴, C. Lee⁵. 1) Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 2) University of North Carolina at Charlotte; 3) University of Michigan; 4) University of Chicago; 5) The Jackson Laboratory for Genomic Medicine.

Large consortium efforts like the International HapMap Project and the 1000 Genomes Projects, have generated a detailed catalog of genetic variants in human populations. The two major classes of genetic variants include single nucleotide polymorphism (SNPs) and copy number variation (CNVs). The effect of SNPs on cellular phenotypes and disease, has been shown by various studies. However, the effects of CNVs, the more recently recognized class of germline genetic variants, on gene regulation and phenotype is not well-elucidated. CNVs can influence gene expression by overlapping either genes or non-coding regulatory regions in the genome. In addition to being genetically influenced, gene expression can also be epigenetically regulated. One such epigenetic regulatory mechanism is DNA methylation. Similar to CNVs, DNA methylation has been reported to vary between individuals and among different tissues within an individual. Studies have shown that monozygotic twins exhibit inter-individual variability in both DNA copy number and methylation. Despite these similarities, we know very little about the relationship between germline CNVs and variation in DNA methylation. We investigated the association between germline CNVs and DNA methylation by genome-wide methylation quantitative trait locus (mQTL) analysis. Using CNV genotypes and DNA methylation data from individuals profiled in the 1000 Genomes and the HapMap projects, we identified 851 genome-wide CNV-CpG associations (CNV-mQTLs, p -value < 0.01 , permutation test). We found instances where a single CNV was associated with CpG methylation of multiple genes and vice-versa. CNV-associated methylation changes were also correlated with gene expression. We found that CNV-mQTLs were enriched for regulatory regions as defined by the ENCODE consortium including transcription factor binding sites. Analysis of Hi-C data, to assess for long-range physical interactions between CNVs and the associated CpGs, revealed that the associated CNV-CpG pairs were enriched for Hi-C interactions suggesting a potential functional association. Finally, we found that some of the CNV-mQTLs and/or associated genes were previously reported to confer disease risk by genome-wide association studies. In summary, we demonstrate for the first time, that germline CNVs are associated with variable DNA methylation in the human genome. These findings suggest a distinct universal mechanism by which structural variation may affect phenotype.

438F

The effects of structural variation on 3D chromatin structure. O. Shanta¹, A. Noor², D. Gorkin^{3,4}, Y. Qiu⁴, B. Ren^{3,4}, J. Sebat². 1) Department of Electrical and Computer Engineering, University of California San Diego, La Jolla, CA; 2) Beyster Center for Genomics of Psychiatric Diseases, Department of Psychiatry, University of California San Diego, La Jolla, CA; 3) Department of Cellular and Molecular Medicine, University of California San Diego, School of Medicine, La Jolla, CA; 4) Ludwig Institute for Cancer Research, La Jolla, CA.

Human chromosomes exhibit three-dimensional spatial organization whereby regions of the genome form highly self-interacting regions termed Topological Associating Domains (TADs). TADs are separated by discrete boundaries in which the chromatin interactions cease abruptly. We hypothesize that structural variants (SVs) in the human genome contribute to heritable variation in chromatin structure. We investigated the effects of SVs on 3D chromatin organization by chromosome conformation capture sequencing (Hi-C) of lymphoblastoid cell lines from 19 subjects enrolled in the 1000 genomes project. The correlation of deletions with the frequency of long-range contacts was tested by principal component regression. We demonstrate that quantitative variation in TADs is strongly influenced by confounding variables including local copy number variation and genome-wide variation in Hi-C data, which can be controlled for statistically. After eliminating these confounders, significant effects of deletions on TAD structure were evident and such effects were strongest for deletions >10 kb in length and deletions that occur in the boundary region between TADs. Deletions result in the disruption of interactions in the flanking regions within ~ 200 kb and an increase in contacts that span the deletion. These results indicate that deletion of the boundary regions between TADs promotes interactions between discrete chromatin domains. However, the overall structure of domains remains largely intact, suggesting that regulatory elements within the boundaries are not required to establish TADs.

439W

Discovery and genotyping of new short and variable number of tandem repeats in the human genome. A. Sulovari¹, P.A. Audano¹, M.J.P. Chaisson¹, C. Lee^{2,3}, J.O. Korbel⁴, E.E. Eichler^{1,5}, *The Human Genome Structural Variation Consortium (HG SVC)*. 1) Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA 98195, USA; 2) The Jackson Laboratory for Genomic Medicine, Farmington, CT 06032, USA; 3) Department of Graduate Studies – Life Sciences, Ewha Womans University, Ewhayeodae-gil, Seodaemun-gu, Seoul, South Korea; 4) European Molecular Biology Laboratory, Genome Biology Unit, Heidelberg, Germany; 5) Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA.

Short and variable number of tandem repeats (STRs and VNTRs) are an important source of disease-causing variation, but they have been problematic to resolve in the reference genome and to genotype with short-read technology. We recently sequenced three human parent–child trios of Yoruban, Chinese and Puerto Rican descent using a multiplatform approach involving short- and long-read sequencing as well as linking technologies. Using Phased-SV, we systematically detected and sequence-characterized STRs and VNTRs of variable length with a particular emphasis on those missing from the human reference genome. We sequence-resolved 21,442 tandem repeats representing 2,767 STRs (1–8 bp) and 3,319 VNTRs (≥ 9 bp) per haplotype, all missing from the human reference genome, GRCh38. 1,300 insertions (1.5 Mbp) were shared by all six haplotypes suggesting that the reference human genome carries a smaller minor allele or is in error at these locations. 7,087 loci overlapped a gene region and 458 intersected an exon. These include large repeats such as a 56-copy hexamer located 4 kbp upstream of *RNASE10*, a 6-copy 5,608-mer overlapping with *LPA*, 2–3 kbp protein-encoding expansions in mucin repeats, and expanded trinucleotide repeats in a variety of disease-associated genes. 45% of VNTRs and 18% of STRs are located in the last 5 Mbp of the telomeric ends representing an enrichment of 5.5- and 2-fold for VNTRs and STRs, respectively. This distribution correlates with biases in chromosomal double-strand breaks (average $r = 0.7$, $p < 2 \times 10^{-16}$), suggesting a mechanistic relationship. We developed a VNTR genotyping approach for short-read data using these alternate alleles. To test its accuracy, we used high-coverage Illumina genome data (150 bp paired-end reads) from the same trio individuals and focused on the top 3% most polymorphic VNTRs across the six haplotypes. We show that most of these VNTRs can be genotyped with $\geq 80\%$ accuracy, which allows expansions and contractions to be readily identified, while approximately 35% of VNTRs (average size of 5.6 kbp) can be genotyped with an accuracy $\geq 99\%$. We are now applying this strategy to $>2,000$ previously uncharacterized STRs and VNTRs in over 10,000 autism genomes and will present our findings for events that are hyperexpanded in cases compared to unaffected siblings. The strategy we have developed can be more broadly applied to discover expanded STRs and VNTRs for other Mendelian and common complex diseases.

440T

De novo human genome assemblies reveal novel sequences in diverse populations. K.H.Y. Wong¹, M. Levy-Sakin¹, P. Kwok^{1,2,3}. 1) Cardiovascular Research Institute, University of California, San Francisco, San Francisco, CA; 2) Institute for Human Genetics, University of California, San Francisco, San Francisco, CA; 3) Department of Dermatology, University of California, San Francisco, San Francisco, CA.

The human reference genome is used extensively in modern biological research. However, a single consensus genome representation is inadequate to provide a universal reference structure because it is a haplotype among many in the human population. Using 10x Genomics “Linked-Read” technology, we performed whole genome sequencing and *de novo* assembly on 17 individuals across five populations: African, American, East Asian, European, and South Asian. Using an alignment- and *de novo* assembly-based approach, we identified 1,842 breakpoint-resolved Non-reference Unique Insertions (NUIs) that, in aggregate, added up to 2.1Mb of new genomic content not currently represented in the human reference genome. These NUIs displayed an underlying population structure, as principal component analysis could faithfully stratify these individuals based on their geographic origins. Furthermore, 64% of these NUIs aligned to non-human primate references, suggesting that they are ancestral sequences. We found that at least 14% of these NUIs likely originated from Alu-mediated recombination. Additionally, 37% of the NUIs aligned to either entries in the expressed sequence tag database or reads from RNA-sequencing experiments, indicating that they are in the human transcriptome. Our results demonstrate that as the genomic variant catalogs continue to expand, the human reference genome must include a more comprehensive set of alternative haplotypes to depict the complete spectrum of genetic diversity across populations.

441F

Patterns and mechanisms defining tissue specificity of human regulatory variation. S.E. Castel^{1,2}, F. Aguet^{3,8}, E. Flynn^{1,2,8}, S. Kim-Hellmuth^{1,2,8}, P. Mohammadi^{4,8}, Y. Park^{5,8}, B. Strober^{6,8}, V. Wucher^{7,8}, T. Lappalainen^{1,2}, GTEx Consortium. 1) New York Genome Center, New York, NY; 2) Columbia University, New York, NY; 3) Broad Institute, Cambridge, MA; 4) The Scripps Research Institute, La Jolla, CA; 5) University of Pennsylvania, Philadelphia, PA; 6) Johns Hopkins University, Baltimore, MD; 7) Center for Genomic Regulation, Barcelona, Spain; 8) equal contribution, listed in alphabetical order.

Quantitative trait loci (QTL) studies that map regulatory variation can bridge the gap between genetic associations to disease and molecular phenotypes. In this work, we studied the tissue specificity of regulatory variation in the v8 release of the Genotype Tissue Expression (GTEx) project, consisting of 838 individuals with 15,201 samples across 49 human tissues, to shed light on its biological mechanisms and interpret disease associations. We quantified tissue specificity of regulatory variation by cross tissue meta-analysis of *cis* and *trans* expression QTLs, as well as splicing QTLs. We also estimated *cis* regulatory effect sharing using individuals with allelic expression data from multiple tissues. Comparison of tissue clustering by shared regulatory effects, gene expression, and inferred cell type composition indicated shared cellular mechanisms but also complex relationships between these factors that define tissue identity. Leveraging on estimates of gene expression sharing between GTEx tissues and other data sets, we demonstrated the use of GTEx tissues as a proxy for diverse tissues. To study the mechanisms underlying eQTL tissue specificity we modeled it using genomic annotations and chromatin states from relevant ENCODE cell types while controlling for confounding effects. This model could predict whether an eQTL was shared between a discovery and replication tissue with an AUC of 0.73. Post-transcriptional effects were more likely to be shared than transcriptional effects, with stop gain ($p=4.74e-31$) and splice variants ($p=1.89e-19$) having the strongest sharing. On the transcriptional level, when compared to eQTLs in unannotated regions, those in promoters had a higher probability of sharing ($p=1.49e-46$), while those in enhancers had a lower probability ($p=6.55e-224$). Furthermore, active chromatin states were associated with higher tissue sharing ($p = 0.011$) while inactive states were associated with lower sharing ($p=9.98e-110$). Finally, we studied two main dimensions of tissue-specificity: eQTL strength and eGene expression level across tissues. These were correlated in 3,477 out of 26,499 tested eQTLs (FDR<5%), and in lower expressed genes eQTL effect size tended to increase with expression level, while highly expressed genes showed the opposite pattern ($p=6.58e-107$). Altogether, our findings illustrate the power of studies that span broad tissues for determining the tissue specificity of regulatory variation and elucidating its mechanism.

442W

New methods for discovery and interpretation of allelic diversity in human genomes. R.S. Fulton^{1,2}, T. Graves-Lindsay^{1,2}, A. Regier^{1,3}, N. Das¹, C. Tomlinson¹, E. Belter¹, N. Stitzel^{1,2,3}, I. Hall^{1,2,3}. 1) McDonnell Genome Institute at Washington University School of Medicine, St. Louis, MO; 2) Department of Genetics, Washington University School of Medicine, St. Louis, MO; 3) Department of Medicine, Washington University School of Medicine, St. Louis, MO.

Interpretation of genomic sequencing data is currently reliant upon the human genome reference sequence. The public reference is a linear representation of a mosaic human genome with associated alternate alleles representing regions of significant allelic diversity. Although adequate for many regions of the genome, allelic diversity presents challenges where genomes being interrogated differ from the reference, or that variation lies in regions recalcitrant to short read mapping. Additionally, the common tool chain utilized in most genomic studies does not take advantage of alternate alleles and thus does not exploit many of these alternate representations. In an effort to unveil additional allelic diversity and to develop methods to discover such diversity through reference-free approaches, we at the McDonnell Genome Institute have initiated a project as part of our NHGRI Center for Common Disease Genomics (CCDG) program to generate and characterize >500 human genome assemblies from individuals of diverse ancestry and ethnicity. We are currently generating haplotype-resolved whole genome assemblies from ~60X long linked-read WGS data from the 10X Genomics platform using the Supernova assembly algorithm. A key aspect of this project is our effort to develop optimized laboratory processes that enable the highest quality assemblies possible from these data. We are constructing a pan-genome representation of genetic variation identified in these assemblies, using highly sensitive reference-free assembly comparison methods. We will then use variant graph alignment methods to genotype these variants in a larger cohort comprising tens of thousands of short-read WGS datasets generated for CCDG trait mapping studies. This reference-free trait association experiment will allow us identify and genotype variants that would otherwise be missed by current reference-based analysis methods due to biases caused by regional fluctuations in ancestry and allelic diversity. This presentation highlights early findings and progress towards the ultimate goal of carrying out comprehensive and unbiased WGS-based human genetics studies that improve our understanding of disease risk in diverse human populations.

443T

Systematic dissection of transcription factor co-binding as a positive predictor of regulatory activity. A.A. Hardigan^{1,2}, R.C. Ramaker^{1,2}, E.C. Partridge², S.B. Chhetri^{2,3}, K.M. Newberry², S. Goh⁴, B. Wold⁴, E. Mendenhall^{2,3}, S.J. Cooper², R.M. Myers². 1) University of Alabama at Birmingham, Birmingham, AL; 2) HudsonAlpha Institute for Biotechnology, Huntsville, AL; 3) University of Alabama at Huntsville, Huntsville, AL; 4) California Institute of Technology, Pasadena, CA.

Transcription factors (TFs) are proteins that regulate gene expression through binding to regulatory DNA regions in the genome such as enhancers and promoters. Multiple TFs often interact at these regulatory regions and determining the nature of this cooperative binding on gene expression and genome regulatory architecture is critical to better understand TF biology. Using chromatin immunoprecipitation sequencing (ChIP-seq), our group recently determined the genome-wide occupancy of 208 TFs in the HepG2 cell line. We found a striking range of co-occupancy amongst TFs in this data set, with ~80% of loci containing <10 TFs, while 10% of loci contained >50 TFs. Using ENCODE ChIP-seq data, we also identified these highly bound loci, or “high occupancy target” (HOT) regions, in the K562 and GM12878 cell lines and identified shared and cell-type specific HOT regions. We found that the number of TFs bound at a given locus correlated significantly with gene expression levels of the nearest gene, and IDEAs promoter and strong enhancer annotations. Using DNase hypersensitive site sequencing (DNase-seq) data from ENCODE and TF footprinting, we replicated our analysis of TF occupancy and HOT regions and found a similar correlation of the number of TF footprints to gene regulatory metrics. However, only a minority of TFs with ChIP-seq binding sites in a given loci also had a DNase-seq TF footprint, indicating that the majority of factors bound at a site do so indirectly. To further dissect the importance of individual motifs and sequence variation in these HOT regions, we performed a massively parallel reporter assay using the self-transcribing active regulatory region (STARR-seq) method. We designed a custom library of 245 unique 130-bp test regions with tiled single nucleotide variants (SNVs) and 5-mer mutations across the length of the elements. We identified critical nucleotides governing TF motif and overall element activity, with a significantly higher proportion of mutations in DHS TF footprints, and found a significant correlation of between LS-GKM predicted motif disruptions and differential activity of mutant elements. Finally, we found a significant negative correlation between the number of motifs in an element and its overall vulnerability to mutation. This work helps support the important regulatory role of HOT regions, and furthermore provides a more complete understanding of the regulatory dynamics of TF occupancy and cooperation in the genome.

444F

Latest improvements in the human genome reference assembly (GRCh38). T. Rezaie¹, T. Graves-Lindsay², K. Howe³, P. Flicek⁴, V.A. Schneider¹ for the Genome Reference Consortium. 1) National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD; 2) McDonnell Genome Institute at Washington University, St. Louis, MO; 3) Wellcome Sanger Institute, Hinxton, Cambridge, UK; 4) European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, Cambridge, UK.

The Genome Reference Consortium (GRC) provides users with updates to the human reference genome assembly. In this era of cutting-edge sequencing and assembly technologies, big data, and diagnostic genome sequencing, the reference genome continues to serve as a foundation for many types of research. It provides the coordinate system for genomic-based analyses and gene annotations, supports the identification of disease-associated variants and offers representation for the diversity of the human population. Our ongoing curation of this reference, the highest quality model of the human genome, provides assembly corrections and new variants, improving the representation of highly complex and clinically important genomic regions. Since the 2013 release of GRCh38, the latest coordinate-changing update to the human reference, the GRC has provided a total of 12 public assembly updates in the form of non-coordinate changing patch releases. As of GRCh38.p12 (GCA_000001405.27), the reference includes 70 fix and 70 novel patch scaffolds, which represent chromosome path changes and alternate representations of sequences, respectively. These patches, which are stand-alone accessioned scaffolds, are given chromosome context via alignments that are available for download with the rest of the assembly. The current suite of patches covers 48 Mb (~1.5% of the total chromosome length), of which 2.23 Mb is novel sequence, and includes coding regions. We will describe our latest analyses to identify and update rare alleles (MAF<5%) affecting coding sequences in the reference. We will present examples of the use of sequence from highly contiguous WGS assemblies and optical maps data from different populations to close remaining GRCh38 gaps and add additional diversity to the reference. We will present GRC policies for data contributions to the reference and our plans for future assembly updates. We will show examples of recent updates, including those not yet released, that close assembly gaps, improve gene representations at clinically important loci, add genomic diversity and discuss their usage in basic and clinical analyses. The GRC is committed to making its curation efforts publicly accessible. Information about assembly issues and updates are available at <https://www.genomereference.org> and in genome browsers via the GRC track hub. We will review these resources available at the GRC website and mechanisms for contacting the GRC with assembly questions and curation requests.

445W

A comprehensive characterization of mitochondrial DNA variants in 1,256 healthy Northern Virginian families. *W. Zhu¹, W. Wong¹, X. Liu¹, J. Deeken^{1,2}, J. Niederhuber^{1,3}.* 1) Bioinformatics, Inova Translational Medicine Institute, Fairfax, VA; 2) Virginia Commonwealth University School of Medicine, Richmond, VA; 3) Johns Hopkins University School of Medicine, Baltimore, MD.

Mitochondria play critical roles in cellular energy production and metabolism, and fully characterization of mitochondrial variants is essential for any large genome study. In this study, we whole-genome sequenced (~40X) 3,768 healthy subjects from 1,256 Northern Virginian trios (newborns and their parents), who participated in the "Childhood Longitudinal Study" at the Inova Translational Medicine Institute. Over 30% (N=810) of the total 2,512 parents from those families were born in Central America (N=652) or South America (N=158). We called mitochondrial variants using the bioinformatic tools MTool-Box and freebayes, and assigned haplogroup via the program HaploGrep2 (<http://haplogrep.uibk.ac.at>) with the mtDNA tree Build 17. Altogether, we found 17,456 mitochondrial variants in the 2,512 parents (an average of 6.9 per individual), including 16,468 homoplasmies and 988 heteroplasmies, from various haplogroups leading by A2 (20.1%, N=504) and B2 (6.8%, N=172). 88.2% of the identified homoplasmies are haplogroup markers, while only 26.3% of the heteroplasmies are haplogroup markers. Some homoplasmies like 8896A and 4227G have not been fixed in any known haplogroups, but were uniquely found in the Native American haplogroup A2w. The result indicated the emergence of novel subhaplogroups in the A2w lineage. Most of heteroplasmies found in the mothers were also inherited by their children but with fluctuated allele frequencies, suggesting the genetic bottleneck during the transmission. It is believed that there is co-evolution between mitochondrial DNA and nuclear DNA, and the disruption of such mitonuclear crosstalk may lead to the mitochondrial diseases. The disease-associated mitochondrial DNA mutations are well curated in the MitoMap database (<http://mitomap.org>). We found 3.4% of the homoplasmies were mitochondria disease related according to the MitoMap database, which is slightly lower than 5.4% in the heteroplasmies. Interestingly, some of the disease-associated variants are also haplogroup markers, suggesting that the pathogenic effects of those mitochondrial variants might be conditioned upon the nuclear DNA context. Further study may warrant enhancing our understanding in the association between the mtDNA variants and the mitochondrial diseases.

446T

Shared patterns of variation between human individuals and cells:

Somatic and germline mutation rates usefully predict each other. *W. Meyerson^{1,2}, M. Gerstein^{1,3,4}.* 1) Program in Computational Biology & Bioinformatics, Yale University, New Haven, CT; 2) Yale School of Medicine, Yale University, New Haven, CT; 3) Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT; 4) Department of Computer Science, Yale University, New Haven, CT.

Human genetic variants occur in two settings: germline variants between individuals in a population and somatic variants between cells in our bodies. The former explains much human phenotypic diversity and disease risk; the latter drives cancer and affects aging and autoimmunity. To identify pathogenic mutations in either setting, we must understand the normal range of human variation as well as sequencing artifacts. While much is known about neutral somatic and germline variants, few studies have compared germline and somatic variants with each other. Such comparative analysis offers unique insight into ubiquitous mutational processes and sequencing artifacts and enables the principled combination of somatic and germline variants to yield a much greater effective sample size for mutation modeling, which is key for discovering the long tail of pathogenic variants of small effect at the frontier of genomics. The completion of two vast, uniformly processed whole-genome sequencing databases - the genome aggregation database (gnomAD) of 15,000 healthy germline samples and the PanCancer Analysis of Whole Genomes (PCAWG) of 2,500 somatic and germline samples from cancer patients - permits the interrelation of somatic and germline variant patterns at unprecedented scale. After excluding sites that are non-uniquely mappable, poorly covered in either database, or sites of common human variation, both gnomAD and PCAWG germline had a single nucleotide variant (SNV)-level overlap with PCAWG soma 2-3 times higher than expected by chance, higher still in cancers with a large prevalence of clock-like Signature 1. A similar overlap with soma was found using *de novo* germline variants from denovo-db, indicating that these results do not derive from the misclassification of true PCAWG germline variants as somatic variants. Mutation rates by heptamer showed high concordance between soma and germline, sometimes with surprising consistency. Megabase-level mutation rates also showed a high somatic-germline correlation, especially in low meiotic recombination regions. Most importantly, when predicting germline mutation rates by region from epigenetic features, somatic mutation rates enhanced accuracy and *vice versa*. These findings demonstrate a large shared component in somatic and germline SNV call-sets and highlight the utility of somatic and germline variants for modeling each other, a critical task in the search for variants that matter in human disease.

447F

Nuclear genetic regulation of the mitochondrial transcriptome across multiple tissue types. A. Hodgkinson, A. Ali. Department of Medical and Molecular Genetics, School of Basic and Medical Biosciences, King's College London, London, United Kingdom.

Mitochondria play an important role in fundamental cellular processes and disease, however little is known about how the transcriptional regime of the mitochondrial genome varies across individuals and tissues. By analyzing >13,000 RNA sequencing datasets across 36 tissue/cell types, we find considerable variation in mitochondrial gene expression along the mitochondrial transcriptome, across tissue types and between individuals, highlighting the importance of cell-type specific and post-transcriptional processes in shaping cellular mitochondrial RNA levels across tissues. Using whole-genome genetic data we identify 64 nuclear loci associated with the expression levels of genes encoded in the mitochondrial genome, including missense variants within genes involved in mitochondrial function, implicating genetic mechanisms that act in *trans* across the two genomes. We replicate ~21% of associations with independent tissue-matched datasets and find genetic variants linked to these nuclear loci that have been associated with cardio-metabolic phenotypes and Vitiligo, supporting a role for variable mitochondrial gene expression in complex disease.

1824W

DNA methylation age in type 1 diabetes. *D. Roshandel¹, Z. Chen², B.H. Braffett³, J.M. Lachin³, A.J. Canty⁴, S.B. Bull^{5,6}, R. Natarajan², A.D. Paterson^{1,6}, DCCT/EDIC Research Group.* 1) Genetics and Genome Biology Program, The Hospital for Sick Children, Toronto, ON, Canada; 2) Department of Diabetes Complications and Metabolism, Beckman Research Institute of City of Hope, Duarte, CA, USA; 3) The Biostatistics Center, The George Washington University, Rockville, MD, USA; 4) Department of Mathematics and Statistics, McMaster University, Hamilton, ON, Canada; 5) Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, ON, Canada; 6) Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada.

The levels of DNA methylation (DNAm) at some sites across the genome show age-related changes and have been used to accurately predict age in general population across a broad spectrum of human tissues. We investigated DNAm age and the factors affecting it in subjects with type 1 diabetes (T1D). T1D subjects were randomly assigned to intensive or conventional treatment groups in the Diabetes Control and Complications Trial (DCCT) in 1983–9. At DCCT close-out in 1993, intensive therapy was recommended for all and participants have since been followed through Epidemiology of Diabetes Interventions and Complications (EDIC). DNAm was measured in 63 subjects using Infinium HumanMethylation450K BeadChip in whole blood (WB) collected in 1992–3 and monocytes (MC) collected in 2009–10 (3 missing). These included 32 cases from conventional group with mean HbA1c>9.1% during DCCT and significant progression of retinopathy/nephropathy from DCCT closeout to EDIC year 10, and 31 controls from intensive group with mean HbA1c<7.3% and no progression of retinopathy/nephropathy. DNAm age and age acceleration (AA), the difference between DNAm and chronological age, were estimated using the Horvath method. The associations of sex, age at T1D diagnosis, T1D duration, and treatment group with AA were tested by multivariable linear regression. WB DNAm age was on average greater than chronological age in both cases and controls ($p<0.0001$): AA=6.77 (range -0.09–16.50) and 5.14 yrs (range -0.27–17.08), respectively. Similarly, MC DNAm age was 4.31 (range -3.75–15.08) and 1.62 yrs (range -8.69–9.48) greater than chronological age in cases ($p<0.0001$) and controls ($p=0.068$), respectively; but the difference was only significant in cases. WB and MC AA were not significantly different between cases and controls ($p=0.14$ and 0.08 , respectively). T1D duration from DCCT eligibility to DNAm measurement was associated with WB AA (β (SE)=0.07(0.03), $p=0.017$). Age at T1D diagnosis was associated with MC AA (β (SE)=-0.23(0.08), $p=0.008$). WB and MC AA were highly correlated within individuals (spearman $\rho=0.62$, $p<0.0001$). In conclusion, DNAm age was on average higher than chronological age in T1D. Longer T1D duration was associated with higher WB AA. Older age at T1D diagnosis was associated with lower MC AA. Caution in generalizing the results is required as the subjects were highly selected in terms of glycemic control and diabetic complications; and conventional therapy is no longer recommended.

1825T

Allelic differences in adipocyte chromatin accessibility identify putative functional variants and mechanisms at loci associated with adipose tissue gene expression. *K.W. Currin¹, H.J. Perrin¹, C.K. Raulerson¹, A. Ko², J. Luo³, M. Wabitsch⁴, P. Pajukanta^{2,5}, M. Laakso⁶, K.L. Mohlke¹.* 1) Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; 2) Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA; 3) Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; 4) Department of Pediatrics and Adolescents Medicine, Division of Pediatric Endocrinology and Diabetes, University of Ulm, Ulm, Germany; 5) Molecular Biology Institute at UCLA, Los Angeles, CA, USA; 6) Department of Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland.

To identify functional variants and mechanisms at cis-expression quantitative trait loci (eQTLs), we identified sites of allelic imbalance (AI) in chromatin accessibility in adipocytes derived from the Simpson Golabi-Behmel Syndrome (SGBS) cell strain. We genotyped SGBS cells using the Infinium Multi-Ethnic Global array, imputed additional genotypes (1000G phase 3), excluded non-diploid regions, and analyzed ~46M genotypes with posterior probability > 0.9. We profiled SGBS chromatin accessibility after day 14 of differentiation using ATAC-seq (3 replicates), resulting in 138M–156M non-duplicated nuclear alignments (mapq>=20). We removed ATAC-seq reads showing allelic mapping biases using WASP, resulting in 3.08M–3.45M reads overlapping heterozygous SNPs. To identify AI, we performed a beta-binomial test separately in each replicate using 13,127 SNPs with at least 20 total overlapping reads and at least 1 read per allele in all replicates. After combining AI p-values across replicates (Stouffer) and correcting for multiple testing (Benjamini-Hochberg), 623 SNPs showed significant AI (FDR<5%). All SNPs showed the same direction of imbalance in all replicates, and 497 (80%) were nominally significant ($p<0.05$) in at least 2 replicates. The majority of AI SNPs (440, 70%) were found in promoter or enhancer chromatin states in Roadmap tissue-derived adipocytes. Using adipose tissue cis-eQTLs from the METabolic Syndrome in Men study ($n=434$), we identified 85 AI SNPs that were an eQTL for 1 or more of 79 genes (FDR<5%) and were in high linkage disequilibrium ($r^2\geq 0.8$) with the lead eQTL variant for the same gene. 14 genes had lead eQTL variants in high LD with multiple AI SNPs, resulting in multiple candidate functional variants at these loci. For 69 AI SNPs that were eQTLs for only 1 gene, the allele with increased chromatin accessibility was most often associated with increased gene expression ($n=51$, 74%), suggesting AI SNPs mainly occur in regions of allelic binding of transcriptional activators; the remaining loci may occur in regions of allelic binding of transcriptional repressors or insulators. In most of the 16 cases when an AI SNP was an eQTL for 2 genes, the expression of both genes changed in the same direction ($n=12$, 75%). Our results suggest mapping chromatin accessibility AI in cultured cells can identify genetic variants that influence transcriptional regulatory elements and gene expression in tissue and help uncover how these variants function.

1826F

Type 2 diabetes epigenome data repository and resource. Y. Sun, P. Kudtarkar, S. Morabito, K. Gaulton, AMP-T2D consortium. Department of Pediatrics, University of California San Diego, School of Medicine, La Jolla, CA.

Genetic risk of type 2 diabetes (T2D) maps largely to the non-coding genome and affects gene regulation. The availability of functional genome and epigenome data from diabetes-relevant tissues has increased dramatically with the rapid progression of high-throughput molecular assay and sequencing technologies. However, this leaves a major challenge in connecting the epigenome with T2D genetic data to enable mechanistic and therapeutic discovery. Here we describe the Diabetes Epigenome Atlas (www.t2depigenome.org) project which is a web service and database based on open source software developed by the ENCODE DCC. The goal of this project is to collect and centrally organize epigenome data from diabetes and complications-relevant tissues, provide a resource for researchers to query, obtain and analyze these data, and serve as a collaborative space for researchers to share and analyze proprietary data. The database currently hosts more than 1,050 published experiments of ChIP-seq, ATAC-seq, RNA-seq, WGBS, and scRNA-seq from pancreatic, adipose, liver, muscle, and brain tissue, primary cells, iPSC-derived cells and cell lines. Many of these experiments were fully re-processed from raw data using ENCODE pipelines. In addition, the database hosts 44 curated annotations of chromatin state, accessible chromatin sites, transcription factor binding sites, gene expression level in islets and other tissues. Examples of additional experimental and annotation data being added over time include chromatin interactions, high-throughput allelic assays, and expression and chromatin quantitative trait loci (QTL), and data from different disease states and cellular conditions. Available and proprietary tools for interacting with these data include querying genetic variant overlap with annotations, visualization of variant regulatory networks, enrichment analyses (e.g. fgwas, LD-score regression), and integration with T2D fine-mapping data. These tools will help prioritization and interpretation of T2D risk variants and identification of disease-relevant annotations. Furthermore, our database provides further integration with T2D genetic data through interface with the T2D Knowledge Portal (www.type2diabetesgenetics.org). In total our resource, data and tools should facilitate studies to understand the genetic mechanisms underlying T2D pathogenesis.

1827W

Parallel accelerated evolution in hibernating mammals identifies novel non-coding regions for shaping behavioral and metabolic traits linked to various human diseases. E. Ferris, C. Gregg. Neurobiology and anatomy, University of Utah, Salt Lake City, UT.

Introduction. Hibernating mammals store large amounts of fat each year to be used as their primary metabolic fuel during torpor. As these mammals transition between metabolic fuels, entering and exiting hibernation, they go through changes including states of reversible insulin resistance and adiposity. Hibernation also involves dramatic modulation of blood clotting and the immune system. Understanding the genetic basis of these transitions may inform future treatment of human diseases like obesity and diabetes, as well as autoimmune and clotting disorders. Hibernation is an ancient trait and its frequent reappearance in different mammalian lineages suggests that the genes involved may be conserved, but gene regulatory mechanisms changed in hibernators compared to homeotherms. **Methods.** To identify non-coding elements shaping hibernation-related phenotypes, we compare genomic patterns of accelerated evolution in four distantly related and phenotypically distinct hibernators: thirteen-lined ground squirrel, mouse lemur, little brown bat and lesser hedgehog tenrec. We tested for parallel hibernation accelerated regions (ARs) in the genome and determine whether parallel hibernation ARs are enriched at seasonally regulated genes in hibernating species. We examine the nature of human homologs of hibernation ARs. **Results.** We identify 2,370 ARs with evidence for parallel, accelerated evolution in at least two of the four hibernators tested. These ARs are enriched near genes seasonally regulated in hibernating lemurs and squirrels. Human homologs of hibernation ARs are putative functional non-coding elements and are significantly overrepresented near genes linked to body mass index (BMI) in humans and genes differentially expressed in the hypothalamus of people with Prader-Willi syndrome compared to healthy controls. **Conclusion.** Our study identifies non-coding genetic elements that may regulate hibernation related traits in mammals. Further study of the functional properties of these novel elements in the mammalian genome may reveal novel mechanisms for shaping behavioral and metabolic traits involved in various human diseases.

1828T

Functional genomics implicates folliculin as putative causal gene for diabetic retinopathy. M. Grassi¹, S. Jung², A. Skol³, S. Chen⁴, S. Faza¹, D. Cao¹, B. Stranger¹, The DCCT/EDIC Study group. 1) University of Illinois at Chicago, Chicago, IL; 2) Computation Institute, The University of Chicago, Chicago, IL; 3) Department of Medicine, The University of Chicago, Chicago, IL; 4) Cellular Screening Center Institute for Genomics & Systems Biology The University of Chicago, Chicago, IL.

Diabetic retinopathy is the leading cause of irreversible vision loss in the western world. We sought to identify and validate genes contributing to the genetic basis of diabetic retinopathy through changes in gene expression. We hypothesized that genes that display differential responses to glucose in diabetic subjects with and without retinopathy may contribute to genetic risk of diabetic retinopathy. To test this, we identified genes differentially expressed in diabetic patients with and without retinopathy. We tested for differential responses between diabetic subjects with proliferative diabetic retinopathy and without retinopathy matched for age, gender and ethnicity in EBV-transformed lymphoblastoid cell lines (LCLs). We measured gene expression before and after exposure to high glucose, defining a response to glucose phenotype for each LCL. Using genotype data from over 2.5M SNPs in each of 2829 diabetic subjects, we determined whether the association of expression quantitative trait loci (eQTLs), was consistent with our experimental observations of gene expression in the LCLs. Our analysis of a total of 15,569 genes identified 103 genes that exhibited significant differential expression in the response to glucose (p -value < 0.01) between subjects with proliferative retinopathy and no retinopathy. We identified all of the genetic variants (eQTLs) that control gene expression for these genes and tested them for association to proliferative diabetic retinopathy. The 7253 eQTLs for these 103 genes demonstrated significant association with diabetic retinopathy (q -value < 0.05 , p -value = 0.003 vs. all eQTLs, and p -value = 0.006 vs. all SNPs). We identified *Folliculin* (*FLCN*), which plays a role in modulating AMPK signaling, as one of the putative susceptibility genes for diabetic retinopathy. *FLCN* expression was up-regulated in response to glucose in subjects with proliferative diabetic retinopathy compared to diabetic subjects without retinopathy ($\log_2FC=0.276$, p -value=0.003). rs11867934, an eQTL of *FLCN* was significantly associated with proliferative diabetic retinopathy in two large cohorts of subjects with type 1 diabetes (meta p -value 6.7×10^{-6} , Bonferroni-corrected p -value < 0.05). Therefore, *FLCN* may be a putative causal gene for diabetic retinopathy.

1829F

Loss of star strand miR-192 resulted in liver steatosis to compensate for hyperglycemia in diabetes. KK. Miu, X. Zhang, S. Hao, HH. Cheung, W. Chao, B. Feng, WY. Chan. School of Biomedical Sciences, The Chinese University of Hong Kong, NT, Hong Kong, China.

Liver steatosis is a sign of lipid overload often alongside impaired insulin signaling. In parallel, steatosis in diabetic mice is manifested by the microvesicular feature associated with poor disease prognosis. Circulating levels of miR-192 (5p, the guiding strand) was demonstrated for its prognostic value in diabetic complications, whilst being of high *in situ* expression suggested it has functions in liver. Reports indicated that it affects hepatic lipid metabolism via targeting SREBF1 and SCD1 along the *de novo* lipogenesis (DNL) pathway. Since the miRNA is tightly linked to diabetes, we screened and noted a lower expression of miR-192* (3p, the star strand) but not 5p in the liver of diabetic mice when compared to heterozygous lean control. As its expression in visceral adipose tissue was reported to correlate with serum fat, its loss in the liver may represent a cause for liver steatosis—a co-morbidity in diabetes. To recapitulate the loss of miRNA in the diabetic liver, adenoviral sponge overexpression was used to deplete it in wild type mice given 30% fructose in drinking water. In these animals, hepatic lipid accumulation was increased with mild microvesicular steatosis after 2 months. To examine the action of its replenishment, we constructed an artificial shRNA stem-loop construct for star-strand specific overexpression. Diabetic mice with the construct demonstrated significant reduction of hepatic steatosis despite impaired glucose tolerance. To investigate the rationale for reduced steatosis, hepatic cell lines with star overexpression had reduced DNL under oil red staining when exogenously supplemented with fatty acids (FAs). We also observed that these cell lines promoted mitochondrial oxidation of FAs measured by Seahorse XF cell mito stress test, therefore potent in dissipating lipid load. Careful examination of transcriptome analysis in hepatic cell line overexpressing the construct revealed that it modulated the expression of key genes along the insulin signaling pathway such as IRS2, PEPCK & INSIG1, etc., in addition to those affected by 5p. In fact, forced overexpression of the star miRNA also raised the level of its guiding counterpart. Gene ontology analysis revealed altered AMPK signaling resulting in lipid mobilization. With accumulating reports suggesting malfunctioned glucose utilization as a potential cause for liver steatosis, we argue the overlooked importance of miR-192* in withholding glycerolipid homeostasis in the liver.

1830W

The accessible chromatin landscape of pancreatic islet cell populations.

M. Okino¹, J. Chiou², S. Huang¹, K. Gaulton¹. 1) Department of Pediatrics, UC San Diego School of Medicine, La Jolla, CA, 92093; 2) Biomedical Sciences Graduate Program, UC San Diego, La Jolla, CA, 9209.

Pancreatic islets are comprised of multiple endocrine cell types that regulate glucose homeostasis, and genetic variants affecting islet regulatory activity are a major contributor to type 2 diabetes (T2D). Islet cell types have distinct functions and secrete different hallmark hormones including beta (insulin) and alpha (glucagon) cells which together account for around 90% of the total cell mass. Rarer islet cell types include delta (somatostatin), gamma (PP), and epsilon (ghrelin) cells. While previous studies have mapped accessible chromatin in alpha and beta cells sorted by surface markers, the chromatin landscape of rarer cell types is currently uncharacterized. To address this question, we fixed primary islets using formalin and FACS-sorted fixed cells using antibodies targeting intracellular markers for insulin, glucagon and somatostatin. We then optimized an ATAC-seq protocol for fixed cells to map regions of accessible chromatin from the resulting sorted cell populations. Preliminary analyses of accessible chromatin generated from alpha cells identified 66.5k peaks and passed ENCODE quality metrics (e.g. FRiP>0.3) and showed high correlation to published alpha cell data (Spearman $r>0.8$). We identified clear peaks at the promoters of alpha cell-specific genes such as *GCG* and *DPP4* and no signal at promoters for marker genes of other cell types such as insulin (*INS*) or somatostatin (*SST*). At distal enhancer peaks, we observed motif enrichment for known transcription factors important for alpha cell function and glucagon expression such as *FOXA1* and *MAFB*. We integrated alpha cell peaks with genetic fine-mapping data for 107 T2D signals and found high probability candidate causal variants in alpha cell peaks at 10 signals including *TCF7L2*, *ADCY5*, *CDC123* and *ZMIZ1*. We are currently sequencing sorted alpha, beta and delta cells from an initial three islet donors and are extending and scaling up this strategy to sort cells from all islet cell types in larger sample numbers. Our data will provide high-quality reference maps of common and rare islet cell populations to define cellular regulatory programs and understand their specific roles in T2D risk. Furthermore, our results provide an optimized method for generating accessible chromatin profiles from fixed cells that should be widely applicable to studies of human samples.

1831T

Impact of pre-gestational body mass index and midgestational weight gain on DNA methylation of candidate CpG sites. *J.O. Opsahl¹, G.H. Moen², K.I. Birkeland^{1,3}, C. Sommer².*

1) Institute of Clinical Medicine, University of Oslo, Oslo, Norway; 2) Dep. of Endocrinology, Morbid Obesity and Preventive Medicine, Oslo University Hospital, Oslo, Norway; 3) Department of Transplantation Medicine, Oslo University Hospital, Oslo, Norway.

Weight status prior to pregnancy and gestational weight gain during pregnancy is associated with increased risk of gestational diabetes (GDM). GDM increases the possibility of adverse pregnancy outcomes and increased risk of developing type 2 diabetes for both the mother and her offspring. The underlying mechanisms for the association between gestational weight status and gestational diabetes are largely unknown, but epigenetic alterations such as DNA methylation may be important. There are few published studies that have explored how pre-pregnant BMI and/or gestational weight gain are associated with DNA methylation in women, and the sample sizes are generally quite small. Here we explore how pre-gestational body mass index and mid-gestational weight gain from week 15-28 are associated with DNA methylation of candidate CpG (5'-C-phosphate-G-3') sites in European and South Asian women. The Stork Groruddalen study is a population-based multi-ethnic cohort of 823 pregnant women performed in Groruddalen, Oslo, Norway, 2008-2010. We have quantified DNA methylation at 850 000 CpG sites (Infinium MethylationEPIC BeadChip, Illumina) in the European (n=314) and South Asian (n=166) women who participated in the study. We here report candidate CpG sites detected in an epigenome-wide association study (EWAS) as well as candidate CpG sites found in studies outside of pregnancy. This study will provide a unique possibility to explore epigenetic factors, and will increase our understanding of how pre-pregnant BMI and weight gain are associated with DNA methylation. We will discuss the results of the analysis and their potential implications in the understanding of weight gain in pregnancy and pre-pregnancy BMI. Our results may increase our understanding of the mechanisms behind the detrimental effect of obesity on diabetes risk.

1832F

Systematic analysis of genetic variants with allelic effects on transcription factor binding.

Y. Qiu^{1,2}, J. Yan^{1,3}, A.M. Ribeiro-dos-Santos^{1,4}, Y. Yin⁵, Y. Li^{1,5}, N. Vinckier⁶, N. Nariai⁷, A. Raman^{1,2}, E. Morgunova³, A. Mahajan⁸, K.J. Gaulton^{7,9}, M.I. McCarthy^{8,10,11}, K.A. Frazer^{7,9}, M. Sander^{6,9,12}, J. Taipale^{3,13}, B. Ren^{1,12}. 1) Ludwig Institute for Cancer Research, 9500 Gilman Dr, La Jolla, CA 92093, USA; 2) Bioinformatics and Systems Biology Graduate Program, University of California San Diego, 9500 Gilman Dr, La Jolla, CA 92093, USA; 3) Department of Medical Biochemistry and Biophysics, Division of Functional Genomics and Systems Biology, Karolinska Institutet, 17165 Stockholm, Sweden; 4) Universidade Federal do Pará, Institute of Biological Sciences, Belém, PA 66075-110, Brazil; 5) MOE Key Laboratory of Bioinformatics, Center for Synthetic and Systems Biology, School of Life Sciences, Tsinghua University, Beijing 100084, China; 6) Pediatric Diabetes Research Center, Department of Pediatrics, University of California San Diego, La Jolla, CA 92093, USA; 7) Department of Pediatrics, University of California San Diego, 9500 Gilman Dr, La Jolla, CA 92093, USA; 8) Wellcome Trust Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, OX3 7BN, UK; 9) Institute for Genomic Medicine, University of California San Diego, 9500 Gilman Dr, La Jolla, CA 92093, USA; 10) Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of Medicine, University of Oxford, Oxford, OX37LE, UK; 11) Oxford NIHR Biomedical Research Centre, Oxford University Hospitals Trust, Oxford, OX37LE, UK; 12) Department of Cellular and Molecular Medicine, University of California San Diego, 9500 Gilman Dr, La Jolla, CA 92093, USA; 13) Genome-Scale Biology Program, P.O. Box 63, FI-00014 University of Helsinki, Finland.

Genome-wide association studies (GWAS) have identified thousands of SNP associations to complex diseases, and the vast majority of them are located in intergenic regions. Molecular characterization of GWAS signals can therefore reveal key gene regulatory mechanisms underlying complex disease risk. Recent studies have linked specific risk loci to disrupted transcription factor (TF) binding and chromatin accessibility. However, most risk variants have not been systematically characterized for their effects on TF binding. Here, we adapted a modified HT-SELEX pipeline using 40bp human genomic sequences as ligands to quantitatively survey allelic imbalance of TF binding, termed SNP-SELEX. In total, we reported allelic binding specificities for 270 distinct human TFs at 95,887 candidate variants. The SNP-SELEX results showed a strong correlation with *in vivo* allele-biased TF binding and explained allelic imbalance in chromatin accessibility and enhancer activity. We then extended results of SNP-SELEX to predict regulatory variants affecting TF binding genome-wide through machine learning models. Specifically, we adapted deltaSVM to train a model for each TF using enriched sequences in SELEX experiments which we benchmarked using candidate variants assayed experimentally. In total, we trained models for 167 TFs with median AUROC (area under the receiver operating characteristic curve) 0.96 and AUPRC (area under PR curve) 0.75. We showed that SNP-SELEX models outperformed PWM models, and that predicted allelic TF binding is equivalent to experimental SNP-SELEX allelic TF binding with regard to *in vivo* allelic TF ChIP-seq binding, allelic imbalance in chromatin accessibility, and enhancer activity. We applied SNP-SELEX models to T2D GWAS data and found that variants affecting TF binding are enriched for T2D risk variants. Finally, we performed *in situ* Hi-C in HepG2 cells and normal human islet tissue to impute the target genes of variants with allelic TF binding. We found that allelic TF binding variants predicted by SNP-SELEX models are correlated with allelic gene expression of target genes, suggesting that SNPs affecting TF binding also affect gene expression. In sum, our study addresses an important gap in our knowledge about the impact of variants on TF binding and provides a valuable resource to interpret regulatory functional variants at disease risk signals through allelic TF binding.

1833W

DNA methylation of TGF- β 1 gene in T2DM associated nephropathy.

S. Singh¹, S. Sonkar², G. Sonkar², A. Mahdi¹. 1) Biochemistry, King George's Medical University, Lucknow, India; 2) Medicine, King George's Medical University, Lucknow, India.

Diabetes is a systemic disorder, over the course of time it affects almost every vital organ and ranked among leading causes of morbidity and mortality due development of micro and macro vascular complications in various organ. Diabetes also affects kidney and is now become the leading cause of End Stage Renal Disease. A total of 98 subjects were recruited in study including 37 T2DM, 22 DN and 39 healthy controls. Estimation of serum TGF- β 1 was done by ELISA kit according to manufacturer's protocol. Methylation study was performed using methylation specific PCR. There was significant association of family history with T2DM. About 62.12% patients with T2DM had positive family history for diabetes. In 32.43% patients with T2DM and 27.27% patients with DN had tobacco chewing habit. Systolic pressure was found significantly increased in T2DM as well as DN group as compared to control group ($p = 0.006$ and 0.001 respectively) while diastolic pressure was significantly increased in DN group as compared to control (0.002). Glycosylated Haemoglobin was significantly decreased in T2DM and DN in our study group while serum creatinine was significantly raised in DN group. Circulatory levels of TGF- β 1 was significantly higher with mean value 256.07 ± 62.72 ng/ml in T2DM ($p=0.006$), in DN 349.81 ± 76.12 ng/ml (<0.0001) as compared to control (213.06 ± 69.57 ng/ml). ROC analysis of TGF- β 1 showed that area under curve was 87.0% and when we take the cut off value 297.5 ng/ml than sensitivity of this ELISA is 81.8% and Specificity is 77.6%. TGF- β 1 had significant correlation with serum creatinine ($p = 0.044$) and HbA1c ($p=0.037$). About 68.18% patients with DN group, 13.51% DM patients had unmethylation in CpG island and hypomethylation of TGF- β 1 was significantly increased ($p = <0.001$) in DN group. TGF β 1 had significant correlation with blood urea, serum creatinine and glycosylated haemoglobin. Therefore, TGF β 1 is better predictor for progression of nephropathy in our study group.

1834T

Of the relationship between the adipose transcriptome and diabetes, insulin resistance and polygenic risk scores: thousands of transcripts in adipose tissue are associated with diabetes status, fasting insulin levels and insulin resistance PRS. K.S. Small¹, C.A. Glastonbury², A.C. Alves¹, J.S. El-Sayed Moustafa¹. 1) King's College London, London, United Kingdom; 2) Big Data Institute, University of Oxford, Oxford, United Kingdom.

Diabetes affects over 400 million adults worldwide, with type 2 diabetes (T2D) accounting for ~90% of cases. T2D GWAS have highlighted the role of beta-cell dysfunction, with increasing recognition of the role of adipose tissue, through its contribution to insulin resistance (IR). Disentangling the relationship between gene expression and diabetes, and underlying quantitative traits such as fasting glucose (FG) and fasting insulin (FI), will be key to understanding the contribution of tissues beyond the beta cell to diabetes. We investigated the relationship between genetics, gene expression and concurrently measured FI, FG, and T2D/IR case-control status using RNASeq gene expression data from adipose tissue, skin and LCLs in ~800 females from TwinsUK. Glycemic trait–gene expression associations were adipose-specific, with no significant associations in skin or LCLs, but 8212, 111, and 2671 genes associated with FI, FG and T2D/IR, respectively in adipose tissue (FDR 5%). Adipose expression-glycaemic trait results replicated in an external adipose dataset (N=770). Expression levels of 60% of T2D and glycaemic trait GWAS genes were associated with FI (enrichment OR=1.9; P=0.001), including *CDKN2B*, *IDE*, *IRS1*, *KCNQ1*, *KLF14*, *PPARG* and *TCF7L2*. Genes associated with each of FI and T2D/IR were enriched for functional categories including NIDDM (FI: $Q_{\text{Bonferroni}}=6 \times 10^{-10}$; T2D/IR: $Q_{\text{Bonferroni}}=6 \times 10^{-9}$), response to glucose (T2D/IR: $Q_{\text{Bonferroni}}=1 \times 10^{-9}$) and cellular response to insulin (FI: $Q_{\text{Bonferroni}}=3 \times 10^{-9}$). To investigate shared genetic aetiologies between T2D/IR and adipose gene expression, we calculated polygenic risk scores (PRS) for both T2D and IR using published GWAS SNPs. In TwinsUK (N=6448), both IR and T2D PRS were associated with T2D (IR PRS $P=1 \times 10^{-4}$; T2D PRS $P=1 \times 10^{-9}$) and T2D/IR case-control status (IR PRS $P=6 \times 10^{-7}$; T2D PRS $P=3 \times 10^{-9}$). After BMI correction, IR, but not T2D PRS was strongly associated with DXA-derived adiposity traits (IR PRS $P_{\text{min}}=9 \times 10^{-22}$). The two PRS showed a strikingly different relationship to the adipose transcriptome; after BMI correction, at an FDR 5% 6,191 genes were associated with the IR PRS, but zero genes with the T2D PRS. This may reflect overrepresentation of beta-cell-driven effects among T2D GWAS hits, or mediation of adipose-specific effects at these loci through BMI. This study helps further elucidate the role of adipose gene expression in risk of type two diabetes and mediation of downstream complications.

1835F

Integration of epigenomic profiles from islet developmental precursors with genome-wide association data reveals contributions of developmental processes to type 2 diabetes risk. A. Wang^{1,2}, R. Geusz^{1,2}, J. Chiou^{1,2}, N. Wetton^{1,2}, J. Wang^{1,2}, S. Kefalopoulou^{1,2}, I. Matta^{1,2}, A. Aylward^{1,2}, B. Ren^{3,4}, K.A. Frazer¹, K.J. Gaulton^{1,2,5}, M. Sander^{1,2,5}. 1) Department of Pediatrics, UC San Diego, La Jolla, CA; 2) Pediatric Research Center, UC San Diego, La Jolla, CA; 3) Ludwig Institute for Cancer Research, La Jolla, CA; 4) Department of Cellular and Molecular Medicine, UC San Diego, La Jolla, CA; 5) Correspondence.

Genetic risk of type 2 diabetes (T2D) is primarily driven by variation within the non-coding genome. T2D risk variants largely impact genomic elements that regulate pancreatic islet function or other tissues involved in glucose homeostasis such as liver, adipose, and skeletal muscle. In this work, we aimed to uncover novel mechanisms of T2D genetic risk beyond the mature tissues typically associated with disease pathogenesis. Specifically, we focused on the embryonic pancreas to investigate whether pancreatic developmental processes contribute to T2D risk. To address this question, we generated genome-wide chromatin accessibility (ATAC-seq), and chromatin state (histone modification ChIP-seq) maps representing early (stem cell, definitive endoderm, gut tube) and late (pancreatic progenitor) stages of an *in vitro* human pluripotent stem cell (hPSC) differentiation model of human pancreas development. We observed enrichment of T2D-associated variants in regions of open chromatin within 'stretch' enhancer elements specific to late pancreatic progenitors which are turned off in differentiated islets. Genes proximal to these elements were predicted to regulate core developmental processes such as cell proliferation, polarity, and morphogenesis; processes which play critical roles in pancreatic cell allocation and determination of islet cell mass. Among known fine-mapped T2D risk signals that mapped in progenitor-specific 'stretch' enhancers and not in islet enhancers was *LAMA1*. Further integration of progenitor-specific enhancer and genome-wide T2D association data revealed novel candidate T2D risk loci with proximal genes predicted to also regulate developmental processes. To functionally validate the stretch enhancers harboring T2D risk variants, we employed CRISPR-Cas9 genome editing in hPSCs and differentiated these hPSCs toward pancreas. We determined that multiple T2D-risk stretch enhancers at both known and novel risk loci are required for proper expression of *LAMA1* and other T2D risk-associated genes in pancreatic progenitor cells. Altogether, our results support a novel role for genetic variants affecting pancreatic developmental regulatory programs in T2D risk, independent of genetic effects on mature islet cells and other T2D-relevant cell types.

1836W

Role of breastfeeding on epigenetic mechanisms underlying early-life growth trajectories. L. Briollais¹, D. Rustand^{1,2}, P. McGowan³, S. Matthews³, S. Lye². 1) Lunenfeld-Tanenbaum Research Institute, Toronto, Canada; 2) ISPED, University of Bordeaux, France; 3) University of Toronto, Toronto, Canada.

Introduction: Previous research has clearly established a link between early environment (pre-natal and post-natal), genetic, and behavioural factors on the development origin of health and diseases (DOHaD). Among the environmental factors, breastfeeding has been advocated in the prevention of overweight/obesity and is recommended by WHO as the "perfect food for the newborn". Each month of breastfeeding results in a 4% reduction in risk of being overweight. The biological mechanisms underlying the role of breastfeeding remain unclear, but a plausible explanation could be through the regulation of epigenetic mechanisms involved in developmental programming. Our goal here was to identify new epigenetic marks associated with birth weight, BMI at birth and longitudinal BMI and to assess whether the duration of exclusive breastfeeding (EXBF) affects these marks during child development. **Methods:** We studied approximately 1,000 children from the British Avon Longitudinal Study of Parents and Children (ALSPAC) cohort and conducted an epigenome-wide association study (EWAS) of birth weight and BMI at birth as well as a longitudinal analysis of child BMI up to 17 years of age in relation to methylation of the EWAS hits, several candidate genes and EXBF. **Results:** The EWAS identified several CpG sites where methylation profile correlated significantly with either birth weight or BMI at birth. We also found several EWAS hits (in genes *NCOR2*, *TMEM30A*, *CHD5*) and CpG sites from candidate genes (*IGF1R*, *IGF2R*, *IGF2BP2*, *IGF2*, *LEP*, *LEPR*, *NPY*, *SLC2A4*, *CDKN2A*) where methylation profile was modulated by EXBF and impacted children longitudinal BMI through infancy and adolescence. Many of the differential methylated genes form part of the mTOR signaling pathway. These CpG sites could be hyper- or hypo-methylated by EXBF but the interaction with EXBF systematically resulted in reduced BMI in the first year of life. The CpG sites involved could have differential age- and sex-specific mechanisms. **Conclusion:** Our study demonstrates that EXBF could regulate early life epigenetic modifications of key genes involved in developmental programming, especially via modulation of the mTOR pathway, which in turn could impact the risks of overweight/obesity in children. Our study could have implications for the future prevention of overweight and obesity in both children and adults.

1837T

Differentially methylated loci in NAFLD cirrhosis are associated with key signaling pathways. J.K. DiStefano¹, I. Malenica¹, L. Laci¹, X. Chu², A.T. Petrick², C.D. Still², G.S. Gerhard³. 1) Translational Genomics Research Institute, Phoenix, AZ; 2) Geisinger Obesity Institute, Danville, PA; 3) Temple University School of Medicine.

Altered DNA methylation events contribute to the pathogenesis and progression of metabolic disorders, including nonalcoholic fatty liver disease (NAFLD). Investigations of global DNA methylation patterns in liver biopsies representing severe NAFLD fibrosis have been limited. The purpose of this study was to investigate global patterns of DNA methylation in NAFLD-related cirrhosis. Patients with biopsy-proven NAFLD fibrosis/cirrhosis (N=14), and age- and sex-matched controls with normal histology (N=15) were recruited from the Bariatric Surgery Program at the Geisinger Clinic Center. Whole genome methylation was performed using the 450K Infinium Methylation BeadChip assay. Methylation levels for each CpG residue were estimated as the ratio of the methylated signal intensity over the sum of the methylated and unmethylated intensities at each locus and differential methylation between fibrotic and normal samples was determined using the City of Hope CpG Island Analysis Pipeline (COHCAP) and the minfi software. Hepatic gene expression was measured using RNA-sequencing. To identify gene-methylation correlations, we compared beta-values for each CpG site to read counts of its corresponding gene obtained from RNA-sequencing analysis. Ingenuity Pathways Analysis software was used to identify canonical signaling pathways and establish network connections associated with dysregulated CpG islands. We identified 208 CpG islands (CGI), including 99 hypomethylated and 109 hypermethylated CGI, showing statistically significant evidence (FDR <0.05) for differential methylation between cirrhotic and normal samples. Comparison of beta-values for each CGI to the read count of its corresponding gene revealed negative correlation (FDR <0.05) for 34 transcripts. Functional enrichment analysis of the differentially methylated CGI identified 113 significantly enriched canonical pathways, many classified as cell signaling. Comparison of our results with previously published studies identified 86 shared sites showing differential methylation, suggesting that genes common to multiple diverse studies may be important in disease pathogenesis. These findings provide additional evidence supporting a role for CpG methylation in the pathogenesis of NAFLD-related cirrhosis, including replication of previously reported differentially methylated CGIs, and contribute new insight into the molecular mechanisms underlying the initiation and progression of liver fibrosis and cirrhosis.

1838F

Characterization of open chromatin and chromosomal interactions in human primary adipocytes reveals environment-responsive cardiometabolic GWAS loci. K.M. Garske¹, Y. Bhagat¹, D.Z. Pan^{1,2}, M. Alvarez¹, P. Pajukanta^{1,2,3}. 1) Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, USA; 2) Bioinformatics Interdepartmental Program, UCLA, Los Angeles, USA; 3) Molecular Biology Institute at UCLA, Los Angeles, USA.

We hypothesized that environmental contexts affect genomic regulatory elements underlying cardiometabolic disease. To detect adipocyte-specific open chromatin regulatory regions that respond to lipid challenge, we first performed Assay for Transposase-Accessible Chromatin (ATAC) -sequencing in human primary adipocytes and their precursors. We found that 32.5% (50,336/154,647) of the ATAC-peaks were more accessible in the adipocytes than in preadipocytes. Compared to all ATAC-peaks, these adipocyte-accessible peaks exhibit an enrichment in both the CCAAT-Enhancer Binding Protein (CEBP) transcription factor motif (p -value= 1.0×10^{-39}) and in imputed enhancer annotations for adipocytes (1.46X larger effect size). Next, to detect regions responding to lipid challenge, we treated the human primary adipocytes with monounsaturated or saturated fatty acids and performed ATAC-seq, which resulted in 1,653 differentially accessible, lipid-responsive (LR) peaks, 572 of which (34.6%) are adipocyte-accessible peaks. A total of 75.1% of the adipocyte-accessible LR peaks fall within enhancers and 14.8% within promoters, suggesting that these regions are highly important for adipocyte gene regulation. Then, to detect active enhancers and their target genes, we performed promoter Capture Hi-C in the adipocytes and found that 277/572 (49.3%) of the adipocyte-accessible LR peaks either fall within an interacting promoter or within a promoter-interacting enhancer fragment. GWAS variants within these regions can give insight into environment-responsive mechanisms contributing to disease phenotypes. We found that 23 cardiometabolic GWAS SNPs fall within the adipocyte-accessible LR peaks involved in chromosomal interactions, including promoters of 8 genes. These SNPs included lipid-associated SNPs in the *FADS1/2* promoter. Supporting our data, SNPs at the *FADS1/2/3* locus have been shown to exhibit positive selection in Inuits, likely due to selective pressure from their high-fat diet. Other environment-responsive GWAS loci we identified include the promoter regions for *HOXC9* (waist-hip-ratio), *CARM1*, and *MLXIPL* (serum lipids). Taken together, our data identify 8 novel environment-responsive metabolic GWAS genes at open chromatin, promoter-enhancer regulatory circuits in human primary adipocytes. This implies that some cardiometabolic GWAS loci are influenced by the environmental context, providing strong candidates for gene-environment interaction analysis in large cohorts.

1839W

Epigenomic map of human liver reveals principles of zonated morphogenic and metabolic control. J. Hampe¹, K. Kattler², A. Herrmann¹, W. von Schönfels³, D. Seehofer⁴, G. Damm⁴, T. Becker⁵, G. Baretton⁵, C. Röcken⁶, M. Muders⁷, M. Matz-Soja⁷, M. Krawczak⁸, G. Gasparoni¹, A. Dahl⁹, C. Schafmayer⁹, J. Walter¹, M. Brosch². 1) Medical Department 1, University Hospital Dresden, Technische Universität Dresden (TU Dresden), Dresden, Germany; 2) Department of Genetics & Epigenetics, Universität des Saarlandes, Saarbrücken, Germany; 3) Department of Visceral Surgery, University Hospital Schleswig-Holstein, Christian-Albrechts-University Kiel, Kiel, Germany; 4) Department of Hepatobiliary Surgery and Visceral Transplantation, University of Leipzig, Leipzig, Germany; 5) Institute of Pathology, University Hospital Dresden, Technische Universität Dresden (TU Dresden), Dresden, Germany; 6) Institute of Pathology, University Hospital Schleswig-Holstein, Christian-Albrechts-University Kiel, Kiel, Germany; 7) Rudolf-Schönheimer-Institute for Biochemistry, University of Leipzig, Leipzig / Germany; 8) Center for Regenerative Therapies Dresden (CRTD), Technische Universität Dresden (TU Dresden), Dresden Germany; 9) Institute of Medical Informatics and Statistics, Christian-Albrechts University, Kiel, Germany.

A deeper epigenomic understanding of spatial and functional organization of cells in human tissues is an important challenge. Here we report the first combined positional analysis of transcriptomes and methylomes across three micro-dissected zones (pericentral, intermediate and periportal) of human liver. We identify pronounced anti-correlated transcriptional and methylation gradients including a core of 271 genes controlling zonated metabolic and morphogen networks. A prominent porto-central gradient of DNA methylation at binding sites for 46 ENCODE transcription factors points towards a methylation gradient dependent control of transcription. The gradient includes specific epigenetic and transcriptional signatures in the wnt pathway supporting the concept of a pericentral hepatocyte regeneration pathway in humans under steady-state conditions. Surprisingly, while metabolic gene expression changes across zones are observed with increasing severity of non-alcoholic fatty liver disease, the relative zonated expression and methylation differences across zones remain unchanged. This observation suggests a maintained molecular zonation during reversible early NASH states. Overall our data provide a wealth of new positional insights into zonal networks controlled by epigenetic and transcriptional gradients in the human liver lobule.

1840T

Genetic variation affects browning potential of white adipose tissue through genome-epigenome interactions at the *Ucp1* enhancer. Y. Hiraike¹, H. Waki^{1,2}, S. Tsutsumi³, H. Aburatani³, T. Yamauchi¹, T. Kadowaki^{1,4}.

1) Department of Diabetes and Metabolic Diseases, Graduate School of Medicine, The University of Tokyo, Bunkyo-ku, Tokyo, Japan; 2) Department of Molecular Science on Diabetes, Graduate School of Medicine, The University of Tokyo, Bunkyo-ku, Tokyo, Japan; 3) Genome Science Division, Research Center for Advanced Science and Technology, The University of Tokyo, Meguro-ku, Tokyo, Japan; 4) Department of Prevention of Diabetes and Lifestyle-Related Diseases, Graduate School of Medicine, The University of Tokyo, Bunkyo-ku, Japan.

While white adipose tissue (WAT) stores energy as lipids and expands in obesity, brown adipose tissue (BAT) dissipates energy as heat, mainly by means of the uncoupling protein-1 (Ucp1) on the mitochondrial inner membrane. Ucp1-positive adipocytes are also found sporadically in WAT depots after cold and sympathetic stimulation, and this phenomenon is called "browning" or "beiging" of WAT. Stimulating development and/or function of brown and beige adipose tissue is highly anticipated as a novel strategy for the treatment of obesity and its complications including diabetes. However, it is well known that there are more than ten-fold variations in BAT activity among human individuals. The same is also true for mice. Inbred mice strains such as A/J and 129 are resistant to diet-induced obesity compared to B6 strain, and browning potential of WAT of A/J and 129 is much higher than that of B6. Here, we have developed a primary beige adipocyte culture system from WAT of B6, 129 and F1 offspring of B6 and 129 to perform comprehensive, genome-wide analyses to dissect the mechanisms underlying the strain difference. Expression levels of Ucp1 is much higher in 129 than in B6 in a cell-autonomous manner. Transcriptome analyses led us to identify a transcription factor Lim homeobox protein 8 (Lhx8), whose expression is significantly higher in 129 than in B6, as a novel browning regulator. Nevertheless, based on the analyses using the F1 cells, we argue that the difference in *Ucp1* expression between B6 and 129 strain primarily comes from the *cis* effect rather than the *trans* effect. In the F1 cells, we observed significant and strong (>70%) allelic imbalance of the *Ucp1* mRNA transcripts favoring 129. At the *Ucp1* enhancers, we also observed 129-favored imbalance of transcription factor binding including PPAR γ and NFIA, a transcription factor which we recently identified as a crucial regulator of the BAT gene program. Through promoter capture Hi-C and computational analyses, we identified candidate SNPs responsible for the strain difference at the *Ucp1* -11.7kb enhancer, as well as a candidate transcription factor which negatively regulates *Ucp1* expression and its binding might be negatively affected by one of the SNPs in the 129 alleles. These results would offer a platform for understanding the mechanisms underlying the substantial variation in human BAT activity and for the implementation of "precision medicine" in BAT-targeted anti-obesity therapy.

1841F

Epigenome-wide association study of change in body mass index and glycaemic trait levels from young- to middle adulthood in 595 Northern Finland Birth Cohort 1966 participants. I. Prokopenko¹, M. Kaakinen^{1,2}, L. Prelot³, M.D. Anasanti⁴, Z. Balkhiyarova¹, M. Wielscher³, S. Sebert^{1,4}, M.-R. Jarvelin^{3,4}, H. Draisma¹.

1) Section of Genomics of Common Disease, Department of Medicine, Imperial College London, London, United Kingdom; 2) Centre for Pharmacology and Therapeutics, Department of Medicine, Imperial College London, London, United Kingdom; 3) Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, United Kingdom; 4) Center for Life Course Health Research, University of Oulu, Oulu, Finland.

Genome-wide association studies (GWAS) have identified thousands of genetic loci for various phenotypes cross-sectionally, but GWAS haven't been successful in identifying *longitudinal* genetic effects. A number of associations of interindividual variation in adiposity and glycaemic trait levels with variation in the degree of DNA methylation have been reported. However, the relationship between *longitudinal* changes in these phenotypes and differential methylation is underexplored. We aimed to develop novel methodology for the detection of longitudinal effects on DNA methylation and tested it on the examples of change in body mass index (BMI), waist/hip ratio (WHR), fasting blood plasma glucose (FG) and fasting blood insulin (FI) levels in individuals without diabetes over time. For 595 individuals from the Northern Finland Birth Cohort 1966 for whom both phenotype and concurrent methylation data were available at both ages 31 (T1) and 46 (T2), we calculated the average change in phenotype value per year between T1 and T2. For all phenotypes we used change residuals corrected for sex to test for association with the degree of methylation measured in blood at T2 for 832,569 markers on the Illumina (San Diego, CA, USA) MethylationEPIC BeadChip using methylSCO-PA software, developed by us. We quality-controlled the methylation data, regressed out the effects of measured (potential) confounders, normalized the methylation signal intensity data, and mapped genomic locations to CGCh37/hg19. We detected epigenome-wide significant associations ($P < 1 \times 10^{-7}$) with BMI change at cg14476101 (chr1:120,255,992; $\beta = -0.05$, $SE = 9.1 \times 10^{-3}$), cg11376147 (chr11:57,261,198; $\beta = -0.01$, $SE = 2.6 \times 10^{-3}$), and cg03067296 (chr17:76,274,577; $\beta = -0.05$, $SE = 8.1 \times 10^{-3}$) representing *PHGDH*, *SLC43A1*, and *LOC100996291* loci, respectively. *PHGDH* and *SLC43A1* are established for concurrent BMI and for BMI change, whereas the association with *LOC100996291* is novel. For FG change, we observed epigenome-wide significant association with cg20367077 (chr11:19,224,117; $\beta = 0.11$, $SE = 0.02$) annotated to *CSRP3*, implicated in obesity-induced insulin resistance in murine and human skeletal muscle. We didn't observe epigenome-wide significant associations with change in either WHR or FI. We implemented a novel method to detect associations between change in phenotype over time and DNA methylation and provide the first account of its use on the example of longitudinal changes in measures of obesity and glycaemic trait levels.

1842W

High-dimensional mediation analysis in causal inference framework. Y. Song¹, X. Zhou¹, M. Zhang¹, W. Zhao², J. Smith², B. Needham², B. Mukherjee^{1,2}. 1) Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI; 2) Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI.

Causal mediation analysis aims to examine the role of a mediator or a group of mediators that lie in the pathway between an exposure and an outcome. Recent studies often involve a large number of potential mediators based on high-throughput technologies. Most of the current analytic methods focus on settings with one or a moderate number of potential mediators. With the expanding growth of omics data, joint analysis of molecular-level genomics data with epidemiological data through mediation analysis is becoming more common. However, this requires methods that can simultaneously accommodate high-dimensional mediators. To address this problem, we develop a Bayesian inference method using continuous shrinkage priors to extend previous causal mediation analysis techniques to a high-dimensional setting. Simulations demonstrate that our method improves the power of global mediation analysis and has decent performance to identify true non-null mediators. We also construct tests for natural indirect effects using a permutation procedure. We applied our method to Multi-Ethnic Study of Atherosclerosis (MESA) and identified DNA methylation regions that actively mediate the effect of neighborhood characteristics / socioeconomic status (SES) on health outcomes.

1843T

Integrating epigenetic maps and genetic fine-mapping implicates tissues of action and effector transcripts at loci associated with type 2 diabetes. J.M. Torres^{1,2}, A. Mahajan¹, M. Abdalla¹, A. Payne¹, J. Fernández-Tajes¹, M. Thurner^{1,2}, V. Nylander², A.L. Gloyn^{1,2}, M.I. McCarthy^{1,2}, DIAMANTE Consortium. 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) Oxford Centre for Diabetes, Endocrinology, and Metabolism, University of Oxford, United Kingdom.

The number of loci associated with type 2 diabetes (T2D) has recently expanded to >400 in Europeans. However, resolving the molecular mechanisms that drive genetic risk to T2D largely remains a challenge as most genetic associations are non-coding and their relevant tissue(s) of action is unclear. To meet this challenge, we developed a framework for integrating tissue specific gene expression and epigenetic maps to obtain tissue-of-action (TOA) scores by systematically dividing SNP-level posterior probabilities of association (PPA) from bayesian fine-mapping. PPA signals are partitioned by membership in a set of genomic annotations that are weighted by their enrichment for T2D-associated variants. We applied this scheme to a set of 380 loci from a large and densely-imputed GWAS meta-analysis (>74K T2D cases and >824K controls; imputed to a reference panel of >30K haplotypes) and obtained TOA scores for each locus for islet, liver, skeletal muscle, and subcutaneous adipose tissue. Unsupervised clustering analysis of TOA scores showed primary separation of loci with high islet scores from those with high scores for peripheral insulin-responsive tissues. We next assigned each locus to a most likely tissue-of-action and found that most were classified as islet-driven (45%) followed by adipose (14%), liver (12%) and muscle (7%). Proximal genes annotated to each set of tissue-classified loci exhibited significant co-expression in the corresponding tissue. Moreover, loci classified as islet were enriched for loci at which multi-trait profiles indicated a primary effect on insulin secretion (p-value = 0.013). Conversely, loci assigned as adipose, liver, or muscle were most enriched for loci with trait profiles consistent with insulin action and/or dyslipidemia. Credible SNPs for each set of classified loci were enriched for accessible DNA regions and expression quantitative trait loci (eQTLs) in the corresponding tissue. Furthermore, only islet loci showed improved fine-mapping in a functional GWAS using chromatin states based on islet DNA methylation, ATAC-seq peaks, and histone modifications (p = 0.005). Lastly, we used these TOA scores in conjunction with an eQTL colocalization analysis to jointly prioritize relevant tissues and effector transcripts at T2D loci. This integrative framework guides mechanistic inference by directing functional validation studies to the most relevant tissues and can be incorporated into process-specific genetic risk scores.

1844F

Exploration of DNA methylation sites associated with adiponectin levels based on a gene co-expression network and DNA methylation data analysis.

M. Nakatochi¹, S. Ihichara², K. Yamamoto³, T. Matsubara⁴, M. Yokota⁵.
 1) Center for Advanced Medicine and Clinical Research, Nagoya University Hospital, Nagoya, Japan; 2) Department of Environmental and Preventive Medicine, Jichi Medical University School of Medicine, Shimotsuke, Japan; 3) Department of Medical Biochemistry, School of Medicine, Kurume University, Kurume, Japan; 4) Department of Internal Medicine, School of Dentistry, Aichi Gakuin University, Nagoya, Japan; 5) Department of Genome Science, School of Dentistry, Aichi Gakuin University, Nagoya, Japan.

Aims Abnormal secretion of adiponectin is strongly correlated with the development of lifestyle-related diseases. The serum adiponectin level is known to be affected by single nucleotide polymorphisms (SNPs); however, such SNPs account for only a small proportion of the heritability of the adiponectin concentrations. In this study, we explored DNA methylation (DNAm) sites associated with the adiponectin levels using a gene co-expression network analysis and integration analyses with DNAm levels. **Methods** We applied a three-step approach to identify DNAm sites associated with adiponectin. First, we performed a gene co-expression network analysis using seven publicly available microarray datasets to extract a module most significantly associated with the expression level of the *ADIPOQ* gene that encodes adiponectin protein (*ADIPOQ* module). In the seven datasets, a total of 855 adipose tissue samples were included. The co-expression network analysis was performed using the weighted correlation network analysis. Second, we explored DNAm sites significantly associated with *ADIPOQ* expression levels from DNAm sites located near *ADIPOQ* module genes. The exploration analysis was performed using 626 adipose tissues from a Caucasian population. In addition, we verified whether these DNAm sites were associated with the serum adiponectin levels in 191 whole blood samples from a Japanese population. Finally, we assessed whether the DNAm sites and adiponectin-associated SNPs were independently associated with serum adiponectin levels. **Results** In the gene co-expression analysis, we obtained an *ADIPOQ* module most significantly associated with the *ADIPOQ* expression level. An integration analysis identified three DNAm sites, which were located near *ADIPOQ* module genes, associated with both *ADIPOQ* and serum adiponectin levels. Out of the three sites, DNAm site X was located near *PPARG*, which regulates the expression of *ADIPOQ* ($\beta=0.375\pm0.105$, $P=0.0004$ in adipose tissues, $\beta=0.340\pm0.149$, $P=0.024$ in whole blood). Two DNAm sites, including DNAm site X, remained significantly associated with serum adiponectin levels after adjustment for adiponectin-associated SNPs (DNAm site X: $\beta=0.362\pm0.163$ $P=0.028$). **Conclusions** A gene co-expression analysis and integration analyses with DNAm data identified three DNAm sites associated with both *ADIPOQ* expression levels and serum adiponectin levels. These results provide new insights into the regulation of adiponectin via epigenetic effects.

1845W

Association of miR-421 expression with adiponectin homeostasis and insulin resistance in metabolic syndrome patients.

R.H. Bortolin^{1,2}, A.A. Braga^{1,2}, G.M. Bastos², M.E.G. Saldarriaga^{1,2}, R.C.C. Freitas^{1,2}, H-T.L. Wang², J.B. Borges^{1,2}, L.R. Castro², M.F. Sampaio², J.I.D. França², R.D.C. Hirata^{1,2}, M.H. Hirata^{1,2}.
 1) Faculdade de Ciências Farmacêuticas, Universidade de São Paulo, São Paulo, SP, Brasil; 2) Laboratório de Investigação Molecular em Cardiologia, Instituto Dante Pazzanese de Cardiologia, São Paulo, SP, Brasil.

Metabolic syndrome (MetS) is a cluster of metabolic disorders that can affect the over 20% of global adult population. Effective treatment of MetS and its complications have been challenging. Understanding the molecular mechanisms involved in the adipokines homeostasis may contribute to identify novel biomarkers and/or drug targets in MetS. We investigated the association of plasma miRNA profile with metabolic and adipokines variables in MetS Brazilian patients. MetS (n=88) and non-MetS (n=144) subjects were classified according International Diabetes Federation criteria. Biometric information was recorded and blood samples were obtained for clinical laboratory testing and miRNA profile analysis. Higher values of waist-to-hip ratio, fat (%), body weight, triacylglycerols, VLDLc, glucose, glycated hemoglobin (Hb1Ac), HOMA-IR, usCRP, PAI-1, fibrinogen, TNF- α , IL-6, resistin, and lower values of HDLc, apolipoprotein AI and adiponectin were found in MetS than in non-MetS subjects ($p<0.05$). Insulin resistance (RI) was found in 85.4% of the MetS patients (85.4%). miRNA PCR array (372 targets) was used for initial screening in a small group of MetS patients (n=6) and health control subjects (n=6). Seven miRNAs (miR-542-5p, miR-424-3p, miR-326, miR-421, miR-301a-3p, miR-183-5p and miR-574-3p) were differentially expressed ($p<0.05$; FC<-2 and >2) and the miR-421 and miR-183-5p remained upregulated in the larger group of MetS when compared to non-MetS group ($p<0.05$). Plasma miR-421 expression was inversely correlated with and adiponectinemia ($r=-0.242$, $p=0.038$). Linear regression analysis showed that plasma miR-421 expression explains variability in 9.5% the adiponectin ($R^2=0.095$, $p<0.05$), 12.5% the HOMA-IR ($R^2=0.125$, $p<0.05$) and 7.5% the HbA1c ($R^2=0.075$, $p<0.05$). Interestingly, the plasma miR-421 expression was also upregulated in RI in comparison with non-RI subjects ($p=0.029$). These results are suggestive that miR-421 may have a potential role in adiponectin homeostasis and insulin resistance. Even though further functional studies are necessary to confirm these interesting findings.

1846T

Investigation of the impact of a nutritional intervention in the first year of life on DNA methylation in a cohort of children. V.S. Mattevi¹, J.K. Castilhos¹, P.D.B. Campagnolo², S. Almeida¹, M.R. Vitolo¹. 1) Federal University of Health Sciences of Porto Alegre, Porto Alegre, RS, Brazil; 2) Universidade do Vale do Rio dos Sinos, São Leopoldo, RS, Brazil.

Experimental models of fetal programming have demonstrated that changes in the gestational and neonatal periods alter DNA methylation, affecting gene expression and being associated with processes involved in the energy balance. Although the effect of the fetal and early childhood environment on health outcomes in adult life has been widely discussed, the empirical assessment of this hypothesis in humans remains a challenge. Therefore, the present study evaluated DNA samples collected at 4 years old from children submitted to a nutritional intervention in the first year of life in order to evaluate the effect of this intervention on global DNA methylation levels. This research is part of a cohort study carried out within a randomized field trial that included 500 children born from 2001 to 2002 in the main hospital from the city of São Leopoldo, Southern Brazil, which mainly serves a low-income population. The children were randomly assigned in control and intervention groups at birth. The intervention group followed a diet based on the "Ten Steps to Healthy Feeding", as advised by Brazilian Ministry of Health, during the first year of life. DNA samples were obtained from peripheral leukocytes using a standard salting-out technique from blood samples collected at 4 years old. DNA methylation was quantified using a commercial enzyme-linked immunosorbent assay. Methylation analysis was performed in 237 samples of these children aged 3-4 years old, being 145 samples from the control group and 92 from the intervention group, randomly selected. Higher mean global methylation values were found in the intervention group (intervention group mean: $2.199 \pm 1.306\%$; control group mean: $1.649 \pm 1.114\%$; $P = 0.001$). There were no differences between groups for characteristics as gender, ethnicity and maternal smoking, while exclusive breastfeeding duration was higher in the intervention group. Mean anthropometric variables, such as BMI Z-score, weight at birth, 6 months, 12 months, 3-4 years and 8 years old were also compared between groups and no significant differences were observed. Our results indicate that this nutritional intervention in the first year of life, which had the main objective to increase breastfeeding and healthy weaning, increased the levels of global DNA methylation in children at age 4. These evidences, together with previous findings from other authors, suggest that breastfeeding might be associated with DNA methylation in the human genome.

1847F

Type 2 diabetes risk variants affect NOTCH2 activity in the liver. S.

Huang¹, A. Aylward², J. Chiou², M. Okino¹, K. Gaulton¹. 1) Department of Pediatrics, UC San Diego School of Medicine, La Jolla, CA; 2) Bioinformatics and Systems Biology Graduate Program, UC San Diego, La Jolla, CA; 3) Biomedical Sciences Graduate Program, UC San Diego, La Jolla, CA.

Type 2 diabetes (T2D) is a complex disease that arises from a combination of environmental and genetic factors. Genome-wide association studies have identified hundreds of genetic loci influencing risk of T2D and which primarily map to non-coding sequence. The mechanisms of how risk variants affect gene regulation and contribute to T2D risk are largely unknown, prohibiting a deeper understanding of disease pathophysiology. Among T2D loci with unknown mechanisms are risk variants at the *NOTCH2* locus. We found T2D risk variants at this locus were correlated and co-localized with expression of the *NOTCH2* gene in human liver samples from GTEx but not for other genes or tissues. We prioritized candidate causal variants at this locus using fine-mapping data from the MetaboChIP study, liver regulatory elements from the NIH Roadmap and HepG2 transcription factor binding sites from ENCODE. We identified two candidate variants in liver enhancer elements (rs2641317 and rs17186372) and which had allelic effects on gene reporter activity in HepG2 cells ($P < 0.05$ for both). In preliminary data, we performed ChIP-seq to identify transcription factor binding sites of the NOTCH2 protein in HepG2 cells and which passed standard quality metrics (e.g. $\text{ReiCC} = 2.020$). NOTCH2 binding sites overlapped high probability variants at other fine-mapped T2D risk loci including *ADCY5* (PPA=0.23) and *KCNQ1* (PPA=0.84). The most enriched motif in NOTCH2 sites was for TCF-1 ($P = 1e-15$), a factor which regulates the Notch signaling pathway. We are currently employing strategies to further determine the downstream effects of T2D risk variants on NOTCH2 genome binding and target gene regulation such as CRISPR/Cas9 to modify risk elements in liver and *trans* eQTL mapping. Together, these results will help determine the role of NOTCH2 activity in the liver in the pathogenesis of T2D.

1848W

CRISPR-derived deletion of *Crebrf* exon 5 in 3T3-L1 preadipocytes identifies Rab12 as a major CREBRF target regulating mTORC1 activity and autophagy. S.L. Rosenthal¹, F. Liu¹, E.C. Lawrence¹, C.T. Su^{1,2}, Z. Urban¹. 1) Department of Human Genetics, University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA, USA; 2) National Taiwan University Hospital, Yun-Lin Branch and Taipei, Taiwan.

Recently, our group identified a common variant in *CREBRF* that significantly increases both BMI and height in an additive fashion in Samoans. This missense variant, rs373863828 (G>A), results in a substitution of arginine for glutamine. *CREBRF* is a transcription factor, but little else is known about its function. In this project, we sought to identify downstream targets of *CREBRF*. Using 3T3-L1 mouse pre-adipocytes and the CRISPR/Cas9 system, we generated clones with (wild-type) and without (mutant) exon 5 of *Crebrf*, the location of our variant of interest. Sequencing and gel electrophoresis of RT-PCR products confirmed the in-frame deletion of exon 5 in two mutant clones. We selected two wild-type and two mutant clones for RNA sequencing experiments. Cells were grown in duplicates to 8 days post confluence, and total RNA was isolated with DNase treatment to remove contaminating DNA. Following quality control procedures, 2 micrograms of total RNA per duplicate were subjected to mRNA sequencing using an Illumina platform. All data quality control and analyses were performed using CLC Genomics Workbench. 5' and 3' adapters were trimmed, and FastQC reports were generated for both raw and trimmed reads. Reads were aligned to the mouse genome (*Mus musculus*, Ensembl version 80), and principal component analysis showed expected clustering of replicates within clone type. Following alignment, we compared differentially expressed genes (DEGs) between mutant and wild-type clones. After filtering genes with maximum group means ≥ 1 , there were 1,330 significant ($p_{\text{FDR}} \leq 0.05$) DEGs with at least ± 2 fold change compared to wild-type expression levels. The most significant of these genes was *Pi15* ($p_{\text{FDR}} < 1 \times 10^{-15}$, fold change=888 compared to wild-type), a peptidase inhibitor expressed in subcutaneous fat in mice. The gene most differentially expressed based on fold change was *Rab12* ($p_{\text{FDR}} < 1 \times 10^{-15}$, fold change=1,814), a ubiquitously expressed enhancer of the mechanistic target of rapamycin complex 1 (mTORC1) and inhibitor of autophagy. We have shown previously that *CREBRF* is up-regulated by rapamycin treatment and nutrient starvation, and starvation assays illustrate that deletion of exon 5 decreases starvation resistance. These findings implicate *CREBRF*-*Rab12* axis in a feed forward inhibition of the mTORC1 pathway required for cellular adaptation to nutritional deprivation.

1849T

Evaluation of X-chromosome inactivation in female heterozygotes with Fabry disease: Combination of methodical approaches. M. Reboun¹, A. Cerna¹, M. Magner², L. Golan¹, H. Poupetova¹, H. Wiederlechnerová¹, D. Musalkova¹, A. Linhart², L. Dvorakova¹. 1) Department of Pediatrics and Adolescent Medicine, 1st Faculty of Medicine Charles University in Prague and General University Hospital in Prague, Prague, Czech Republic; 2) 2nd Department of Medicine-Department of Cardiovascular Medicine, 1st Faculty of Medicine, Charles University in Prague and General University Hospital in Prague, Prague, Czech Republic.

Fabry disease (OMIM 301500) is a rare X-linked lysosomal storage disorder caused by α -galactosidase A (α -GalA) deficiency due to mutations in the *GLA* gene. Published data evaluating the impact of X-chromosome inactivation (XCI) on the disease manifestation in heterozygous females show controversial results, partly due to the choice of the studied tissues and applied methods. Our aim was to explore the XCI not only by methylation analysis but also with the use of transcriptional XCI ratio examination. 32 heterozygous female Fabry patients from 22 families were involved in our study. The analyses were performed in white blood cells (WBC) and buccal swabs (BS). Methylation analysis was based on inspection of polymorphic repeat region within the genes *AR*, *RP2*, *CNKS2*. To assess RNA-transcriptional allele ratios single nucleotide variants (SNV) were examined using massive parallel sequencing (NexteraXT, Illumina). Three SNV loci were analyzed: causative mutations in *GLA*, and two common polymorphisms in *IDS* (rs1141608) and *LAMP2* (rs12097). *AR*, *RP2*, *CNKS2*, *IDS*, and *LAMP2* assays were informative in 59%, 59%, 22%, 56%, and 31% of cases, respectively. The expression ratio of the wt/mutated allele in the *GLA* gene was evaluated in 84% of cases, variants affecting transcript stability (e.g. splicing variants or large deletions) were excluded from the analysis. Combining the described assays, XCI could be determined for each patient from the cohort, 94% of patients were informative for 2 or more assays. XCI measured by methylation analysis corresponded to results obtained by transcriptional XCI assessment. A good correlation was observed between the XCI status assessed in WBC and α -GalA activity analyzed in leukocytes. Balanced XCI was found in 25 patients in both WBC and BS. However, in 6 patients with skewed XCI (>80%) in WBC, random XCI was observed in BS. *Vice versa*, XCI was skewed in 1 patient in BS but random in WBC. In conclusion, our results show significant tissue specificity of XCI status in the described cohort of patients. These findings strongly complicate the interpretation of XCI in the prediction of the disease severity in female Fabry patients. Grant support: MZ CR – RVO VFN64165, GAUK 580716, SVV No.260367.

1850F

Chromatin accessibility differences between preadipocytes and adipocytes aid identification of regulatory elements at cardiometabolic trait GWAS loci. H.J. Perrin¹, K.W. Currin¹, S. Vadlamudi¹, J.P. Davis¹, M. Wabitsch², K.L. Mohlke¹. 1) University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) University of Ulm, Ulm, Germany.

GWAS have identified hundreds of cardiometabolic trait loci but most lie in noncoding regions with mechanisms that remain poorly understood. At these loci, chromatin accessibility (CA) in relevant cells and contexts can guide identification of regulatory elements and mechanisms. To identify elements that differ between preadipocytes and adipocytes, we performed ATAC-seq in SGBS preadipocytes before (2 replicates) and after 14 days of differentiation (3 replicates). On average, we identified 158,182 CA regions per sample (MACS2 FDR<5%). We examined 20,767 regions that were significantly more accessible (DESeq2, log2 fold change (LFC)>1, FDR<5%) in preadipocytes, and 9,601 regions that were stronger in adipocytes. Genes near preadipocyte CA regions were associated (GREAT P<1E-10) with five GO biological processes including positive regulation of angiogenesis and TGF β receptor, while genes near adipocyte CA regions were associated with twenty processes including fat cell differentiation and triglyceride, acylglycerol, and cholesterol metabolism. Preadipocyte CA regions were enriched (HOMER P<1E-5) for 58 transcription factor motif families including AP, involved in differentiation control and proliferation, and GATA, involved in inhibiting adipogenesis, while adipocyte CA regions were enriched for 49 motifs including GRE, involved in glucocorticoid response, CEBP, involved in differentiation and immune response, and PPAR, involved in differentiation and metabolism. To identify potential regulatory variants at cardiometabolic GWAS loci, we examined 110 loci for which we previously reported a colocalized adipose eQTL. Of 6,692 candidate variants ($r^2>0.8$), 65 variants at 35 loci were located in a region of differential CA. For example, at a GWAS locus for phospholipids colocalized with an adipose eQTL for *SCD* (P=8.4E-23), we observed an adipocyte CA region (LFC=2.72) containing rs603424. This variant exhibited stronger allelic differences in transcriptional reporter activity in 3T3-L1 differentiated adipocytes (FC=2.09, P=0.01) than in SGBS preadipocytes (FC=1.25, P=0.09). These results demonstrate that CA differences between preadipocytes and adipocytes reflect biological cell context and, although further validation is needed, that epigenomic annotation in different cell contexts can guide discovery of regulatory variants at GWAS loci. Identification of CA regions in other disease-relevant contexts could aid identification of additional regulatory mechanisms.

1851W

Down regulation of miR-574-3p associated with low HDLc concentration in obesity patients. R.C.C. Freitas^{1,2}, A.A. Braga^{1,2}, G.M. Bastos², M.E.G. Saldarriaga^{1,2}, R.H. Bortolin^{1,2}, H-T.L. Wang², J.B. Borges^{1,2}, L.R. Castro², M.F. Sampaio², J.I.D. França², R.D.C. Hirata^{1,2}, M.H. Hirata^{1,2}. 1) Faculdade de Ciências Farmacêuticas, Universidade de São Paulo, São Paulo, SP, Brasil; 2) Laboratório de Investigação Molecular em Cardiologia, Instituto Dante Pazzanese de Cardiologia, São Paulo, SP, Brasil.

Obesity is a complex and multifactorial disease. Obesity-related lipid profile contributes to the development of cardiovascular risk. A better understanding epigenetic mechanism involved in obesity-induced dyslipidemia may contribute for new drugs targets discovery. The association of plasma miRNA profile and metabolic and inflammatory variables was investigated in Brazilian subjects. Non-obese (n=47), overweight (n=84), obesity class I (n=68), obesity class II (n=26) and obesity class III (n=8) patients were classified according to body mass index (BMI) values. Biodemographic information was recorded and blood samples were obtained for clinical laboratory testing and miRNA profile analysis. Seven miRNAs (miR-542-5p, miR-424-3p, miR-326, miR-421, miR-301a-3p, miR-183-5p and miR-574-3p) previously related to adipocyte metabolism and/or lipotoxicity and/or inflammation and/or adipogenic differentiation were analyzed by RT-qPCR. Triacylglycerols, VLDLc, leptin, IL-6, TNF- α plasma concentrations, as well as waist-to-hip ratio (WHR) and body fat (%) values were higher in obese patients than non-obese (p<0.05). Lower concentrations of HDLc and apolipoprotein AI (ApoAI) were also observed in the obese group (p<0.05). Plasma miR-574-3p and miR-183-5p were down regulated in obese class III patients compared to non-obese patients (p<0.05). miR-574-3p expression was positively correlated with HDLc (r=0.169, p=0.020) and ApoAI (r=0.159, p=0.028) concentrations. Linear regression analysis showed that the miR-574 expression explains variability in 3.9% of HDLc in obese patients (R²=0.039, p=0.007). A negative correlation was observed between the miR-574 expression and BMI (r=-0.168, p=0.020), hip (r=-0.187, p=0.010) and waist circumference (r=0.179, p=0.014), body weight fat (r=-0.174, p=0.017) and leptin (women, r=-0.222, p=0.022). The association between downregulation of miR-574-3p and obesity-related traits and low concentrations of HDLc and ApoAI is suggestive of a potential role of miRNA in the pathophysiology of obesity and obesity-related complications. Despite these previous evidences, further functional studies are necessary, and the found may to represent a tool to be explored, in order to better understand the association of epigenetic mechanisms and obesity.

1852T

Identification of a transcriptional metabolic hub in adipose tissue using a *cis* mediated *trans* eQTL analysis. D.Z. Pan^{1,2}, Z. Miao^{1,2}, K.M. Garske², A. Ko^{2,3}, Y.V. Bhagat², J.S. Sinheimer^{2,4}, K.L. Mohlke⁵, M. Laakso⁶, P. Pajukanta^{1,2,3}. 1) Bioinformatics IDP, UCLA, Los Angeles, CA, CA; 2) Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, USA; 3) Molecular Biology Institute at UCLA, Los Angeles, USA; 4) Department of Biomathematics, David Geffen School of Medicine at UCLA, Los Angeles, USA; 5) Department of Genetics, University of North Carolina, Chapel Hill, North Carolina, USA; 6) Department of Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland.

Although a significant proportion of the phenotypic variation in obesity is attributed to genetic variation, the mechanisms underlying this heritable component are still elusive. Even though *cis*-regulation is increasingly understood, there is little insight into long-range gene regulatory networks in human subcutaneous adipose tissue, a clinically relevant tissue for obesity. We hypothesize that a set of *trans*-eQTLs are mediated via expression of *cis*-regulated genes. As long non-coding RNAs (lncRNAs) are implicated in gene regulation, we further hypothesize that lncRNAs explain the effects of these *cis*-mediated *trans*-eQTLs. Using the subcutaneous adipose RNA-seq data (n=335) from the Finnish METSIM cohort, we performed weighted gene co-expression network analysis and identified network modules correlated with body mass index (BMI) (adjusted $p < 0.05$). We selected lncRNAs from these BMI-correlated modules and identified their adipose *cis*-eQTLs, with each *cis*-eQTL (+/-1 Mb from TSS, FDR<0.05) and its target lncRNA replicated in subcutaneous adipose RNA-seq data (n=450) from GTEx. Next, we selected the *cis*-eQTLs which are also *trans*-eQTLs (FDR<0.05). The key identified set includes eQTL variant, rs13120966, in the promoter of the *cis*-gene *LINC00989*, which is also associated with the expression level of the *trans*-gene, *AC005225.2*. The *trans*-gene is an antisense transcript sharing 2 exons with *ACOT6*, a peroxisomal lipid metabolism gene. An additional 353 genes are correlated (genome-wide adjusted $p < 0.05$) with the expression levels of the *cis*-gene and *trans*-gene. These genes are enriched for metabolism-related KEGG pathways (adjusted $p < 0.05$), including those for biological processes involved in lipid metabolism (i.e. fatty acid beta oxidation). In addition, 162 of the 353 genes are enriched for lipid metabolism annotations associated with the 212 genes in lipid GWAS loci using the ToppGene suite (PMID 24097068, $p < 5.00 \times 10^{-324}$). Furthermore, we performed mediation analysis between the eQTL variant, *cis*-gene, and *trans*-gene, which resulted in a significant *cis*-mediated *trans*-eQTL effect (mediated effect=0.096, residual direct effect=0.197; $p=0.001$), suggesting a causal path from SNP rs13120966 to *AC005225.2*, mediated by the expression of *LINC00989*. In conclusion, our data from a BMI-correlated adipose network identified a novel *cis*-mediated *trans* gene that is significantly correlated with expression of 353 metabolically enriched genes in human adipose tissue.

1853F

The long intergenic, non-coding RNA, *OLMALINC*, regulates lipid metabolism by regulating *SREBP2* and *SCD1*. J.N. Benhammou^{1,2}, M. Alvarez², A. Ko^{2,3}, C. Rankin², D. Padua², K. Garske², Y. Bhagat², D. Kaminska⁴, J. Pihlajamäki⁴, J. Pisegna^{1,2}, P. Pajukanta^{2,3}. 1) Vache and Tamar Manoukian Division of Digestive Diseases, University of California Los Angeles, CA; 2) Department of Human Genetics, University of California Los Angeles, CA; 3) Molecular Biology Institute at University of California Los Angeles, CA; 4) Clinical Nutrition and Obesity Center, Kuopio University Hospital, Kuopio, Finland.

Background: The metabolic syndrome (MetS) has reached epidemic proportions in the US with a paralleled increase in nonalcoholic fatty liver disease (NAFLD), which can progress to nonalcoholic steatohepatitis (NASH), fibrosis, and cirrhosis. The pathophysiology of NAFLD includes both genetic and environmental factors. Long intergenic noncoding RNAs (lincRNAs) have emerged as regulators of disease processes. We performed RNA-sequencing (RNA-seq) on 259 liver samples from morbidly obese Finnish patients undergoing bariatric surgery, with the majority having a diagnosis of NAFLD/NASH. Using weighted gene co-expression network analysis (WGCNA), we identified 75 genes associated with statin use and fibrosis (Ko et al. submitted). Among those, we observed a lincRNA, *OLMALINC*, and hypothesize that it regulates cholesterol and lipid pathways. **Methods:** To confirm whether *OLMALINC* is a sterol and statin-responsive gene, we measured its expression together with cholesterol and lipogenic genes under sterol-rich and sterol-depleted conditions with statins by RT-PCR in HepG2 liver cells. To determine which genes are regulated by *OLMALINC*, we used transient knock-down with siRNA and performed RNA-seq. Inversely, we also over-expressed *OLMALINC* by generating activating CRISPR dead Cas9 (aCRISPR dCas9) cell lines and measured target gene expression by RT-PCR. **Results:** *OLMALINC* expression increases in sterol-depleted conditions with statins, consistent with up-regulation of the *SREBP2* gene pathway. RNA-seq after *OLMALINC* knock-down demonstrates the upregulation of *SREBP2*-dependent genes by up to 2-fold (FDR<0.05). Endogenous over-expression of *OLMALINC* using the aCRISPR dCas9 cell line demonstrates an increase in *SREBP2* and *SCD1* expression. Promoter capture Hi-C data from our lab and public data suggest that the *OLMALINC* promoter/enhancer region regulates *SCD1* by interacting in *cis* with the *SCD1* promoter. **Conclusions:** *OLMALINC* is a lincRNA we identified to be co-expressed with 74 genes responding to statin use in the liver RNA-seq data from obese patients. We show that *OLMALINC* expression is sterol- and statin-sensitive. Its knock-down results in an increase in the cholesterol biosynthesis pathway genes, while its over-expression also increases *SREBP2* and *SCD1* gene expression. Taken together our human liver RNA-seq and experimental data suggest that *OLMALINC* regulates *SCD1* in *cis* at the DNA and *SREBP2* at the mRNA level, and identify *OLMALINC* as a new statin-responsive gene.

1854W

Identification of functional susceptibility variants in non-coding regions of genome and exploration of their regulatory mechanisms for asthma. T.L. Yang, Y.Y. Duan, M. Yang, H. Chen, S.S. Dong. Key Laboratory of Biomedical Information Engineering of Ministry of Education, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, Shaanxi, China.

Genome-wide association studies (GWASs) have revealed numerous single nucleotide polymorphisms (SNPs) associated with asthma. However, over 90% of these SNPs are located in non-coding regions, leaving big challenge to the interpretation of asthma genetic mechanism. Therefore, we aimed to identify the potential functional noncoding variants and interpret the functions and mechanisms of these variants by integrating a series of bioinformatics analyses and functional assays. Firstly, 89 asthma associated SNPs ($P \leq 5 \times 10^{-6}$) were obtained from GWAS Catalog database, through LD analysis and ANNOVA annotation, we found 1882 SNPs were residing in the noncoding region. By performing eQTL analysis in whole blood and lung tissue in Genotype-Tissue Expression project (GTEx) and another peripheral whole blood sample, we found that 482 SNPs might affect the expression of 21 genes. We further used the Hi-C data of IMR90 cell line and the interaction data of GM12878 cell line to analyze the interactions of the 482 SNPs and their affected genes. We found that 284 SNPs might interact with the promoter of 16 genes. Especially, 82 SNPs located in the 2q12.1 region were associated with expression of 3 important genes, *IL18R1*, *IL1RL1*, and *AC007278.3*, which are related to asthma. By performing epigenetic annotation of the SNPs in the 2q12.1 region, we found that 9 candidate functional SNPs, namely, rs72998585, rs11692065, rs3771180, rs72823646, rs13431828, rs56179005, rs3771166, rs2293223, rs2272127, are located in the regulatory region in asthma related cell lines. Dual luciferase reporter assay showed that 6 SNPs has the allele specific regulatory activity, which are rs11692065, rs3771180, rs72823646, rs13431828, rs56179005, rs2272127. By performing CRISPR/Cas9 in A549 cell line to knock out the regions where rs3771180 and rs56179005 located in, and the results showed that the knock-out of the two regions lead to significant decrease of the mRNA level of the target gene *IL18R1*, and *AC007278.3* respectively. We further analyzed the transcription factors (TFs) that can be allele-specific binding to the 6 SNPs, and detected the activity of the reporter gene after knock-down of the TFs in A549 cell line. We found that the different allele of SNP rs2272127 showed different regulatory activity via alteration of TF MLX binding. Our finding may illuminate the dark road of regulatory mechanisms of noncoding variants leading to underlying the development and progression of asthma.

1855T

Changes in DNA methylation identify response to treatment with methotrexate and TNF inhibitors among rheumatoid arthritis patients. C. Adams¹, K. Marker¹, D. Quach¹, H. Quach¹, M. Krueger², L.F. Barcellos¹, L.A. Criswell¹. 1) Genetic Epidemiology and Genomics Laboratory, School of Public Health, UC Berkeley, Berkeley, CA; 2) Departments of Medicine and Orofacial Sciences, University of California, San Francisco.

Epigenetic modifications including DNA methylation are implicated in the development and progression of autoimmune diseases, such as rheumatoid arthritis (RA [MIM 180300]). Exogenous exposures such as smoking, diet, and medication can change DNA methylation and may contribute to immune cell autoreactivity. Methotrexate (MTX) and anti-TNF agents (a-TNFs) are effective treatments for RA. It is hypothesized they reverse epigenetic modifications which cause T-cell autoreactivity, improving RA symptoms. We conducted a study to identify DNA methylation profiles that may serve as biomarkers of response to treatment with MTX and a-TNFs. Blood samples, clinical data, and disease severity scores (DSS) were collected from 30 treatment naive subjects with RA at baseline and after 3-6 months of treatment with: MTX (n=10), a-TNFs (n=10), or MTX and a-TNFs (n=10). DSS included the Simple Disease Activity Index and Clinical Disease Activity Index. Genome-wide methylation profiles were generated using Illumina's Infinium Human MethylationEPIC BeadChip from PBMC samples. Quantile normalization and background subtraction with dye-bias normalization were performed using minfi and CpGs with high detection p-values, cross-reactive probes, and CpGs measuring SNPs were excluded. Treatment response was defined as difference in DSS from baseline to post treatment measurements. Differentially methylated positions (DMPs) and differentially methylated regions (DMRs) associated with, i) pre-post treatment, and ii) treatment response, were identified with limma and bumpHunter, respectively, using surrogate variables to adjust for confounders such as blood cell proportions, batch effects, and genetic ancestry. Next, baseline methylation levels in the top 1000 DMPs and 100 DMRs (by p-value) were used to predict observed treatment response. Principal component analysis (PCA) was done with the top DMPs and DMRs to identify clusters of pre- and post-treatment and "good" and "poor" response to treatment. Preliminary results found 38 DMPs associated with pre/post-treatment ($p < 0.0001$). PCA found distinct clusters for study subjects pre- and post-treatment. Additionally, subjects clustered by treatment arm using baseline methylation only. These results indicate that treatment with MTX and a-TNFs alters DNA methylation and changes in DNA methylation are potential biomarkers of treatment response. Further research is needed with larger sample sizes and diverse study subjects.

1856F

Genetic regulation of *FCGR2A* expression and its implications for immunity and disease. J. Dahlqvist¹, C. Fulco^{1,2}, J. Ray¹, C. de Boer¹, T. Eisenhaure¹, N. Hacohen^{1,3}. 1) Broad Institute of Harvard and MIT, Cambridge, MA; 2) Department of Systems Biology, Harvard Medical School, Boston, MA; 3) Center for Cancer Research, Massachusetts General Hospital, Department of Medicine, Boston, MA.

Antibody-antigen immune complex (IC) formation is a common phenomenon in autoimmune disorders, such as SLE, and antibody-based treatment of malignancies. Fc gamma receptors (FCGR's) are significant factors in eradication of IC's, a capacity that is crucial in respect to clinical outcome of disease and therapy. We hypothesized that expression levels of specific FCGR's are important determinants of IC depletion, and might influence disease pathogenesis, and hence set out to investigate the regulation of the SLE-associated *FCGR2A* gene. We analyzed the FCGR locus for open chromatin using ATAC-seq and followed up on open regions using CRISPRi and CRISPR screens, identifying regulatory elements of importance for *FCGR2A*. Candidate transcription factors of importance for *FCGR2A* were identified and analyzed for common pathways. Genetic analysis of the putative promoter/enhancer regions raised hypotheses about disease-associated regulatory variants. Taken together, these results suggest a relationship between immune environment and genetic variation with an effect on *FCGR2A* levels and thus a role for this receptor in the complex networks underlying autoimmune disease.

1857W

Functional dissection of cis-regulatory variants associated with autoimmunity. A. Jajodia¹, N. Javelle¹, A. Mishra¹, R.D. Hawkins^{1,2}. 1) Medical Genetics and Genome Sciences, University of Washington, Seattle, WA; 2) Institute for Stem Cell & Regenerative Medicine.

Variants from genome-wide association studies (GWAS) are often significantly enriched at functional non-coding regions. The current study focuses on the dissection of disease-susceptibility variants that influence enhancer function and identification of their target genes. We retrieved associated variants and those in linkage disequilibrium from 19 and 32 GWAS for type 1 diabetes (T1D) and rheumatoid arthritis (RA), respectively. We utilized H3K4me1 ChIP-seq data for enhancers from primary T helper cells to identify a total of 1832 enhancer variants. We measured the allelic effect on enhancer activity using massively parallel reporter assays (MPRA) in human, activated CD4+ T cells. We identified over 250 enhancer variants with a statistically significant difference in allele-specific expression of the reporter gene. Interestingly, 47%(T1D) and 30%(RA) of validated variants demonstrate an increase in enhancer activity relative to the reference allele. Further analysis suggests that functionally validated variants disrupt binding of transcription factors critical for CD4+ T-cell regulation and differentiation, including GATA and STAT factors, IRF2, and FOXO, as well as understudied factors like MEF2A and HNF4. Quantitative changes in predicted TF binding correlate with allele-specific enhancer activity. We leveraged promoter capture HiC and eQTL analysis for activated CD4+ T cells and whole blood, respectively, to identify 20(T1D) and 108(RA) target genes for validated enhancer variants, including known genes implicated in each disease. Pathway analysis of target genes identifies regulation of T-cell receptor signaling, T-cell activation, and inflammation-mediated cytokine signaling to be overrepresented. For RA, we identified important target genes, such as hypoxia-inducible factor-1 α (HIF1A), a promising target for autoimmune disease therapies; and CD6 a potential immunotherapeutic target. For T1D, target genes include PTPN2, a tyrosine phosphatase that modulates JAK/STAT signaling; and SOCS1, a negative regulator of cytokine signaling. We find that multiple validated enhancer variants can interact with a common target gene. This suggests that a genetic interaction may be mediated through physical looping interactions. A subset of validated enhancer variants also interacts with multiple target genes. This study provides a comprehensive view of disease-associated regulatory variants and a deeper understanding of the genetic complexity of autoimmunity.

1858T

Genetic control of transcription factors in immune cells. *N. Soranzo*^{1,2}, *S. Watt*¹, *L. Vasquez*¹, *K. Walter*¹, *A. Mann*¹, *B.-M. Javierre*³, *K. Kundu*², *L. Chen*⁴, *P. Fraser*⁵, *M. Frontini*², *T. Pastinen*⁶, *M. Spivakov*⁷. 1) Human Genetics, Wellcome Sanger Institute, Hinxton, United Kingdom; 2) Haematology, University of Cambridge, Cambridge, United Kingdom; 3) Josep Carreras Leukaemia Research Institute, Barcelona, Spain; 4) Sichuan University, Sichuan, China; 5) Florida State University, United States; 6) Children's Mercy, Kansas City, United States; 7) MRC LMS, London, United Kingdom.

Results from genome wide association studies have revealed that the majority of causal genetic variants are located within intergenic regulatory regions. Recent studies have shown that DNA sequence variation is a major source of variability in chromatin states and gene expression within human populations. However, the extent to which this variability is mediated through DNA binding proteins and long-range DNA interactions remains largely unknown. We aimed to gain insights into how intergenic genetic variation perturbs transcription factor binding and its effect upon the downstream transcriptome. In this study we consider common genetic variation as a perturbation assay. For our model system we used primary immune cells; neutrophils and monocytes collected from healthy blood donors as part of the Blueprint Human variation project. We performed quantitative trait loci (QTL) analysis on transcription factors; PU.1 and CEBP β , key proteins involved in the differentiation of the myeloid lineage and CTCF associated with chromatin architecture. In addition we generated two hQTL datasets for the histone modifications; active H3K4me3 and repressive H3K27me3. Finally, to map enhancer-gene long-range interactions in both cell types we generated Promoter Capture HiC data in six reference individuals with genotypes available. These data sets were integrated with recently published whole genome sequence data, enhancer and transcriptome QTL results. By association and through integration with long-range interaction data, we map genetic variants in distal regulatory elements to their target genes, therefore assessing the contribution of tfQTLs in modulating gene expression. In addition, by studying patterns of cell type specific transcription factor binding, chromatin structure and associated quantitative trait loci we identify genomic regions that may differentially influence functional changes in cellular phenotype, both neutrophil-specific and shared with monocytes. .

1859F

Genome-wide cell-type proportion deconvolution using DNA methylation sequencing in purified and heterogeneous white blood cells. *W.A. Cheung*^{1,2,3}, *X. Shao*^{2,3}, *T. Martin*⁴, *D. Aïssi*⁵, *E. Bouzigon*⁶, *S. Bernantsy*⁷, *M. Lathrop*^{2,3}, *S.G. Wilson*^{1,8}, *J. Bell*¹, *M.F. Moffat*⁹, *W.O.C.M. Cookson*¹⁰, *F. Demenais*⁶, *V. Siroux*⁶, *I. Pin*^{6,11}, *B. Richards*², *D. Manousaki*², *O. Yu*², *R. Biswell*¹, *N. Miller*¹, *T. Pastinen*^{1,2,3}, *E. Grundberg*^{1,2,3}. 1) Children's Mercy Hospital, Kansas City, MO, USA; 2) Dept of Human Genetics, McGill University, Montreal, Canada; 3) McGill University Genome Quebec Innovation Centre, Montreal, Canada; 4) Dept of Twin Research & Genetic Epidemiology, King's College London, London, UK; 5) Inserm U1209, CNRS, University Grenoble Alpes, Institute for Advanced Biosciences, Grenoble, France; 6) INSERM U946, Université Paris Diderot, Université Sorbonne Paris Cité, Paris, France; 7) Division of Clinical Epidemiology, McGill University, Montreal, Canada; 8) University of Western Australia, Perth, Australia; 9) National Heart and Lung Institute, Imperial College London, London, UK; 10) Genomic Medicine, Respiratory Sciences, Asmarley Centre for Genomic Medicine at the National Heart and Lung Institute, Imperial College London, London, UK; 11) Pédiatrie, Centre Hospitalier Universitaire (CHU) Grenoble Alpes, Grenoble, France.

Cellular heterogeneity is a critical factor in genome-wide high-throughput functional sequencing analyses. Variable CpG methylation has been previously applied to HumanMethylation450 BeadArray data to estimate cell proportions, however, only limited direct biological validation has been performed and methods have not been adapted for next-generation sequencing (NGS) data. Deconvolution of cell-type proportions is critical in association studies, separating effects due to the gain or loss of specific cell-types from actual functional changes. Combining 219 genome-wide and capture methylation sequencing samples in 20 cell purification datasets, to 1215 heterogeneous whole blood samples in 7 datasets, with matching cell proportions from blood lab tests, we present the first genome-wide identification and validation of CpG methylation sites predicting cell-type proportion using sequencing. We use a combination of genome-wide (up to 26M CpG sites) as well as targeted (up to 6M sites) methylation sequencing datasets. Using these CpGs, we identify sites showing extreme methylation rate (mean methylation < 20% or > 80%) in one cell-type while having the opposite extreme methylation rate (> 80% or < 20%, respectively) in all other available cell types. We note improved correlation of hypomethylated CpGs with measured blood cell proportions of neutrophils (up to $R=-0.93$, $n=2,756$), lymphocytes ($R=-0.90$, $n=2883$), and monocytes ($R=-0.81$, $n=4944$). Find hypermethylated CpGs to be informative for lymphocyte proportions showing almost perfect correlations ($R=0.97$, $n=2459$). Initially, putative eosinophil-derived hypomethylated CpGs ($n=1133$) were not predictive of cell proportions in blood samples but instead correlated to the neutrophil ($R=-0.89$) cell proportions. However, by refining our analysis using whole blood methylation data, we identify a subset of 51 eosinophil-specific hypomethylated CpGs ($R=-0.80$). Finally, we also show that the performance of our predictions directly relates to available sequencing depth – for the most highly correlated neutrophil-specific CpG we observe $R=-0.56$ at min. 10X coverage increasing to $R=-0.98$ at min 90X coverage. In conclusion, we show how methylation at cell-type specific CpG sites can deconvolute white blood cell proportions. Our method leverages data from purified and heterogeneous white blood cell samples, and allows for proper accounting of differing cell-type proportions among samples, to identify true changes in the methylome. .

1860W

Slipping through the cracks: Migration through a small pore disrupts inactive chromatin organization in neutrophil-like cells. *E. Jacobson¹, J.K. Perry¹, M.H. Vickers¹, D.S. Long², J.M. O'Sullivan¹.* 1) Liggins Institute, University of Auckland, New Zealand; 2) Department of Biomedical Engineering, Wichita State University.

Mammalian cells are flexible and can rapidly change shape when they contract, adhere, or migrate. Their nucleus must be stiff enough to withstand cytoskeletal forces, but flexible enough to remodel as the cell changes shape. This is particularly important for cells migrating through a constricted space, when the nuclear shape must change in order to fit through. This occurs many times in the life cycle of a neutrophil, which must protect its chromatin from damage and disruption associated with migration. We hypothesize that chromatin reorganization during migration with constriction has a role in regulating transcriptional programs during the immune response. Neutrophil-like HL60 cells are 7µm in diameter and were migrated through pores that required constriction (5µm) or not (14µm), and collected within 30 minutes of migrating. Total RNA-seq was performed in triplicate, and differentially expressed genes were identified (FDR<0.05). The transcriptional changes associated with migration with or without constriction, compared to control, were enriched for inflammation and chemotaxis gene ontology terms (FDR<0.1). Differentially expressed genes specific to migration with constriction were enriched for cytoskeletal remodeling (FDR<0.1). Hi-C was used to compare genome organization in control and migrated cells. Chromatin did not switch between the active (A) and inactive (B) compartments after migration. There was a global depletion of mid- to long-range contacts after migration with constriction. Disrupted contacts (FDR<0.1) at this range were enriched in compartment B (p<0.0001), and contained lowly transcribed genes (p<0.05). We propose a model whereby mid- to long-range contacts are preferentially lost within inactive chromatin, thus protecting transcriptionally active contacts from the disruptive effects of constriction. This is consistent with current hypotheses implicating heterochromatin as the mechanoresponsive form of chromatin. Under this model, the lobed nucleus increases the force dispersion due to its higher surface area, providing a possible explanation for neutrophils' unusual nuclear shape. The protection vs flexibility compromise of nuclear stiffness is a problem for all migratory cells. Further investigating the contributions of heterochromatin to stiffness, flexibility, and protection of nuclear function is important for understanding cell migration in processes including brain development, bone healing, metastasis, and the immune response.

1861T

Immune disease variants modulate gene expression through enhancers specific to CD4 regulatory T cells. *L. Bossini-Castillo¹, D. Glinos¹, N. Kunowska¹, G. Golda¹, A. Lamikanra², D. Roberts², G. Trynka¹.* 1) Wellcome Sanger Institute, Hinxton, Cambridgeshire, UK; 2) John Radcliffe Hospital, Oxford, Oxfordshire, UK.

Despite hundreds of variants associated to immune-mediated diseases, little is known about the mechanisms through which these variants lead to disease. Previously we showed that immune SNPs are enriched in promoters and enhancers active in CD4 regulatory T cells (Tregs). Tregs play a critical role in controlling the immune response, however because of their very low numbers, less than 0.001% of whole blood cells, there are no genomic resources available for this cell type. To understand the role of immune-mediated disease variants in Treg function, we isolated Tregs from 100 healthy blood donors and generated a comprehensive resource of genomic assays for SNP fine-mapping (RNA-seq, ChIP-seq for H3K4me3 and H3K27ac, and ATAC-seq). We detected genetic effects on expression of 5,460 genes (eQTL), 3,543 accessible chromatin regions (caQTLs), 3,297 promoter regions (promQTLs) and 7,180 active regions (actQTL) (at FDR < 0.05). Over 35% of the eQTLs and over 55% of the actQTLs were Treg specific and not observed in naïve CD4 T cells from the BLUEPRINT project. We observed 69 immune-mediated disease signals colocalizing with Treg eQTLs, three times more than previously observed in macrophages. This allowed us to fine-map the association signal at 33 loci down to 134 potential functional variants. Twenty seven signals that colocalized with eQTL variants were also acting as actQTLs and caQTLs and directly resided within peaks. For example, one of the inflammatory bowel disease (IBD) associations colocalized with 47 SNPs with eQTL effect on the *BACH2* gene, three of these variants acted as actQTLs while one, rs62408233, also controlled chromatin accessibility. This suggests that the IBD protective allele of the rs62408233 may disrupt an active enhancer resulting in reduced chromatin accessibility, reduced enhancer activation, and consequently decreased *BACH2* expression. We observed that the Treg eQTL genes colocalizing with immune variants were significantly less tolerant to loss of function mutations (median pLI=0.26) compared to the non-colocalizing eQTL genes mapping to GWAS loci (median pLI=0.06). Our findings strongly implicate that a proportion of immune variants regulates gene expression through modulating activity of cell type specific enhancers, building evidence for the causal role of Tregs in the development of immune-mediated diseases.

1862F

DNA methylation signatures of Crohn's disease largely a consequence of inflammation. H.K. Sominen¹, S. Venkateswaran¹, V. Kilaru², U.M. Marigorta³, K. Mondal¹, D.T. Okou¹, J.S. Hyams⁴, L.A. Denson⁵, D.J. Cutler⁶, G. Gibson³, K.N. Conneely⁶, A.K. Smith², S. Kugathasan^{1,6}. 1) Department of Pediatrics, Emory University School of Medicine & Children's Healthcare of Atlanta, Atlanta, GA; 2) Department of Gynecology & Obstetrics, Emory University, Atlanta, GA; 3) Center for Integrative Genomics, Georgia Institute of Technology, Atlanta, GA; 4) Division of Digestive Diseases, Hepatology, and Nutrition, Connecticut Children's Medical Center, Hartford, CT; 5) Division of Pediatric Gastroenterology, Hepatology, and Nutrition, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 6) Department of Human Genetics, Emory University, Atlanta, GA.

Background: Crohn's disease (CD) is a life-long relapsing-remitting disorder with flares of inflammation and a heterogeneous clinical course. A subgroup of patients with CD progress from an inflammatory phenotype (B1) to stricturing (B2) behavior. DNA methylation differences have been reported in CD but their role remains unelucidated; thus, it is critical to investigate the temporal relationship between methylation and disease to establish whether the methylome plays a causal role and can be leveraged for therapeutic benefits. **Methods:** Utilizing a subset of subjects recruited in the pediatric RISK study we generated genome-wide DNA methylation data using the Illumina HumanMethylationEPIC 850K array in blood DNA samples of 74 controls and 164 newly diagnosed CD cases (B1), of which 55 progressed to B2 within 36 months from diagnosis. We used genetic association and the concept of Mendelian randomization to test for causal relationships between methylation changes and CD. **Results:** We found differential methylation at 1189 CpGs associated with CD at diagnosis (FDR < 0.05). Estimated effects of methylation changes at these CpGs showed a strong correlation with plasma C-reactive protein levels, a marker of inflammation ($R = 0.91$; $P < 2.2 \times 10^{-16}$). When we performed within cases analysis comparing their methylation profiles at diagnosis to follow-up, the sign of the effect has reversed, while the magnitude of the change has remained the same at nearly all 1189 sites ($R = -0.93$, $P < 2.2 \times 10^{-16}$). We noted similar results even after stratifying patients based on disease progression to B2 ($R = -0.91$, $P < 2.2 \times 10^{-16}$ for progressors; $R = -0.90$, $P < 2.2 \times 10^{-16}$ for non-progressors). We further demonstrated that, while some methylation changes associated with CD might be causal ($n = 6$ CpGs), the vast majority are a transient consequence of inflammation and thus a result of disease. **Conclusion:** We show convincing evidence that the methylation signatures of inflammation appear to primarily be a symptom of the disease rather than a cause, suggesting that treatment of inflammation signatures may fundamentally be treating the symptoms of CD rather than the cause. This may at least partially explain why CD often remains a life-long relapsing and relapsing disorder, despite effective treatment of the inflammation symptoms.

1863W

Cis regulatory variation determines dynamic HLA-DQB1 allelic expression in response to T cell activation at RNA and protein levels. M. Gutierrez-Arcelus^{1,2}, Y. Baglaenko^{1,2}, S. Hannes^{1,2}, J. Arora³, N. Teslovich^{1,2}, Y. Luo^{1,2}, K. Slowikowski^{1,2}, H.J. Westra^{1,2}, D.A. Rao¹, J. Hermann¹, A.H. Jonsson¹, M.B. Brenner¹, S. Raychaudhuri^{1,2,4}. 1) Division of Genetics and Rheumatology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; 2) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA. 3) Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, 60 Fenwood Road, Boston, MA 02115, USA; 4) Max Planck Institute for Evolutionary Biology, Plön, Germany; 4) Faculty of Medical and Human Sciences, University of Manchester, Manchester, UK.

HLA class II genes contribute the greatest risk for autoimmune diseases, and *cis* regulatory HLA effects might play a role in CD4+ T cells. Most of the non-MHC causal variants for autoimmune diseases affect CD4+ T cell gene regulation (Trynka *et al.* 2013). Hence, understanding the genetic regulatory effects acting at CD4+ T cell states is critical to define autoimmune mechanisms. Since allelic expression is largely driven by *cis* genetic regulatory variation (Buil *et al.* 2015), it can be exploited to define regulatory effects in multiple cell states without the need of a large sample size. We queried CD4+ memory T cells with RNA-seq after non-antigenic stimulation at 8 time points over 72 hours in 24 individuals. We looked for cell-state specific *cis* regulatory effects by identifying **allele specific expression that changes with time**. We tested 224,924 heterozygote sites, and identified 841 sites with dynamic allele specific effects ($P < 2.2 \times 10^{-7}$, Bonferroni), representing 364 unique SNPs spanning 163 genes. Many genes are key in T cell activation (i.e. IL2, LTA, IL7R). Our most striking signal is *HLA-DQB1*. In detailed analyses, where we mapped reads to personalized HLA genomes, we observed 16/24 individuals had significant time-dependent allelic expression ($P < 1 \times 10^{-18}$, fold change > 2.7). For example, in one 0503/0301 heterozygote individual, the 0503 allele represents 61% of overall *DQB1* counts at 0 hours, and increases to 96% of counts at 72 hours. We observed that the same *DQB1* alleles in different individuals had similar expression patterns ($P < 1 \times 10^{-4}$). **DQB1's dynamic allelic profiles during T cell activation cluster into 3 main regulatory programs.** The putative regulatory SNP (rSNP) driving the program with late activation over-expression is in the promoter. To assess whether these effects translate to protein levels, we applied flow cytometry to quantify HLA-DQ surface expression. We observed that the **protein levels recapitulate allelic RNA programs with a delay in time**. That is, at 7 days following stimulation, the median HLA-DQ protein levels were significantly higher in individuals homozygous for the highly expressing rSNP allele compared to the homozygous for the lowly expressing allele ($P = 0.028$). This study brings **two main contributions**: 1) we identified **163 genes with time-dependent allelic expression**, 2) We discovered and characterized **dynamic cis regulatory variation for DQB1, a major autoimmune disease gene, at both RNA and protein levels.**

1864T

Systematic identification of DNA regulatory elements for *GATA3* in human T cells. *H. Chen, P. Fiaux, I. Luthra, G. McVicker.* Salk Institute, La Jolla, CA.

GATA3 is a zing-finger transcription factor that is a master regulator of T helper Type 2 cell differentiation. The dose and timing of *GATA3* expression are controlled precisely throughout T cell development, and are critical for type 2 immune responses that are required for defense against extracellular pathogens such as helminthes. Dysregulation of type 2 immune responses leads to allergy and asthma, and potentially may be caused by improper expression of *GATA3*. Yet little is known about how *GATA3* expression is controlled in the human genome. While some *GATA3* regulatory elements have been identified in the mouse genome, these regions are not evolutionarily conserved. To systematically identify human regulatory elements that control *GATA3* expression, we used a new CRISPR/Cas9 regulatory screen known as CREST-seq. CREST-seq uses pairs of CRISPR guide RNAs to tile deletions across the genome and subsequently flow sorts cells into different pools based on the expression of a gene of interest. The guide RNAs that are present in each expression pool are then sequenced and tested for enrichment by comparing counts across pools. We performed CREST-seq in Jurkat T cells to systematically identify regulatory elements in a 2MB window centered at *GATA3*. We performed a statistical test for guide pairs that are enriched in the *GATA3* low-expression pool, and identified 44 putative regulatory elements (FDR is smaller than 10%). 26 of these elements fall within the topological-association domain where *GATA3* is located, and 8 elements are located in the introns of *GATA3* and neighboring genes. Some of the elements overlap with H3K27 acetylation signals, however others fall within regions that are devoid of epigenetic marks. The former elements are likely to be enhancers for *GATA3*, while the latter may regulate *GATA3* through different mechanisms. Our data suggest that there are a large number of elements that control the expression of *GATA3* and that its regulation is more complicated than anticipated.

1865F

Characterizing human immune response with implications for understanding autoimmune trait architecture. *D. Calderon¹, M.L.T. Nguyen², A. Mezger^{3,4}, J.V. Ribado³, A. Kathiria³, B. Wu⁵, L.A. Criswell⁶, W.J. Greenleaf^{3,6,7}, A. Marson^{2,7}, J.K. Pritchard^{3,8,9}.* 1) Program in Biomedical Informatics, Stanford University, Stanford, CA; 2) Department of Microbiology and Immunology, University of California at San Francisco, San Francisco, CA; 3) Department of Genetics, Stanford University, Stanford, CA; 4) Laboratory of Molecular Neurobiology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, 17177 Stockholm, Sweden; 5) Rosalind Russell/Ephraim P. Engleman Rheumatology Research Center, University of California at San Francisco, San Francisco, CA; 6) Department of Applied Physics, Stanford University, Stanford, CA; 7) Chan Zuckerberg Biohub, San Francisco, CA; 8) Department of Biology, Stanford University, Stanford, CA; 9) Howard Hughes Medical Institute, Stanford University, Stanford, CA.

The human immune system depends on a balanced interplay among a wide variety of specialized cell types transitioning between resting and stimulated states that can become dysregulated in autoimmune disorders. While there are large-scale efforts to map functional regions of resting immune cells, a limited number of cell types have been characterized in stimulated states. Thus, we collected ATAC and RNA-seq data from 25 cell types under resting and stimulated conditions from the blood of 4 healthy individuals. In general, we observed strong chromatin accessibility and transcriptional changes attributable to cell and stimulation-specific effects. Strikingly, stimulation-specific chromatin responses tended to be broadly shared within cell subsets and even across B and T cell lineages. To identify disease-relevant cell types and quantify the importance of stimulation, we performed a GWAS enrichment analysis on 39 traits with publicly available summary statistics. In addition to validating previously identified disease-relevant cells, we discovered a class of autoimmune traits, including rheumatoid arthritis ($P = 3e-4$) and ulcerative colitis ($P = 2e-3$), with enriched heritability in shared stimulation-response peaks. Finally, emphasizing the value of stimulation effects for clarifying disease mechanisms, we fine mapped rs6927172, which links stimulated T cell-specific chromatin dysregulation in the *TNFAIP3* locus to ulcerative colitis and rheumatoid arthritis. Overall, our results highlight the importance of characterizing immune stimulation and classify a set of complex disorders associated with shared stimulation response.

1866W

lncRNA-HBBP1 promotes the imbalance degradation of α and β globin by ubiquitin-dependent pathway to aggravate anemia in β -thalassemia. J. Ma¹, Q.X. Zhang², N. Jiang¹, J. Zhang¹, X. Du¹, D. Ma¹. 1) Fudan University, Shanghai, China; 2) The Second People's Hospital, Shenzhen, China.

β -thalassemia is a common monogenic life-threatening blood disorder and is caused by mutations of β -globin gene *HBB*. However, different degrees of anemia among β -thalassemia major patients can not completely be explained by *HBB* mutations. We have reported differential expression of lncRNAs and mRNAs among β -thalassemia major, β -thalassemia minor patients, and matched normal controls. What most interests us among so many changed lncRNAs in that study is the expression of hemoglobin beta pseudogene 1 (*HBBP1*, also called lncRNA-HBBP1) of human β -globin gene cluster was significantly higher in β -thalassemia major patients compared with other two groups. This difference was also verified in expanding clinical samples. We up-regulated the expression of lncRNA-HBBP1 in zebrafish and found that the amount of red blood cells and hemoglobin level were observed to be decreased significantly. Then we found that lncRNA-HBBP1 could bind to α and β globin (HBA and HBB) and some proteins of ubiquitin-proteasome system (UBC, UHRF1, UBE2K and UB2V2) in K562 erythroleukemic cells tested with chromatin isolation by RNA purification (ChIRP) technology. The binding of lncRNA-HBBP1 to HBA and HBB was confirmed by RNA binding protein immunoprecipitation (RIP). Through overexpressing and interfering the expression of lncRNA-HBBP1 in K562 cell, we found lncRNA-HBBP1 may degrade HBA and HBB unbalancedly. What's more, using Co-IP, we identified the degradation of HBA and HBB may due to their combination with more Ub in K562 cell with high lncRNA-HBBP1 expression. Subsequently by interfering the expression of above-mentioned three ubiquitin proteins in overexpressed lncRNA-HBBP1 K562 cell, we found HBA and HBB bound to UBC and UBE2K respectively. With the prediction of transcription factors binding to regulatory sequence of lncRNA-HBBP1, the HIF1 α was detected and verified via ChIP-qPCR experiment. We also determined that high expression of HIF1 α may promote the transcription of lncRNA-HBBP1 in K562 cell. Finally, we observed significant negative correlation existed between the expression level of lncRNA-HBBP1 and some hematological parameters (Hb, HCT, MCV, MCH, MCHC and RDW) associated with anemia in large β -thalassemia major whole blood samples by statistical analysis. In Conclusions, more HIF1 α in β -thalassemia major leads to increase lncRNA-HBBP1 which by combining with α/β globin and ubiquitin-proteasome complex, aggravate the imbalance of α and β globin and increase the anemia. .

1867T

miRNAs may contribute to pediatric-onset multiple sclerosis by affecting immune signaling. B. Rhead^{1,2}, X. Shao², J.S. Graves³, H.C. von Buedingen³, T. Chitnis⁴, A. Waldman⁵, T. Lotze⁶, T. Schreiner⁷, A. Belman⁸, B. Greenberg⁹, B. Weinstock-Guttman¹⁰, G. Aaen¹¹, J.M. Tillema¹², J. Hart¹³, J. Ness¹⁴, J. Rubin¹⁵, M. Candee¹⁶, L. Krupp⁸, M. Gorman¹⁷, L. Benson¹⁷, S. Mar¹⁸, I. Kahn¹⁹, J. Rose²⁰, S. Roalstad²¹, T.C. Casper²¹, H. Quach², D. Quach², C. Schaefer^{22,23}, E. Waubant², L.F. Barcellos^{1,2}. 1) Computational Biology Graduate Group, University of California, Berkeley, Berkeley, CA; 2) Genetic Epidemiology and Genomics Laboratory, University of California, Berkeley, Berkeley, CA; 3) Department of Neurology, University of California, San Francisco, CA; 4) Partners Pediatric Multiple Sclerosis Center, Massachusetts General Hospital for Children, Boston, MA; 5) Division of Neurology, Children's Hospital of Philadelphia, Philadelphia, PA; 6) Blue Bird Circle Multiple Sclerosis Center, Baylor College of Medicine, Houston, TX; 7) Children's Hospital Colorado, University of Colorado, Denver, CO; 8) Lourie Center for Pediatric MS, Stony Brook Children's Hospital, Stony Brook, NY; 9) Department of Neurology, University of Texas Southwestern, Dallas, TX; 10) Pediatric Multiple Sclerosis Center, Jacobs Neurological Institute, SUNY Buffalo, NY; 11) Pediatric MS Center at Loma Linda University Children's Hospital, Loma Linda, CA; 12) Mayo Clinic's Pediatric MS Center, Rochester, MN; 13) University of California, San Francisco, Regional Pediatric MS Center, Neurology; San Francisco, CA; 14) University of Alabama Center for Pediatric-onset Demyelinating Disease, Children's Hospital of Alabama, Birmingham, AL; 15) Department of Pediatric Neurology, Northwestern Feinberg School of Medicine, Chicago, IL; 16) University of Utah, Primary Children's Hospital, Salt Lake City, UT; 17) Boston Children's Hospital, Boston, MA; 18) Pediatric-onset Demyelinating Diseases and Autoimmune Encephalitis Center, St. Louis Children's Hospital, Washington University School of Medicine, St. Louis, MO; 19) Children's National Medical Center, Washington, DC; 20) Department of Neurology, University of Utah School of Medicine, Salt Lake City, UT; 21) Department of Pediatrics, University of Utah School of Medicine, Salt Lake City, UT; 22) Kaiser Permanente Division of Research, Oakland, CA; 23) Research Program on Genes, Environment and Health, Kaiser Permanente, Oakland, CA.

Multiple sclerosis (MS; MIM 126200) is a disabling chronic inflammatory autoimmune disease that affects the central nervous system. For approximately 5% of MS cases, onset occurs in individuals less than 18 years old (pediatric MS or "ped-MS"). Epigenetic influences are strongly implicated in MS pathogenesis in adults, including the contribution from miRNAs, small noncoding RNAs that target specific messenger RNAs and block their translation into proteins. SNPs can affect miRNA function in different ways depending on their location. These miRNA-related SNPs (miRSNPs) can impact miRNA expression levels or influence miRNA target interactions; either mechanism can affect disease susceptibility. Our objective was to assess miRNA involvement in the pathogenesis of ped-MS. We tested 52,586 miR-SNPs in 486 ped-MS cases and 1,362 controls drawn from the U.S. Pediatric MS Network and Kaiser Permanente Northern California membership using logistic regression models adjusted for genetic ancestry. miR-SNPs within the major histocompatibility complex (MHC) were also adjusted for the presence of the *HLADRB1*15:01*, an MHC allele that is the strongest known genetic risk factor for MS. Additionally, we conducted a pilot gene expression study and examined levels of 1,979 miRNAs in four sorted peripheral immune cell types from eight treatment-naïve post-pubertal ped-MS female cases and eight female controls matched on age and race/ethnicity. No individual miR-SNPs were associated with ped-MS after accounting for multiple hypothesis tests. However, pathway analysis of top hits ($p < 0.01$) implicated genes involved in immune signaling, specifically in pathways involving the histamine H₁ receptor, the serotonin 5-HT₂ receptor, and interferon gamma, all pathways that have plausible roles in MS pathogenesis. There were 88 miRNAs that were differentially expressed in immune cells of ped-MS cases compared to controls (unadjusted $p < 0.1$), and these miRNAs were enriched for those that have previously been associated with MS in human adult or mouse studies. We conclude that miRNAs may play a role in ped-MS pathogenesis by affecting immune-signaling genes, and that miRNAs relevant to MS are expressed at different levels in ped-MS patients than healthy controls.

1868F

The long non-coding RNA gene *NeST* influences basal immune state, antiviral responses, and ulcerative colitis phenotype in mice. O.M. de Goede¹, N.J. van Buuren², A.T. Satpathy^{3,4}, S.B. Montgomery^{1,3}, H.Y. Chang^{1,4}, K. Kirkegaard^{1,2}. 1) Department of Genetics, Stanford University School of Medicine, Stanford, CA; 2) Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA; 3) Department of Pathology, Stanford University School of Medicine, Stanford, CA; 4) Center for Personal Dynamic Regulomes, Stanford University School of Medicine, Stanford, CA.

In mice, the lncRNA gene *NeST* (or *Tmevpg1*, *Ifng-as1*) influences immune responses to Theiler's virus and *Salmonella* infection, in part through *trans*-regulation of the overlapping gene *Ifng*. Here, we further classify functional effects of *NeST* through a knockout mouse model (*NeST*^{-/-}), and define the changes in gene expression and chromatin accessibility that may be driving these responses. Compared to *NeST*-floxed (*NeST*^{fl/fl}) controls, *NeST*^{-/-} mice had significantly higher baseline numbers of splenic regulatory T cells and T-bet⁺/CD8⁺ T cells, suggesting a more attenuated immune environment. No other baseline changes in immune cell compartments or cytokine levels were observed. RNA sequencing of CD8⁺ T cells from *NeST*^{-/-} and *NeST*^{fl/fl} mice identified 211 differentially expressed (DE) genes, which were enriched for decreased activity in *NeST*^{-/-} mice in pathways related to oxidative phosphorylation, responses to viral infection, and the cell cycle. Targeted examination of immunosuppressive genes revealed several that were upregulated in *NeST*^{-/-} CD8⁺ T cells, including *Tbx21* (T-Bet). ATAC-seq analysis identified 28 differentially accessible (DA) regions in unstimulated CD8⁺ T cells, and 31 DA regions in CD8⁺ T cells following *in vitro* stimulation with PMA/ionomycin. 19 of these DA regions were shared between the unstimulated and stimulated conditions, and 2 overlapped with DE genes. These baseline differences with *NeST* knockout yield dramatic phenotypes in response to immune challenges. *NeST*^{-/-} mice intraperitoneally injected with poliovirus had significantly higher splenic viral titers, reductions in both number and proliferative capacity of CD8⁺ T cells, and lower survival rate. This is consistent with the decreased gene expression in metabolic and cell cycle-related pathways we observed in *NeST*^{-/-} CD8⁺ T cells. Male *NeST*^{-/-} mice also had an exacerbated DSS-induced ulcerative colitis phenotype, with greater decrease in body weight, shorter colon length, and more severe colon histology than their *NeST*^{fl/fl} counterparts. This may be related to a region of increased accessibility in *NeST*^{-/-} CD8⁺ T cells located near *Ap1m2*, an epithelia-specific gene for which reduced expression has been associated with spontaneous colitis in mice and Crohn's disease in humans. While the functions of many lncRNAs are unknown, here we demonstrate that they can have broad immunological impacts influencing both antiviral responses and ulcerative colitis susceptibility.

1869W

Characterizing host-pathogen gene interaction networks using weighted gene co-expression network analysis (WGCNA). A.C. Shetty¹, C. McCracken¹, H. Liu², S.G. Filler^{2,3}, V. Bruno^{1,4}. 1) Institute for Genome Sciences, University of Maryland Baltimore, Baltimore, MD; 2) Division of Infectious Diseases, Los Angeles Biomedical Research Institute, Harbor-UCLA Medical Center; 3) David Geffen School of Medicine at UCLA, Los Angeles, California; 4) Department of Microbiology & Immunology, University of Maryland School of Medicine, Baltimore, Maryland.

Background: There is an increasing interest in using dual-species transcriptomics to deduce potential host-pathogen gene interactions and understand molecular mechanisms of virulence and disease progression. Correlation network analysis is a powerful tool to study such multi-dimensional biological datasets. Weighted gene correlation network analysis (WGCNA) can find clusters of highly correlated genes in transcriptomic datasets. **Hypothesis:** WGCNA applied to the host and pathogen gene networks can be further correlated to identify novel host-pathogen interactions. **Methods:** We performed RNAseq analysis on *in-vivo* murine models of pulmonary aspergillosis in which mice were infected with pathogenic fungus, *Aspergillus fumigatus*. Total RNA, containing mouse and *A. fumigatus* transcripts, was collected from the lungs of infected mice at 2, 4, or 6 days post infection (DPI) in biological triplicate. RNAseq libraries were generated and also enriched for fungal transcripts using an Agilent SureSelect XT custom bait library designed against *A. fumigatus*. Unenriched and SureSelect captured libraries were sequenced to assay the host and fungal transcriptomes, respectively. Raw reads were aligned to the Mouse mm10 and *A. fumigatus* Af293 reference genomes using a splice aware aligner. We quantified and normalized gene expression values for both the host and pathogen genes. Normalized expression estimates were used to identify highly correlated gene clusters for the host and pathogen genome separately using WGCNA. Gene clusters significantly correlated with DPI (p-value<0.02) were selected. For each cluster, genes correlated with DPI (>0.8) showing high cluster membership (>0.8) were selected for host-pathogen interaction analysis. Using methods in WGCNA, we tested these host and pathogen genes for correlation with pathogen and host gene clusters respectively to identify potential host-pathogen gene coexpression. Finally, highly correlated host-pathogen gene pairs (>0.8) were selected to build interaction networks. **Results:** We observe 322 Mouse and 80 *A. fumigatus* genes correlated with DPI, respectively. Of these, 43 pathogen genes show expression that is highly correlated with the expression of 116 host genes. Gene pairs are visualized as an interaction network. **Conclusions:** Dual-species WGCNA facilitates the formulation of testable hypotheses regarding molecular mechanisms governing the interaction between pathogen and host in the context of pulmonary aspergillosis.

1870T

Deciphering and modelling of the regulatory network controlling dendritic cell differentiation from human monocytes. A. Medina-Rivera¹, K. Nuñez-Reza¹, P. Pozos¹, P. Gonzalez², J.E. Sotelo Fonseca², A. Naldi³, C. Molina-Aguilar⁴, D. Gutierrez-Reyna⁴, S. Spicuglia⁵, M.A. Santana⁴, M. Thomas-Chollier⁵, D. Thieffry⁵. 1) LIIGH, Universidad Nacional Autónoma de México, Querétaro, Querétaro, México; 2) Functional and Evolutionary Genomics of Regulatory RNA, Laboratorio Nacional de Genómica para la Biodiversidad, Irapuato, México; 3) Computational Systems Biology team, Institut de Biologie de l'Ecole normale supérieure – PSL Université, Paris, France; 4) Centro de Investigación en Dinámica Celular (IICBA), Universidad Autónoma del Estado de Morelos, Cuernavaca, México; 5) INSERM, TAGC, Marseille, France.

The aim of this project is to determine the regulatory mechanisms that govern the differentiation of monocytes to dendritic cells (DCs) and integrate the regulatory information in a logical model for DC differentiation. In peripheral tissues, differentiation from monocytes into DCs occurs mostly in inflammatory conditions (Villadangos and Schnorrer, 2007). This differentiation can be reproduced *in vitro* (Segura et al, 2013). Albeit progresses in the field of DC differentiation, we still lack a comprehensive understanding of the underlying signaling pathways and regulatory circuits involved. We analyzed bulk ChIP-seq and RNA-seq of DCs and monocytes from the BLUEPRINT consortium (Stunnenberg et al, 2016). Regarding bulk RNA-seq analysis, we used the methods described in Law et al (2016). Our analysis pipeline for ChIP-seq includes quality control with FastQC tool, along with ENCODE QC and IDR analysis (Bailey et al, 2013). Peak calling was performed using MACS2 (Hung and Weng, 2017). Chromatin states were then defined using ChromHMM (Ernst and Manolis, 2012). RSAT tool peak-motifs was used to identify transcription factor binding sites (Thomas-Chollier et al, 2011). Using the software GINsim (Naldi et al, 2018), we are currently building a logical model of the regulatory network controlling dendritic cell differentiation from monocytes. This model includes the pathways previously described for CSF-2 and IL-4 signalling, taking into account the cross-talks between these pathways. CSF-2 and IL-4 activate the Jak/Stat pathway, leading to activation of the transcription factors STAT3, STAT5, STAT6, IRF4, NFkB2 and CEBa α , all reported to be involved in monocyte to dendritic cell differentiation. The results of ChIP-seq and bulk RNA-seq analysis will enrich our model, taking into account novel transcription factors and regulatory interactions. Model analyses will serve as a screening platform to select components and interaction to assess experimentally (Collombet et al, 2017; Rodriguez et al, unpublished data). DCs play an important role in the activation of the adaptive immune response, making them excellent candidates for cancer treatments and vaccines development. The construction of a predictive dynamical model of monocytes to dendritic cell differentiation should enable a better understanding of this process, to identify clinically relevant intervention points and ultimately design novel therapeutic strategies.

1871F

Inflammatory bowel disease causal SNP rs1887428 regulates JAK2 expression via allele-specific binding of RBPJ and CUX1. C.J. Cardinale¹, A. Gutiérrez Uzquiza¹, L.A. Spruce², S.H. Seeholzer², H. Hakonarson¹. 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Proteomics Core Facility, Children's Hospital of Philadelphia, Philadelphia, PA.

Recently, 18 genetic variants conferring risk for inflammatory bowel disease were fine-mapped to single-variant resolution with greater than 95% confidence. One SNP so identified, rs1887428, is located 600 bp upstream of the transcriptional start site of the JAK2 gene. We used affinity purification-mass spectrometry (AP-MS) with oligonucleotide baits containing either the risk or protective allele of this SNP to purify two transcription factors with allele-specific binding from Jurkat cell nuclear extracts. Notch pathway member RBPJ binds preferentially to the risk allele, while transcriptional repressor CUX1 binds to the protective allele. This binding preference was confirmed *in vitro* using electrophoretic mobility shift assay (EMSA) and affinity purification. We constructed subclones from Jurkat using CRISPR/Cas9 genome editing with homology-directed repair (HDR) to carry either the risk or protective allele of rs1887428, and showed that the risk allele increases expression of JAK2. We used a dual luciferase reporter assay to assess the function of this SNP on the expression of a reporter driven by the 1.2-kb promoter region of JAK2. Depletion of CUX1 with siRNA raised expression of the reporter, as did co-expression of constitutively-active NOTCH1, partner of RBPJ. We conclude that causal noncoding SNPs, even in the absence of canonical transcription factor binding motifs, can be recognized by allele-specific proteins, resulting in differential gene expression.

1872W

Does spacing matter? A comprehensive study on the spatial constraint of transcription factor binding sites in mouse macrophages. Z. Shen^{1,2}, J. Tao^{2,3}, G.J. Fonseca², C.K. Glass^{2,3,4}. 1) Department of Bioengineering, University of California, San Diego, La Jolla, CA; 2) Department of Cellular and Molecular Medicine, School of Medicine, University of California, San Diego, La Jolla, CA; 3) Bioinformatics and Systems Biology Program, University of California, San Diego, La Jolla, CA; 4) Department of Medicine, University of California, San Diego, La Jolla, CA.

Motivation. Transcription factors regulate gene expression by binding to DNA regulatory sequences. In macrophages, lineage-determining transcription factors (LDTFs) bind collaboratively in order to select functional enhancers. Such collaboration relies on the arrangement of transcription factor binding sites that are embedded in DNA sequences, sometimes called “grammar”. Natural genetic variants like single nucleotide polymorphisms (SNPs) and short insertions and deletions (indels) will interfere with the binding of LDTFs by directly affecting the sequence of binding sites of LDTFs or their collaborative transcription factors. Short indels can also alter the spacing between binding sites. Previous studies showed that the altered spacing could significantly affect gene regulation in *Ciona* embryos. Currently, there is no systematic study on the spatial relationships of transcription factors. It remains unknown to what extent the change of spatial relationship will affect the binding events in mammalian cells. **Method.** We used ChIP-seq to measure the binding of macrophage LDTFs, including PU.1, C/EBP beta, and c-Jun in murine bone marrow-derived macrophages. The binding sites were further annotated by searching for highly confident motifs within binding regions. We first revealed different spacing patterns of binding sites among LDTFs and of those between LDTFs and collaborating transcription factors. In addition, we analyzed the effects of altered binding sites and spacing on transcription factor binding by looking at small indels across 5 genetically diverse mouse strains, which have between 0.8 and 4 million indels depending on the pair-wise comparisons. Of the 50,000 binding regions of PU.1 in each strain, comparative studies were conducted between 690 regions with indels that affect LDTF binding sites and 1120 regions with indels that fall between LDTFs, changing the spacing. **Result.** We found that collaborative LDTFs have a tendency to bind close together with certain spatial relationships. Taking PU.1 as an example, we observed positive spatial relationships for 14 factors that include transcription factors previously established to be important for macrophage development and function. However, the spatial relationships are relatively broad. Consistent with this, at a genome-wide level indels that alter transcription factor binding sites are more likely to result in decreased binding than those that alter the spacing between binding sites.

1873T

Identification of genes escaping X-inactivation and the variability of escape across cells, tissues and twin pairs. A. Zito, K. Small. Department of Twin Research & Genetic Epidemiology, King's College London, London, United Kingdom.

To equalize the X-linked transcriptional dosages between the sexes, one X-chromosome is inactivated in somatic tissues of female mammals. However, X-inactivation (XCI) is incomplete; 15-20% of genes escape XCI and maintain X-linked bi-allelic expression. Escape from XCI has high biomedical relevance and may underlie sexual dimorphisms in autoimmunity and cancer incidence, and clinical phenotypes in poly-X karyotypes. The extent to which genetic and environmental factors influence escape, and the extent of variability across individuals and cell types is not well characterized. To address this, we investigated escape from X-inactivation in a multi-tissue sample of deeply phenotyped twins. We quantified escape with a metric (EscScore) that combines a gene's allele-specific expression with the overall XCI skew of the sample, and ranges from 0 to 1, with 0 = no escape (no expression from inactive X) and 1 = complete escape (equal expression from the active and inactive X). We applied our score to skewed samples (XCI-skew > 80%) with paired RNA-seq and WGS data; our dataset included LCLs (N = 145), whole blood (N = 20), fat (N = 45), skin (N = 54) and purified monocytes, B, CD4⁺, CD8⁺ and NK cells from a monozygotic twin pair. Using previously published escape designations for ChrX genes, we find that our score distinguishes between genes designated as silenced (mean EscScore = 0.2, IQR = 0.24), variably escape (mean EscScore = 0.28, IQR = 0.46) and fully escape (mean EscScore = 0.57, IQR = 0.41). We identify 51 genes with novel tissue-specific escape patterns, 27 of which were previously classified as silenced and 24 classified as escapees but with no previous evidence of tissue-specificity. Within the purified immune cells, we identify 25 new escape genes, of which 14 show cell-type specific escape. We see a high degree of variability in EscScore(s) across individuals in a subset of genes, and are investigating whether this is concordant across twin pairs or correlates with phenotypic measures from these individuals. This study comprehensively quantifies escape in multiple immune cell types and tissues in a large cohort of women with skewed XCI patterns. By identifying new putative escapees, we conclude that escape from XCI is a complex and widespread event that can modulate gene expression in a cell-type-restricted manner.

1874F

Epigenome-wide association study for IgA nephropathy in Chinese population.

Y. Lin¹, Z. Zhu¹, P. Yin², Y. Peng², L. Liang¹, X. Yu². 1) Program in Genetic Epidemiology and Statistical Genetics, Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA; 2) Department of Nephrology, The First Affiliated Hospital of Sun Yat-sen University, Guangzhou, China.

IgA nephropathy (IgAN) is the most common primary glomerular disease in China and worldwide. It is recognized that up to 50 percent of IgAN patients might end up with the end-stage renal disease, the etiology and pathogenesis is unclear. Accumulated evidences suggested that genetic factors and environmental factors were both involved in the pathogenesis of IgAN. Epigenetic modifications such as DNA methylation might play a significant role in gene regulation and disease susceptibility. However, few epigenome-wide association studies (EWAS) have investigated the relationship between DNA methylation variation genome-wide and IgAN. In this study, we applied the Infinium Methylation EPIC BeadChip, which interrogates 850,000 CpG sites, on 184 samples of CD19⁺ B cells and 24 whole blood cells from Chinese descent. Data was normalized and adjusted for cell types, age, gender effect and residual batch effects via smart surrogate variable analysis (smartSVA). Linear regression model was used to analyze differentially methylated CpG sites associated with IgAN in both cell types, respectively. We use false discovery rate (FDR) less than 10% to detect significant CpGs in relation to IgAN. After normalization and adjusting clinical covariates and technical surrogate variables, we found that 3 CpGs were significantly associated with IgAN in CD19⁺ B cells samples and the B-cell and whole blood combined analysis. Among them, one CpG was associated with *PLEKHA4*, and the rest CpGs were in close proximity to genes including *CLEC11A* and *C1orf127*. (FDR < 0.1). Specifically, *PLEKHA4* plays an important role in lipid binding and nucleic acid metabolism which may have potentiality to have effect on the genesis of IgAN. Moreover, *CLEC11A* was coding for stem cell growth factor which may have impact on the progressive damage associated with chronic kidney failure from IgAN. These findings suggested that DNA methylation variation might involve in the pathogenesis of IgAN. We will further illustrate the potential functional impact from the identified DNA methylation site and seek for replication in a large samples undergoing sample collection and methylation typing.

1875W

Cyclin-dependent kinase inhibitor 1B (*CDKN1B*) promoter variant and Systemic Lupus Erythematosus (SLE) susceptibility. B. Singh, J. Molineros, J. Kaplan, B. McDaniel, C. Sun, S. Nath. Oklahoma Medical Research Foundation, Oklahoma city, OK.

Background: Systemic Lupus Erythematosus (SLE) is an autoimmune disease with a strong genetic component and ethnic differences. Recently, cyclin-dependent kinase inhibitor 1B (*CDKN1B*) was reportedly associated with SLE risk in Han-Chinese. However, its validity and functionality are not thoroughly assessed. **Objectives:** The objectives of this study were to replicate *SLE-CDKN1B* association and characterization of the SLE-associated variants. **Methods:** First, we performed allelic association followed by a meta-analysis of rs34330 (*CDKN1B*) using ten independent cohorts (n = 34,017). Second, we used *in-silico* bioinformatics and epigenetic analyses to identify potential regulatory effects on gene expression and using data on histone marks, eQTLs across multiple tissues. Third, luciferase reporter assays in HEK293 and Jurkat cells were used to measure the allelic effects. Fourth, combination of DNA pulldown, Electrophoretic Mobility Shift Assay, Western blot, and Mass Spectrometry were used to identify DNA-bound proteins followed by ChIP-qPCR to assess the allele-specific binding of interacting proteins. Finally, mRNA and protein expression of candidate genes were compared between risk and non-risk genotypes. **Results:** We replicated genetic association between SLE and rs34330 ($P_{meta}=1.16 \times 10^{-18}$, OR=0.84). Bioinformatics predicted that rs34330, located in active chromatin, has potential promoter/enhancer effect. Gene regulation through rs34330 was further confirmed through methylation association with a CpG probe (cg27423445) and chromatin interactions. SNP rs34330 is an eQTL and risk allele increased expression of *CDKN1B* and a nearby *APOLD1*. Finally, we observed significant allele-specific promoter activity by luciferase in HEK293 and Jurkat cells, significant allele-specific bindings with histones (H3K27Ac, H3K4Me3, H3K4Me1) and transcription factors (Pol-II & IRF1). **Conclusions:** We replicated rs34330 association with SLE in multiple cohorts. Our results showed that rs34330 is a potentially functional variant, predicted to alter *APOLD1/CDKN1B* expression through its enhancer/promoter effects. Allele-specific histone modifications and binding with regulatory proteins may regulate the expressions of *APOLD1/CDKN1B* proteins, which could be a potential mechanism for exacerbating the susceptibility of lupus through rs34330.

1876T

Whole genome sequencing shows contribution of *de novo* noncoding variants to congenital diaphragmatic hernia. A. Kitaygorodsky^{1,2}, H. Q^{2,3}, J. Wynn⁴, L. Yu⁴, V. Duron⁵, G. Aspelund⁶, R. Herman⁴, F.Y. Lim⁷, T. Crombleholme⁸, R. Cusick⁹, K. Azarow⁹, M.E. Danko¹⁰, D. Chung¹⁰, B.W. Warner¹¹, G.B. Mychaliska¹², D. Potoka¹³, A.J. Wagner¹⁴, S. Soffer¹⁵, M. ElFiky¹⁶, J.M. Wilson^{17,18}, F.A. High^{17,18,19}, M. Longoni^{18,19}, P.K. Donahoe^{18,19}, D. Schinde²⁰, P. Kosiński²¹, Y. Shen^{1,2,22}, W.K. Chung^{4,23,24}. 1) Biomedical Informatics, Columbia University, New York, NY, USA; 2) Department of Systems Biology, Columbia University Medical Center, New York, NY, USA; 3) Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY, USA; 4) Department of Pediatrics, Columbia University Medical Center, New York, NY, USA; 5) Department of Surgery, Columbia University Medical Center, New York, NY, USA; 6) Division of Pediatric Surgery, Department of Surgery, Columbia University, New York, NY, USA; 7) Cincinnati Children's Hospital, Cincinnati, OH, USA; 8) Children's Hospital & Medical Center of Omaha, University of Nebraska College of Medicine, Omaha, NE, USA; 9) Department of Surgery, Oregon Health & Science University, Portland, OR, USA; 10) Monroe Carell Jr. Children's Hospital, Vanderbilt University Medical Center, Nashville, TN, USA; 11) Washington University, St. Louis Children's Hospital, St. Louis, MO, USA; 12) Division of Pediatric Surgery, Department of Surgery, University of Michigan School of Medicine, Ann Arbor, MI, USA; 13) Division of Pediatric Surgery, Department of Surgery, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA; 14) Medical College of Wisconsin, Milwaukee, WI, USA; 15) Northwell Health, Great Neck, NY, USA; 16) Department of Pediatric Surgery, Faculty of Medicine, Cairo University, Cairo, Egypt; 17) Department of Surgery, Boston Children's Hospital, Boston, MA, USA; 18) Department of Surgery, Harvard Medical School, MA, USA; 19) Pediatric Surgical Research Laboratories, Department of Surgery, Massachusetts General Hospital, Boston, MA, USA; 20) Children's Hospital Dallas, Dallas, TX, USA; 21) 1st Department of Obstetrics and Gynecology, Medical University of Warsaw, Poland; 22) JP Sulzberger Columbia Genome Center, Columbia University, New York, NY, USA; 23) Department of Medicine, Columbia University Medical Center, New York, NY, USA; 24) Herbert Irving Comprehensive Cancer Center, Columbia University, New York, NY, USA.

Congenital diaphragmatic hernia (CDH) is one of the most common and lethal birth defects. Many CDH patients have additional congenital anomalies or neurodevelopmental disorders. Although previous studies have identified significant contribution of *de novo* coding variants to both isolated and complex CDH, the impact of mutations in noncoding and regulatory regions is still not well understood. To further investigate the genetics of CDH and to better understand the role of noncoding mutations, we analyzed whole genome sequencing (WGS) from a total of 489 CDH parent-offspring trios and used 1606 parental-unaffected siblings trios from the Simons Simplex Collection as controls. We observed on average 72 *de novo* single nucleotide variants (SNV) or small insertions and deletions (indels) per subject in both cases and controls. We investigated the role of *de novo* noncoding variants by comparing the burden between cases and controls in biologically plausible regions, based on a combination of epigenomic features (histone marks), physical location relative to transcription start sites (TSS) or 3' untranslated regions (3' UTR), and proximity to genes that are highly expressed in developing diaphragm or candidate CDH risk genes from previous human or mouse studies. We observed a 2.1-fold enrichment of indels within H3K4me1 histone modification marks (associated with enhancers or promoters) in stem cells ($p=1.6e-4$). When limited to TSS regions of candidate CDH genes, we further observed an 8.8-fold enrichment of SNPs and indels within H3K4me1 marks in stem cells ($p=7.3e-4$); notably, two of these noncoding variants are close to *GLI3*, an important mediator of Sonic Hedgehog signaling that has been previously implicated in CDH. Finally, we observed an excess of 3' UTR regions with noncoding *de novo* variants in multiple cases ($p=0.02$ by permutation; observed 743, expected 682). In particular, there are 4 noncoding *de novo* variants in 3'UTR of *ZNF408* ($p=6e-5$ based on a background mutation model) and *USP12* ($p=7e-5$), both with plausible roles in development. Overall, the whole genome sequencing data suggests a substantial role of regulatory noncoding *de novo* variants in CDH. Future studies with larger sample size and better assessment or prediction of genetic effect of noncoding variants will help to estimate precisely the contribution of noncoding variants and to discover new high risk genes.

1877F

Epigenomic signatures of disease progression in Wilson disease liver and blood. C.E. Mordaunt¹, D.A. Kieffer², N.M. Shibata³, A. Czlonkowska³, T. Litwin³, K.H. Weiss⁴, C.L. Bowlus⁵, S. Sarkar⁶, S. Cooper⁶, Y.J.Y. Wan⁶, M. Ali⁶, V. Medici⁶, J.M. LaSalle¹. 1) Medical Microbiology and Immunology, Genome Center, and MIND Institute, University of California - Davis, Davis, CA; 2) Internal Medicine, Gastroenterology and Hepatology, University of California - Davis, Davis, CA; 3) Institute of Psychiatry and Neurology, Warsaw, Poland; 4) Internal Medicine IV, Heidelberg University Hospital, Heidelberg, Germany; 5) California Pacific Medical Center, San Francisco, CA; 6) Surgery, School of Medicine, University of California - Davis, Davis, CA.

Wilson Disease (WD) is an autosomal recessive disease caused by mutations in the copper transporter encoding *ATP7B* gene, manifesting as variable phenotypes as a result of copper accumulation. WD presentation is characterized by hepatic, neurologic, and psychiatric symptoms without clear genotype-phenotype correlations. Based on prior evidence of effects of methionine metabolism and epigenetic alterations in WD, we hypothesized that alterations in DNA methylation could differentiate patients with WD from healthy subjects and from subjects with other liver diseases and could characterize different WD clinical trajectories. Whole-genome bisulfite sequencing was therefore performed on DNA from liver and whole blood of WD patients compared to healthy subjects, primary sclerosing cholangitis, and non-alcoholic fatty liver disease patients (liver $n=21$, blood $n=72$). In liver, 1840 differentially-methylated regions (DMRs) specifically differentiated WD from both healthy and disease control subjects. Functional annotation of WD-specific liver DMRs revealed a highly significant enrichment for liver-specific enhancers in WD-hypermethylated DMRs ($\log(\text{odds ratio})=11.6$, $\text{FDR } q\text{-value}=5.5E-278$). WD-hypermethylated DMRs were also enriched for binding sites of liver development transcription factors, including Hepatocyte Nuclear Factor 4 alpha (HNF4A), Retinoid X Receptor alpha (RXRA), Forkhead Box A1 (FOXA1), and FOXA2. Genes associated with WD-DMRs were enriched for functions in inflammatory response, as well as lipid and folic acid metabolism. 2458 DMRs discriminated liver samples from advanced compared to early stage WD patients. Genes with higher methylation in advanced WD patients were enriched for functions in complement activation (fold enrichment=18.2, $\text{FDR } q\text{-value}=6.2E-5$). In blood samples from an independent patient cohort, 262 WD-specific DMRs were identified. Genes near WD-specific blood DMRs significantly overlapped with those in liver, including at the genes for the drug-targetable epigenetic regulators histone deacetylase 4 (HDAC4) and HDAC5 (Hypermethylated: odds ratio=6.1, $p=6.8E-40$; Hypomethylated: odds ratio=3.7, $p=4.1E-8$). 2860 DMRs in blood distinguished WD patients with hepatic versus neurologic symptoms and were enriched for traits in mental deterioration (fold enrichment=2.1, $\text{FDR } q\text{-value}=0.04$). In conclusion, this novel epigenomic signature of WD reveals new insights into potential biomarkers and treatments for this complex genetic disease.

1878W**Using epigenetic age to investigate tissue origin in endometriosis.** K.

Leap, J. Martinez-Agosto, S. Horvath. UCLA, Los Angeles, CA.

Endometriosis is a common reproductive disorder defined by endometrial tissue that is found outside of the uterus. Despite its prevalence and high disease burden, it lacks effective treatment for both management of symptoms and recurrence of disease. Perhaps most surprising is that the origin of ectopic tissue is unknown. A focal question in the pathogenesis of the disease is whether ectopic tissue has an embryological origin or, according to Sampson's theory, arises from retrograde menstruation after puberty and menarche. We hypothesized that ectopic tissue would be younger, consistent with an embryological origin and contrary to Sampson's theory of retrograde menstruation. To determine biological age and compare it to chronological age, we used DNA methylation age, an established epigenetic clock. We analyzed Gene Expression Omnibus (GEO) DNA methylation data comparing ectopic diseased endometrial tissue, uterine tissue from affected women, and healthy uterine tissue. Our data analysis established that ectopic tissue is younger on average than both the woman's chronological age and the predicted DNA methylation age of the uterine tissue. We further explored the differences between these tissues through weighted gene co-expression network analysis (WGCNA). Using GEO gene expression data on ectopic and normally situated endometrial tissues in diseased women and healthy tissues, we identified biological networks relevant to differences between healthy and diseased women, as well as ectopic and eutopic tissues. These findings suggest that endometriosis may not result from a migration of uterine tissue to ectopic locations. The pathways and epigenetic state identified in this study will serve to explore additional hypotheses about biological origin of disease, including other conditions in which tissue origin is important in understanding pathogenesis, like cancer.

1879T**Gene expression profiling of liver tissue biopsies, PBMCs, and monocytes to predict treatment response for patients with alcoholic hepatitis.** N. Siddiqui¹, C. Magnan², Z.X. Liu², J.M. Jacobs³, T.R. Morgan⁴, T.M.

Norden-Krichmar¹, the Southern California Alcoholic Hepatitis Consortium. 1) University of California, Irvine, Irvine, CA; 2) Keck School of Medicine, University of Southern California, Los Angeles, CA; 3) Biological Sciences Division/ Integrative Omics, Pacific Northwest National Laboratory, Richland, WA; 4) Hepatology, VA Long Beach Healthcare System, Long Beach, CA.

Alcoholic Hepatitis (AH) is the most severe form of alcohol-induced liver disease and is associated with a high short-term mortality rate. Prednisolone, an established standard of care treatment for AH patients, does not successfully prevent the development of acute liver failure, multi-organ dysfunction, and death in all patients. To better understand this aspect of treatment, this study utilized transcriptomic profiling of human liver tissue biopsies, peripheral blood mononuclear cells (PBMCs), and isolated monocytes to explore how underlying variation in gene expression may be associated with differences in drug response. The protocol for this study was approved of by the IRB and written consent was obtained for all participants. Biological samples were collected at baseline, before the start of treatment administration, by the Southern California Alcoholic Hepatitis Consortium (SCAHC). RNA was extracted from these samples and sequenced at 2x100 paired-end on an Illumina HiSeq. Patients were clinically analyzed seven days after the administration of corticosteroid treatment for indication of drug response. Transcriptomic data for treatment responders (R) and non-responders (NR) for liver tissue (n = 12 R versus 6 NR), PBMCs (n = 9 R versus 7 NR), and monocytes (n = 7 R versus 3 NR) were aligned to the human genome and quantified. Differential expression analysis was conducted via the Cufflinks software, using upper quartile normalization between data files. Normalized differentially expressed genes were filtered to retain only those with an absolute value of log 2 adjusted fold change ≥ 1 , FPKM ≥ 1 , and p-value ≤ 0.05 . Ingenuity Pathway Analysis software was utilized to determine significantly enriched pathways. There were observed differences in gene expression between responders and non-responders in all cases. Common significant pathways included those related to immune function and the inflammatory response. Proteomics data from a high mass accuracy LC-MS/MS platform was available for correlation. The preliminary results suggest that there are characteristic gene expression profiles that can be observed at baseline for responders and non-responders to corticosteroid treatment. These differences may be detected via liver tissue or blood samples with potential ramifications in predicting the outcome of drug response and guiding the administration of future treatment. (Supported by NIAAA #U01AA021838, #U01AA021886).

1880F

Comparative developmental study of mesenchymal stem cell differentiation in primates. G. Housman¹, S. Pott^{1,2}, Y. Gilad^{1,2}. 1) Section of Genetic Medicine, University of Chicago, Chicago, IL; 2) Department of Human Genetics, University of Chicago, Chicago, IL.

Mesenchymal stem cells (MSCs) are multipotent adult stem cells that aid in the repair of tissue damage through their ability to differentiate into several mesodermal cell types, including bone, cartilage, muscle, and adipose. These features, in addition to their ability to be grown in culture, make MSCs ideal for the targeted treatment of skeletal injuries and diseases. However, current standards on the definition and purity measures of MSCs within and between species do not account for transcriptional heterogeneity. Gene expression profiles at the single-cell level may elucidate more accurate characterizations of MSCs. To evaluate this, we collect single-cell RNA expression patterns during the differentiation of human and chimpanzee iPSCs into MSCs, identify cellular subpopulations at different stages of differentiation, assess how well these subpopulations persist across technical replicates, and determine how differentiation-associated changes in gene regulation vary between humans and chimpanzees. We differentiated iPSCs from a human and a chimpanzee into MSCs using a well replicated study design and collected RNA at three time-points during each differentiation starting with the iPSC state (day 0), intermediate state (day 5), and MSC state (day 9). Standard imaging and FACS analyses were performed to confirm the progress of differentiation, and single-cell RNA sequence data (scRNAseq) were collected using a Drop-seq protocol. Cells progressed from round iPSC morphologies to elongated MSC morphologies and became plastic adherent during differentiation. Additionally, as expected, iPSCs only express the cell surface marker CD90, intermediate cells express CD90 and CD105, and MSCs express CD90, CD105, and CD73. Further, the dimensionality of the scRNAseq data reveals an initial increase in cellular heterogeneity during differentiation, as well as a maintenance of this subpopulation structure across technical replicates and conservation across species. Overall, these results begin to clarify the purity of MSCs at multiple comparative levels. This is the first study to assess gene expression patterns at the single-cell level in MSCs and to evaluate such differences across species. Continued work in this area will further clarify the expected heterogeneity of MSCs and thus, their feasibility for clinical applications.

1881W

Identification of novel key genes and pathways related to Hidradenitis suppurativa and its comorbidities using genome-wide DNA methylation analysis. S. Vishweswaraiah¹, N.M. Sayed², U. Ratnamala³, U. Radhakrishna¹. 1) Oakland University William Beaumont School of Medicine, Royal Oak, MI., United States; 2) Biotechnology, Nirma Institute of Science, Nirma University, Ahmedabad, India; 3) Department of Pharmacology, Creighton University, Omaha, NE, United States.

Hidradenitis Suppurativa (HS), usually presents after puberty with deep-seated inflamed lesions in the apocrine gland-bearing areas of the body, including the armpit (axillae), groin, perianal, and breast regions, with a prevalence of 1% to 4% in the general population. HS results from a combination of genetic and environmental factors and is negatively influenced by various lifestyle factors. The condition is characterized by genetic heterogeneity and is traced to specific γ -secretase gene mutations in some families. However, in sporadic cases, these gene mutations are not present. The Infinium Methylation EPIC BeadChip array (Illumina) was performed in 24 HS subjects, previously excluded from γ -secretase mutations and 24 normal controls. Differentially-methylated genes were used to generate a heatmap using the ComplexHeatmap (v1.6.0) R package (v3.2.2). Ward distance was used for the hierarchical clustering of samples. A multiple logistic regression analysis was done using more stringent criteria (FDR $p \leq 0.00001$ and ≥ 2 -fold change), to select candidate genes. Ingenuity Pathway Analysis (IPA) (Qiagen IPA) software was used to identify biological functions or interacting regulatory networks. The study identified a total of 20 cytosine loci which had excellent accuracy (AUC ≥ 0.90) for the detection of HS. IPA detected multiple signaling pathways implicated in the development of skin and other cancers, wound healing, immunological processes, inflammatory disorders, obesity, hypertension, and Type 2 diabetes mellitus. The study provided insights into the pathogenesis of HS by generating novel candidate genes and pathways.

1882T

Global methylation and transcriptional changes following endothelial cell activation during infection and autoimmunity. X. Shao¹, B. Rhead^{1,2}, O. Solomon¹, P. Ghai³, L.F. Barcellos^{1,2}, A.M. Bowcock^{3,4}. 1) Genetic Epidemiology and Genomics Laboratory, University of California, Berkeley, Berkeley, CA; 2) Computational Biology Graduate Group, University of California, Berkeley, Berkeley, CA; 3) Imperial College, London, UK; 4) Icahn School of Medicine at Mount Sinai, New York, NY.

Endothelial cells are a primary site of leukocyte recruitment during inflammation. An increase in tumor necrosis factor- α (TNF) levels as a result of infection, or in some autoimmune diseases can trigger this process and a number of autoimmune diseases including rheumatoid arthritis, psoriatic arthritis and psoriasis are now treated with TNF inhibitors. However, the genomic alterations that occur as a result of TNF-mediated inflammation are not well understood. Psoriasis is a chronic inflammatory disease of the skin that is also accompanied by increased vascularization due to proliferation and differentiation of resident endothelial cells following an inflammatory trigger. We were interested in the molecular targets and networks of this process. To investigate this we examined altered CpG methylation and gene expression in both primary keratinocytes and endothelial cells following TNF α stimulation. We analyzed responses in 40 human umbilical vein endothelial cell (HUVEC) and 35 normal human epidermal keratinocyte (NHEK, 24 fetal and 11 adult) primary cell lines. Genome-wide DNA methylation and gene expression profiles were generated on all 75 cell lines via microarray, before and 24 hours after stimulation with TNF α . Genes with differential expression (DE) in response to TNF α were identified in each cell type separately using paired analyses in *limma*. Differentially methylated regions (DMRs) were identified with *DMRcate*. HUVEC cells showed marked differences after stimulation with TNF α , with 408 DE genes (163 up- and 245 down-regulated) after adjusting for multiple hypothesis tests. There were 330 DMRs in HUVEC cells after stimulation with TNF α , with a strong tendency toward hypomethylation: 309 regions were decreased and 21 regions increased. There were 17 highly significant DMRs in or near a DE gene. These included the promoters of a number of genes upregulated during the inflammatory response and playing a role in leukocyte proliferation, activation or adhesion including the interleukin 7 receptor (IL7R), vascular cell adhesion molecule 1 (VCAM1), CD69, and CCL2 where several upstream enhancers were also DM. Many sites harbored binding sites for NF κ B transcription factors which are important inflammatory mediators. Although NHEK cells did not show significant DMRs or DE genes, which could be due in part to smaller sample size, the findings in HUVECs are providing novel insights into endothelial activation following an inflammatory stimulus.

1883F

Identification of hypermethylation of rDNA as an epigenetic regulator of rRNA expression in brain of Alzheimer's disease's patients. T.C. Faria¹, D.A. Santana¹, L.C. Santos¹, S.L.M. Payao², G. Turecki³, C.O. Gígeki¹, M.A.C. Smith¹, E.S. Chen¹. 1) Department of Morphology and Genetics, Genetics Division, Federal University of Sao Paulo, Sao Paulo, Brazil; 2) Genetics Division, Marília College of Medicine, Sao Paulo, Brazil; 3) Department of Psychiatry, Douglas Hospital Research Center, McGill University, Montreal, QC, Canada.

Alzheimer's disease (AD) is a neurodegenerative, irreversible and progressive disease, and the most common form of dementia. Alterations in rRNA expression have been previously associated with AD pathogenesis. In addition, hypermethylation of the rDNA promoter was associated with AD in parietal and prefrontal cortex. Nonetheless, it is unclear whether DNA methylation is the mechanism involved in rRNA expression regulation in AD. Hence, we aimed to investigate rDNA methylation pattern in 3 brain regions – auditory cortex, hippocampus and cerebellum – from AD and control individuals. We also analyzed gene expression of the DNA methylation machinery as well as the rRNA expression. Methylation was evaluated using bisulfite-conversion followed by NGS and gene expression was performed by qRT-PCR. Hypermethylation in 20 CpG sites (76.9%) of the rDNA promoter region was observed in AD cerebellum compared to controls; likewise, 6 CpGs sites (23%) were hypermethylated in AD hippocampus. Additionally, CpG#9 showed a methylation percentage higher than 40% in both groups in all regions when compared to all other CpGs sites, which presented a maximum average of methylation of 16.3%. Concerning gene expression, all evaluated methylation machinery genes showed increased expression in AD cerebellum and *MBD4* expression level was significantly higher in AD auditory cortex when compared to controls. We also verified increased expression of rRNA18S in AD in all brain regions and of rRNA28S in AD cerebellum. When brain regions were pooled, increased methylation in rDNA promoter and increased rRNA28S and *MBD2-3* expression were observed in AD compared to controls; moreover, a slight correlation between rDNA methylation and rRNA18S expression was verified. Additionally, strong correlations between rRNA18S and 28S expression levels were verified in AD in all brain regions. Furthermore, it was observed difference in rRNA18S expression level in AD between cerebellum and auditory cortex and, cerebellum and hippocampus; likewise, rRNA28S expression level differs in AD between cerebellum and auditory cortex and, auditory cortex and hippocampus. In conclusion, rDNA hypermethylation might be a mechanism involved in the regulation of rRNA expression; however, it may not be sufficient to silence rRNA expression in AD, suggesting that other mechanisms might be acting in rRNA expression. Also, some CpGs sites may have a relative importance in the dynamic methylation of rDNA promoter.

1884W

Alzheimer disease associated changes of RNA editing in temporal cortex and whole blood affects targetable pathways. O.K. Gardner^{1,2}, L. Wang^{1,2,3}, D. Van Booven^{1,2}, P.L. Whitehead^{1,2}, K.L. Hamilton-Nelson^{1,2}, L.D. Adams^{1,2}, S. Rolati^{1,2}, N.K. Hofmann^{1,2}, J.M. Vance^{1,2}, M.L. Cuccaro^{1,2}, W.S. Bush⁴, E.R. Martin^{1,2}, G.S. Byrd⁵, J.R. Gilbert^{1,2}, J.L. Haines⁴, G.W. Beecham^{1,2}, M.A. Pericak-Vance^{1,2}, A.J. Griswold^{1,2}. 1) John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL, USA; 2) Dr. John T Macdonald Foundation Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL, USA; 3) Department of Public Health Sciences, University of Miami Miller School of Medicine, Miami, FL USA; 4) Institute for Computational Biology and Department of Population & Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH, USA; 5) North Carolina A&T State University, Greensboro, NC, USA.

Though genetic studies have implicated several biological processes in the pathogenesis of Alzheimer Disease (AD), little is known about the epigenetic mechanisms that regulate them. This hinders our ability to target these mechanisms for preventative and therapeutic effects. RNA editing, the post-transcriptional modification of individual bases, plays a vital role in neuronal development and immune regulation, and dysregulated RNA editing has been observed in cancer and neurological diseases. Our goal was to test whether RNA editing alters genes in potentially druggable biological pathways by pursuing the first transcriptome-wide study describing AD related changes of RNA editing in brain and blood. Whole transcriptome RNA sequencing was performed on temporal cortex samples from 39 AD cases and 9 cognitively normal individuals and on peripheral blood specimens from 121 AD cases and 120 cognitively normal individuals of non-Hispanic White ethnicity. We applied a bioinformatics and statistical pipeline utilizing REDIttools software, quality control filtering, and statistical analysis. This resulted in 89,540 high-quality editing events in brain and 44,985 in blood. Using a logistic regression model, 2,949 positions in 803 genes and 723 positions in 371 genes were differentially edited in brain and blood, respectively. Across both tissues, differentially edited genes were enriched for biological processes in key pathways related to AD pathogenesis including: immune regulation, endocytosis, protein modification, and amyloid precursor protein catabolism. The type-1 interferon, NF-kappa B, and MAP kinase signaling cascades were particularly enriched. Genes composing all three of these pathways are regulated by RNA editing, have been implicated in AD in molecular studies, and are targetable by pharmaceuticals. Some differentially edited sites within these genes, including *TBKBP1* and *KDM6B*, are predicted to result in nonsynonymous coding changes or disrupted miRNA binding sites, further supporting the functional relevance of dysregulated RNA editing on these pathways. RNA editing in AD shows dysregulation of targetable inflammatory pathways in both blood and brain. As such, RNA editing studies can help identify new targets for therapeutics, and editing signatures may serve as biomarkers for pathogenic processes in AD.

1885T

DNA methylation analysis of cognitive decline and neuropathology. A. Lu¹, K. Nho^{2,3}, A. Vasanthakumar¹, J. Waring¹, J. Wade Davis⁴, A. Saykin^{2,3,5}, M. Levine⁶, S. Horvath^{1,7}. 1) Human Genetics, UCLA, Los Angeles, CA; 2) Center for Neuroimaging, Department of Radiology and Imaging Sciences, Indiana University School of Medicine, Indianapolis, IN, USA; 3) Indiana Alzheimer Disease Center, Indiana University School of Medicine, Indianapolis, IN, USA; 4) AbbVie Inc., North Chicago, Illinois, USA; 5) Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, USA; 6) Department of Pathology, Yale School of Medicine, New Haven, CT, 06510, USA; 7) Biostatistics, School of Public Health, University of California Los Angeles, Los Angeles, USA.

Nearly 36.6 million people live with dementia. This number is expected to more than triple by 2050 according to World Health Organization. Despite significant progress, the molecular underpinnings of dementia remain poorly understood. A recent DNA methylation based biomarker of aging (known as DNAm age or epigenetic clock) has revealed a relationship between Alzheimer's disease (AD) related neuropathology and epigenetic aging rates in brain tissue but not yet in blood. Here we relate DNA methylation levels and DNAm age in blood to dementia related traits, using the Alzheimer's Disease Neuroimaging Initiative (ADNI) longitudinal cohort. Our study involved 332 individuals with mild cognitive impairment (MCI) and 218 controls. The mean age was 74 years at baseline and the mean follow-up period was 3.6 years. Our study focused on i) an imaging biomarker of cerebral glucose metabolism, FDG-PET, sensitive to dementia and ii) a rate of change in cognitive assessments based on the Mini-Mental State Examination (MMSE). We defined a measure of MMSE decline on the basis of a linear mixed model. Specifically, cognitive decline was defined as minus the estimated random slope effect with respect to MMSE on time. DNA methylation levels were measured on the Illumina EPIC array (~ 850k CpGs). Our epigenome-wide association study (EWAS) of MMSE decline (adjusted for age) identified two loci at a suggestive significance level of $P < 10^{-6}$. The top EWAS hit is within the *SORCS1* gene ($P = 3.4E-6$). Interestingly, *SORCS1* is known to alter amyloid precursor protein processing. The second hit is near the *KCNH5* gene ($P = 9.6E-6$), a member of voltage-gated K⁺ channels implicated in epilepsy. Our epigenetic clock analysis of longitudinal changes of FDG-PET assessed the-pan tissue epigenetic clock by Horvath (353 CpGs), the blood based clock by Hannum (71 CpGs) and the estimator of DNAm PhenoAge by Levine (513 CpGs). Our growth curve analysis showed that FDG-PET was inversely related to all three clocks ($=0.034$, $=0.053$, and $=0.019$). A secondary analysis revealed that FDG-PET exhibited weak associations with age-adjusted DNAm based blood cell counts: a positive correlation with B cells ($P = 0.039$) and a negative correlation with CD8⁺ T cells composition ($P = 0.042$). In summary, our DNA methylation studies in blood a) identify two loci for which methylation levels relate to cognitive decline, and b) demonstrate that epigenetic age acceleration relates to cerebral glucose metabolic changes.

1886F

DNA hypomethylation in blood links *B3GALT4* ganglioside and *ZADH2* prostaglandin pathways to Alzheimer's disease.

A. Madrid^{1,2}, K. Hogan^{3,4}, L. Papale¹, L. Clark^{4,5,6}, S. Asthana^{4,6}, S. Johnson^{4,5,6}, R. Alisch¹. 1) Department of Psychiatry, University of Wisconsin School of Medicine and Public Health, Madison, WI, USA; 2) Neuroscience Training Program, University of Wisconsin School of Medicine and Public Health, Madison, WI, USA; 3) Department of Anesthesiology, University of Wisconsin School of Medicine and Public Health, Madison, WI, USA; 4) Wisconsin Alzheimer's Institute, University of Wisconsin School of Medicine and Public Health, Madison, WI, USA; 5) Wisconsin Alzheimer's Disease Research Center, University of Wisconsin School of Medicine and Public Health, Madison, WI, USA; 6) Geriatric Research Education and Clinical Center, William S. Middleton Memorial Veterans Hospital, Madison, WI, USA.

Alzheimer's disease (AD) is the most common neurodegenerative disease afflicting the elderly population. Late-onset AD (LOAD) is characterized by symptoms presenting after the age of 65 that include dementia and memory loss. While the etiology of LOAD remains largely unknown, evidence suggests that environmental influences increase disease risk, particularly through interactions with the epigenome. Recent epigenome-wide association studies (EWAS) using brain tissue report differential methylation in known and newly recognized LOAD genes, thereby underscoring the utility of EWAS in disclosing novel genes and pathways associated with LOAD pathogenesis. As an alternative to the study of donor brain tissues, investigation of DNA methylation in accessible peripheral tissues may provide an opportunity to improve clinical diagnosis and estimates of prognosis. To identify differentially methylated positions (DMPs) in blood that distinguish persons with and without LOAD, genomic DNA from the whole blood of 45 LOAD patients and 39 matched controls were interrogated using the Illumina HumanMethylationEPIC array platform, which detects DNA methylation levels at >850,000 sites spanning the genome. DMPs were observed at 477 loci corresponding to 106 genes. Remarkably, 17 of these DMPs were shared between six clinical markers of LOAD that were analyzed independently as continuous variables, comprising Rey Auditory Verbal Learning Test (RAVLT) scores, cerebrospinal fluid total tau (t-tau) and phosphorylated tau 181 (p-tau₁₈₁) levels, and t-tau/Amyloid beta (AB₁₋₄₂ (AB₄₂)), p-tau₁₈₁/AB₄₂, and AB₄₂/AB₁₋₄₀ (AB₄₀) ratios. In patients with LOAD, 12 of the 17 shared DMPs were hypomethylated in *B3GALT4* (Beta-1,3-galactosyltransferase 4), a gene previously associated with LOAD in the superior temporal gyrus of the brain. The remaining five shared DMPs were hypomethylated in *ZADH2* (Prostaglandin reductase 3), a novel LOAD-associated gene. Taken together, these data reinforce the use of blood as an accessible tissue of value in the identification of DMPs associated with dementia onset and progression. Our data lend further support to the role of enzymes in ganglioside and prostaglandin metabolism in the pathogenesis of Alzheimer's disease, and may provide novel diagnostic, prognostic, and modifiable therapeutic targets.

1887W

Identification of genes regulated by H3K9 acetylation in brains of Alzheimer's disease patients: A ChIP-Seq approach.

D.A. Santana¹, R.D. Puga², T.C. Faria¹, G. Turecki³, C.O. Gigue⁴, M.A.C. Smith¹, E.S. Chen¹. 1) Genetics Division, Department of Morphology and Genetics, Universidade Federal de Sao Paulo (UNIFESP), Sao Paulo, SP, Brazil; 2) Department of Clinical Research, Hospital Israelita Albert Einstein, Sao Paulo, SP, Brazil; 3) Department of Psychiatry, Douglas Hospital Research Center, McGill University, Montreal, QC, Canada; 4) Department of Pathology, Universidade Federal de Sao Paulo (UNIFESP), Sao Paulo, SP, Brazil.

Alzheimer's disease (AD) is the most common neurodegenerative disease and the leading cause of irreversible dementia. Evidences suggest that epigenetic modifications, such as histone post-translational modifications, may increase the risk of AD. Thus, we aimed to identify genes regulated by H3K9 acetylation, an essential mark of transcriptional activation, in three brain regions – auditory cortex, hippocampus and cerebellum – of AD compared to control individuals. Samples underwent ChIP-Seq procedure and were pooled according to the region and the group for library preparation. Libraries were sequenced using Illumina HiSeq 2500 platform, in Rapid Run mode, producing single-ended reads. Quality control was evaluated using FastQC, and the alignment to reference genome (GRCh37/hg19) was performed using Bowtie2. Peak calling, differential analysis between groups, and identification of motifs in the output data were undertaken by HOMER software. Functional and gene network analysis were also performed using Gene Ontology (GO) and GANT, respectively. ChIP-Seq analysis revealed that H3K9 acetylation was decreased in 17 genes, such as *ACHE* and *ADRA2A*, in AD compared to controls. Moreover, 71 genes, including *IREB2*, *RYR1*, *HMOX2*, *RORA*, *NRF1*, *ARSB* and *SAA1*, showed increased H3K9 acetylation levels in AD group. Gene network analysis demonstrated interactions among these genes and AD-associated genes (*APP*, *PSEN2* and *APOE*). Functional analysis revealed that most genes are involved in cellular process, cell part and catalytic activity. In AD group, motif discovery indicated the enrichment of the transcription factor *ERRα*, which is highly expressed in nervous system. In conclusion, we identified several differentially H3K9 acetylated genes in AD in comparison to controls, which could lead to abnormalities in the expression of such genes. Most identified genes were not previously described in AD. Therefore, our results indicate the importance of studying epigenetic modifications to better understand the pathophysiology of AD.

1888T

Transcriptomic association analysis with Alzheimer's disease related traits in ROS/MAP, MSBB and Mayo Clinic cohorts. C. Zhang, J. Hutz, A. Mitchell. Andover innovative medicines institute, Eisai Inc, Andover, MA.

Dementia, with Alzheimer's disease (AD) as the most common form, is currently affecting 36 million people worldwide and expect to affect 115 million by 2050 without effective therapeutic development. Large-scale RNA-seq analysis from human brain samples is a promising method for biomarker and drug target discovery, which is essential to the development of novel drugs for AD treatment and prevention. The Religious Orders Study and Memory and Aging Project (ROSMAP) study, the Mount Sinai Brain Bank (MSBB) study and the MayoRNAseq study are the three largest late onset AD cohorts with brain samples RNA-seq data linked to rich individual level clinical data in Accelerating Medicines Partnership - Alzheimer's Disease (AMP-AD) consortium. While there are publications associated with each individual study, it is not clear if the genes significantly associated with AD in one study can be replicated in other cohorts. Here, we performed association analysis with AD related traits using uniformly processed RNA-seq data from the three cohorts and found a list of protein coding and non-coding genes consistently associated with AD status, cognitive decline and neuritic plaques across all three cohorts. Some of the genes in this list, including *CCL4L2*, *GABRR2*, *CRH* and *VGF* are already being studied by the community for their neural or neuro-inflammation related functions, while others are not as well studied and can be new biomarkers or targets of interest. The newly discovered genes, including some long non-coding RNAs, *WISP1*, *APLNR*, *SIGLEC1*, *LTF*, *SERPINA5*, *SST* and *FREM3*, have broad range of functions in inflammation, cell signaling, cell proliferation and oxidative stress. We think this gene list can be complimentary to genes found in recent late onset AD genome wide association studies, where neuro-inflammatory genes became of high interest. We also highlight the importance for controlling covariates, such as RIN, batch effects, age, etc., in analysis of this scale.

1889F

A suppressor screen in *Mecp2* mice implicates the DNA damage response in Rett syndrome pathology. A. Enikanolaiye¹, C.M. Buchovecky², M.J. Justice^{1,2}. 1) The Hospital for Sick Children, Toronto, Canada; 2) Department of Molecular Genetics, University of Toronto, Toronto, Canada; 3) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

Rett syndrome (RTT) is an X-linked neuro-metabolic disorder caused by mutations in methyl-CpG binding protein 2 (*MECP2*). RTT is characterized by normal postnatal development, after which previously acquired motor and communicative skills are lost. Re-expression of *Mecp2* in symptomatic mutant mice rescues neurological defects and prolongs survival. This suggests that the disease may be improved or reversed by genetic or pharmacological intervention. To identify novel candidate pathways for such intervention, a dominant random mutagenesis screen for suppressors of RTT symptoms was carried out in a *Mecp2*-null mouse model of RTT. Genes that suppress the disease phenotype were isolated by whole exome sequencing and linkage mapping. Strikingly, multiple mutations in the DNA damage response (DDR), which repairs double-stranded breaks (DSBs) can contribute to the improvement of RTT-like symptoms in mice. Expression studies show that levels DDR and other DNA damage factors are perturbed in *Mecp2*-null mice brain and liver. In particular, γ H2AX, a marker of DSBs is upregulated in the brain and liver of *Mecp2*-null mice. *In vitro* assays that measure the frequency of DSB repair pathways, show that the rate of homologous recombination is upregulated in *Mecp2*-null mice. Because neurons are non-dividing cells, an ongoing question is why mutations in this pathway improve symptoms in mice. Unbiased genetics screens such as these can help uncover unexpected pathways that contribute to RTT pathophysiology and that can provide insight into therapies. DSB repair pathways may also prove to be important in other autism spectrum disorders as well.

1890W

Placental DNA methylation at the interface of genetics and environment in the MARBLES prospective autism study.

Y. Zhu¹, C. Mordaunt¹, K. Dunaway¹, D. Yasui¹, R. Coulson¹, P. Krakowiak², B. Puschner³, Y. Lin³, R. Marathe⁴, T. Totah¹, C. Walker¹, S. Ozonoff^{5,6}, I. Hertz-Picciotto^{2,6}, F. Hormozdizadeh⁷, R. Schmidt⁶, J. LaSalle^{1,6,7}. 1) Medical Microbiology and Immunology and Genome Center, University of California at Davis, Davis, CA; 2) Public Health Sciences, University of California at Davis, Davis, CA; 3) School of Veterinary Medicine, University of California at Davis, Davis, CA; 4) Obstetrics and Gynecology, University of California at Davis, Davis, CA; 5) Psychiatry and Behavioral Sciences, University of California at Davis, Davis, CA; 6) MIND Institute, Sacramento, University of California at Davis, Davis, CA; 7) Genome Center, University of California at Davis, Davis, CA.

DNA methylation acts at the interface of genetic and environmental risk and protective factors relevant for autism spectrum disorder (ASD). Placenta, a tissue normally discarded at birth, is a potentially rich source of DNA methylation patterns predictive of autism. The MARBLES (Markers of Autism Risk in Babies-Learning Early Signs) prospective cohort study enrolls pregnant mothers with at least one child with ASD. Mothers were interviewed about prenatal vitamin use and their blood was sampled at each trimester for environmental exposures, including persistent organic pollutants polychlorinated biphenyls (PCBs) and polybrominated diphenyls ethers (PBDEs). Placentas were collected for MARBLES births and stored frozen until diagnosis of ASD, typical development (TD), or non-typical development (NTD) in the child at age three. Whole genome bisulfite sequencing was performed on DNA isolated from 20 ASD and 21 TD male placentas. 352 differential methylated regions (DMRs) on 553 genes discriminated ASD from TD samples. Percent methylation over these DMRs showed strong positively associated with autism severity scores but not other potential confounding variables. Placental ASD DMRs were significantly enriched in active transcription start sites and its flanking regions enriched in H3K4me1 and H3K4me2 from chromHMM chromatin state predictions, suggesting an impact on gene expression. Two placental ASD DMRs passed a family-wise error rate sample permutation test (adjusted $p < 0.05$), at gene loci *IRS2* and *CYP2E1*. Both *IRS2* and *CYP2E1* loci replicated by an independent method of pyrosequencing ($p = 0.014$ and $p = 0.004$) and increased protein levels of *IRS2* in ASD were confirmed by Western blot. Methylation at *CYP2E1*, encoding a member of the cytochrome P450 superfamily, was primarily driven by genetic alterations (rs1536828). While percent methylation at *IRS2*, encoding insulin receptor substrate 2, was not significantly affected by genotype, multiple PCB and PBDE congeners showed significant associations with *IRS2* methylation, especially PCB 118 ($p = 0.005$). *IRS2* methylation also significant associated with prenatal vitamin intake ($p = 0.012$). ASD diagnosis, *IRS2* genotype and PCB 118 level formed a significant three-way interaction on explaining *IRS2* methylation ($p = 0.019$). This study identified epigenetic signatures of ASD risk in placenta that environmentally sensitive to prenatal vitamin, PCB, and PBDE exposures in pregnancies for ASD.

1891T

Alcohol abuse associated allele specific expression and regulation in human brain tissue.

X. Rao¹, H. Lin¹, K. Hargreaves², J. Ipe³, K. Thapa³, H. Gu³, M. Kapoor⁴, S. Farris⁵, T. Foroud¹, A. Goate⁴, H. Edenberg^{1,3}, R. Mayfield⁶, T. Skaar², Y. Liu¹. 1) Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN; 2) Division of Clinical Pharmacology, Department of Medicine, Indiana University School of Medicine, Indianapolis, IN; 3) Department of Biochemistry & Molecular Biology, Indiana University School of Medicine, Indianapolis, IN; 4) Ronald M. Loeb Center for Alzheimer's Disease, Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, New York; 5) Waggoner Center for Alcohol and Addiction Research, University of Texas at Austin, Austin, TX.

Alcohol abuse and alcoholism are significant public health problems. As a central nervous system depressant, high levels of alcohol intake over a long period of time may alter brain function to promote alcohol addiction and have extensive damaging effects on the brain. Understanding the molecular mechanism of how alcohol affects the brain will be very important to prevent alcohol abuse and to reverse the impact of heavy drinking on the brain. Animal models have been widely used to identify candidate mechanisms, which greatly advanced our knowledge of molecular basis of alcoholism. However, due to ethical considerations, there were very limited molecular evidence from human brain. Through collaboration between the Collaborative Studies on Genetics of Alcoholism (COGA) and Integrative Neuroscience Initiative on Alcoholism (INIA) Consortia, we obtained four regions of brain tissues of 60 human samples from NSW Brain Bank (Australia), of which 30 were heavy drinker and 30 were social/non-drinkers. By integrating deep RNA-seq data with GWAS array results, we identified 90 genes with differential allele-specific expression (ASE) in heavy drinkers, compared to social/non-drinker, in either of four brain regions, many of which have previously been implicated in neurological diseases. Because ASE provides an excellent model to study *cis*-regulatory mechanisms (both alleles are in the same cellular environment), single nucleotide variants (SNVs) in regulatory regions from GWAS array results were further examined, including upstream enhancer/promoter regions, gene bodies, 3'UTRs and downstream regions. Hundreds of SNVs were identified as highly potential *cis*-regulators based on their location in open chromatin regions, identification as expression quantitative trait loci (eQTLs) in GTEx, and prediction of altering transcription binding and miRNA binding sites. Using massively parallel reporter assays, we validated the regulatory role of several SNVs in 3'UTR and will further validate the functional SNVs in enhancer and promoter regions. Our findings will advance the understanding of alcohol abuse associated gene regulation in human brain and serve as a research framework for further study with expanded human brain tissues.

1892F

Screening of drugs conferring an increase in miRNA that is deficient in Huntington's disease. *H. Shimizu, E. Suzuki, M. Fukuoka, H. Hohjoh.* Department of Molecular Pharmacology, National Institute of Neuroscience, National Center of Neurology and Psychiatry, Kodaira, Tokyo, Japan.

Huntington's disease (HD) is an inherited neurodegenerative disorder characterized by involuntary movement, cognitive decline and psychiatric symptoms, and is caused by mutant *Huntingtin (HTT)* genes carrying aberrantly expanded CAG triplet repeats that are translated to abnormal polyglutamine tracts. Mutant HTT proteins adversely affect biomolecules and genes including functional RNA genes such as microRNAs (miRNAs). Our previous study indicated that many miRNAs were markedly decreased in the brain of HD-model mice (R6/2 mice) relative to that of wild-type mice. Of such decreased miRNAs, *miR-132* was more prominent than other miRNAs. In addition, it had been reported to play important roles in neuronal maturation and functions. Based on the findings, we performed a challenging experiment, in which *miR-132* was supplied into R6/2 brains deficient the miRNA. The results indicated that *miR-132* supplementation brought about amelioration to motor functions and prolonged the lifespan of the treated R6/2 mice (Fukuoka et al., 2018). These findings led us to the possibility that recovery of the *miR-132* gene expression back to a normal level might produce symptomatic improvements as in *miR-132* supplementation. We attempted to search for biomolecules and chemical compounds that are capable of inducing the expression of endogenous *miR-132* gene. We constructed an *in vitro* assay system to search for such molecules. Because the expression of *miR-132* may be regulated by CREB transcription factor, and because the *miR-132* gene has a putative CREB binding site in its predicted promoter region, the luciferase reporter gene carrying a consensus CREB binding site was constructed and introduced together with a mutant *HTT* gene into SH-SY5Y cells, a human neuroblastoma cell line. The cells were subjected to exposure with various chemical compounds followed by a conventional luciferase reporter assay; and, molecules conferring enhancement of the expression of the reporter gene even under the presence of mutant HTT were selected. We screened approximately 1,200 chemical compounds and selected several candidates so far. *In vitro* and *in vivo* experiments using the candidates are currently under progress. We would like to discuss the assay system and the data related to the candidates in this meeting at San Diego.

1893W

Leveraging brain cortex-derived molecular data to elucidate epigenetic drivers of neurological function and disease. *C. Hatcher, T. Gaunt, T. Richardson.* MRC Integrative Epidemiology Unit, University of Bristol.

Introduction: We have performed extensive analyses to uncover regions of the genome where transcriptional and epigenetic mechanisms may play a role in the susceptibility of neurological traits and disease. **Methods:** This was undertaken by applying Bayesian multiple-trait colocalization (moloc) to assess the likelihood that genetic variants responsible for trait variation also influence intermediate molecular phenotypes. We harnessed high dimensional data concerning gene expression, DNA methylation and histone acetylation quantitative trait loci derived from prefrontal cortex tissue (available from the xQTLServe and ROSMAP resources). Summary statistics from genome-wide association studies (GWAS) on five neurological traits (Neuroticism, Schizophrenia, Educational Attainment, Insomnia and Alzheimer's disease) were obtained from large-scale consortia and the UK Biobank study. **Results:** There was evidence at 125 independent loci that genetic variants which influence these complex traits may be acting via changes in gene expression within brain tissue. Of these, 81 were also found to influence proximal DNA methylation and/or histone acetylation. These findings support previous evidence implicating DNA methylation as a transcriptional repressor, as we observed enrichment of an inverse relationship between DNA methylation and gene expression across these loci. Our results support previous findings implicating epigenetic factors in neurological trait susceptibility (such as at *KLC1* with schizophrenia), as well as uncovering evidence at dozens of novel loci not previously identified. For instance, variation at *MDGA1* has not previously been identified by GWAS, although our analysis provided evidence of colocalization between molecular phenotypes and insomnia risk at this locus (total posterior probability of association = 86%). This finding has recently been validated by a large-scale GWAS of insomnia (P-value = 4.0×10^{-13}), suggesting that the other novel loci which colocalize with molecular phenotypes in our study are likely to be identified by GWAS as study sizes increase. **Conclusion:** We have identified evidence at dozens of loci that the same causal variant influencing complex traits is responsible for changes in epigenetic factors and gene expression. Integrating molecular data with GWAS can elucidate the biological mechanisms which influence traits, as well as help prioritise both known and novel candidate genes in disease.

1894T

Human neuronal models of *CHD2*-associated epilepsy. *K.J. Lamar^a, A.M.*

Tidball^b, J.M. Parent^c, G.L. Carvill^d. 1) Ken and Ruth Davee Department of Neurology, Northwestern University Feinberg School of Medicine, Chicago, IL; 2) Department of Neurology, University of Michigan, Ann Arbor, MI, USA.

Pathogenic variants in *CHD2* are associated with developmental epileptic encephalopathy (DEE) in humans. The majority of these variants are frame-shifts or deletions and all are heterozygous, having arisen *de novo*, indicating that haploinsufficiency is the mode of pathogenesis. Chromodomain helicase DNA binding protein 2 (*CHD2*) is a chromatin remodeler that is associated with sites of active transcription. To better understand the role of *CHD2* in human DEE, we utilized the CRISPR/Cas9 system to create heterozygous disruptions of *CHD2* in human induced pluripotent stem cells (iPSCs). To examine the effect of *CHD2* disruption in neuronal development, we converted these iPSC lines into neuronal progenitor cells (NPCs) and performed RNA sequencing on *CHD2*^{-/-} NPCs and isogenic wild-type (WT) controls. We found that *CHD2*^{-/-} NPCs exhibited an upregulation of genes involved in neuronal differentiation and a downregulation of genes involved in proliferation. *CHD2* disruption in NPCs also lead to misregulation of other epilepsy genes. Basic fibroblast growth factor (bFGF) is a potent mitogen for NPCs and suppresses differentiation. In order to test for a possible proliferation defect in *CHD2*^{-/-} NPCs, *CHD2*^{-/-} and WT NPCs were grown with or without bFGF and labeled with BrdU for 30 minutes. When grown in the presence of bFGF, *CHD2*^{-/-} NPCs did not have a significant difference in the percentage of BrdU+ cells compared to WT NPCs. However, when grown without bFGF for 3 days prior to BrdU labeling, the percentage of BrdU+ cells in *CHD2*^{-/-} NPCs was significantly lower than that of WT NPCs, indicating that *CHD2*^{-/-} cell lines have a proliferation defect at the neural progenitor stage and that bFGF may mask this phenotype. Overall, our results suggest that *CHD2* haploinsufficiency results in cell proliferation defect at the NPC stage which likely contributes to premature neuronal differentiation. In the future, we will examine expression changes in *CHD2*^{-/-} and WT lines during neuronal differentiation. Moreover, we aim to identify *CHD2* targets and epigenetic changes during development. Ultimately, we seek to understand the mechanisms leading to gene expression changes when *CHD2* is disrupted and the role of *CHD2* in regulating other epilepsy genes. The role of chromatin remodeling and control of gene expression is largely unexplored in epilepsy, but it is a dynamic process with many therapeutic targets and presents a novel opportunity to develop new treatments for patients.

1895F

Three-dimensional chromatin organization and gene regulation in human brain cells. *M. Song^{1,2}, X. Ren¹, C. Wang⁴, S. Lu⁵, M. Traglia², L. Maliskova¹, I.*

Jones¹, T. Tam¹, J. Dixon⁶, L. Weiss^{1,7}, J. Yao⁵, L. Gan⁴, Y. Shen^{1,2,3}. 1) Institute for Human Genetics, University of California, San Francisco, San Francisco, CA, USA; 2) Pharmaceutical Sciences and Pharmacogenomics Graduate Program, University of California, San Francisco, San Francisco, CA, USA; 3) Department of Neurology, University of California, San Francisco, San Francisco, CA, USA; 4) Gladstone Institute for Neurological Disorders, San Francisco, CA, USA; 5) State Key Laboratory of Membrane Biology, Tsinghua-Peking Center for Life Sciences, School of Life Sciences, Tsinghua University, Beijing, China; 6) Salk Institute for Biological Studies, La Jolla, CA, USA; 7) Department of Psychiatry, University of California, San Francisco, San Francisco, CA, USA.

Mutations in gene regulatory elements have been associated with a wide range of neurological disorders. However, due to their inherent cell type specificity and difficulties in characterizing their regulatory targets, our ability to identify causal genetic variants has remained limited. To address these constraints, we perform integrative analysis of chromatin interactions (promoter capture Hi-C), open chromatin regions (ATAC-seq), and transcriptomes (RNA-seq) in three functionally distinct cell types associated with neurological disorders: iPSC-induced cortical excitatory neurons, hippocampal dentate gyrus (DG)-like neurons, and human primary fetal astrocytes. We identify tens of thousands of cell type-specific, long-range *cis* interactions between promoters and distal promoter-interacting regions (PIRs), enabling us to link regulatory elements to their target genes and identify putative pathways and genes that are dysregulated in disease. We perform validation of novel PIRs linked to neuronal function using CRISPR-Cas9 genome editing in human neurons. Finally, we demonstrate that physical chromatin interactions can mediate genetic interactions in autism spectrum disorder (ASD). Our study presents a catalog of regulatory interactions in the human brain and advances our understanding of how regulatory variants contribute to complex neurological disorders.

1896W

Investigating cell type-specific promoter usage to uncover new biology:

Defining the role of SOX10 in myelination. E. Fogarty¹, A. Antonellis^{1,2}. 1) Neuroscience Graduate Program, University of Michigan, Ann Arbor, MI; 2) Department of Human Genetics, University of Michigan, Ann Arbor, MI.

The variety of cellular identities in a multicellular organism is dependent on genomic complexity achieved partly by differential use of multiple transcription start sites (TSSs) at individual loci. TSS use can be regulated by cell type-specific transcription factors and defining these interactions provides insights into cell type-specific biology. Schwann cells and oligodendrocytes are the myelinating glia of the peripheral and central nervous systems, respectively. SOX10 is a transcription factor required for the development and maintenance of both cell types and SOX10 target genes include loci critical for myelination and genes implicated in demyelinating disease. We and others have described SOX10-responsive promoters at loci with multiple TSSs. We sought to leverage these features of SOX10 activity to uncover novel pathways and gene isoforms important for myelination. Toward this, we performed cap analysis gene expression (CAGE) on cultured Schwann (S16) cells and intersected these data with marks of SOX10 binding, active promoters, and open chromatin to identify 4,032 candidate SOX10-regulated promoter elements. Next, we used RNA-Seq to assay the expression of transcripts associated with these promoters in S16 cells ablated for SOX10 expression (Δ SOX10) and found that 488 TSSs have reduced expression with loss of SOX10, while 362 TSSs have increased expression. Gene ontology analyses show that TSSs with reduced expression in Δ SOX10 S16 cells reside at loci associated with pathways known to be SOX10-regulated including myelination and glial differentiation. Interestingly, these data also show an enrichment for terms not previously associated with SOX10: cell junctions, adhesion, and GTPase signaling. A target of interest is *FGD4*, a known demyelinating disease gene that encodes a GTPase exchange factor. Only one of three *FGD4* TSSs is utilized in Schwann cells and it is associated with a SOX10-regulated promoter element. Luciferase assays show that the promoter is active in S16 cells and deletion of a conserved SOX10 binding site reduces the activity to 25%. These findings support a model where loss of the protein specifically expressed from the SOX10-regulated TSS causes demyelinating disease. In sum, our data support the utility of leveraging cell type-specific transcription factors in defining TSS use genome-wide toward investigating isoform-specific gene function in a cell-type specific manner.

1897T

Single cell analysis of Alzheimer's disease. M. Kellis^{1,2}, H. Mathys^{1,3}, J. Davila-Velderrain^{1,3}, S. Mohammad^{1,3}, D. Bennett⁴, L.-H. Tsai^{1,2}. 1) Massachusetts Institute of Technology, Cambridge MA 02139; 2) Picower Institute for Learning and Memory, Cambridge MA 02139; 3) Computer Science and Artificial Intelligence Lab, Cambridge MA 02139; 4) Department of Neurological Sciences, Rush Medical College, Rush Alzheimer's Disease Center, Chicago, Illinois 60612.

Alzheimer's disease (AD) is a devastating disorder characterized by cognitive decline and neuronal loss, whose complex etiology involves multiple brain cell types, but remains poorly characterized, due to the mostly-bulk-tissue-based analyses across disease cohorts. Here, we provide the first single-cell view of AD pathology across 48 individuals, including 24 AD and 24 non-AD individuals from the ROSMAP cohort, both male and female post-mortem samples, using 80,660 droplet-based single-nucleus transcriptomes. We clustered all cell jointly across individuals, resulting in 20 transcriptionally-distinct cell clusters capturing excitatory and inhibitory neurons, astrocytes, oligodendrocytes, microglial cells, and biologically-meaningful cellular subpopulations within them. We found systematic expression differences between AD and non-AD individuals, pinpointing specific candidate driver genes, including *LINGO1*, *SLC26A3*, and *RASGEF1B*, upregulated in all cell types, *APOE* upregulated in microglia only, *MT1E* upregulated in astrocytes, and *CRYAB* upregulated in oligodendrocytes. For microglia, up-regulated genes in AD were enriched in THE complement system, and down-regulated genes were enriched in insulin signaling. For astrocytes, AD-upregulated genes were enriched in cell-cell interactions and adhesion, and down-regulated genes were enriched in vesicular transport. For oligodendrocytes, AD-upregulated genes were enriched in translation and proteostasis, and down-regulated genes were enriched in metabolism. For excitatory neurons, AD-upregulated genes were enriched in calcium signaling, and down-regulated genes were enriched in immune pathways and metabolism. For inhibitory neurons, AD-upregulated genes were enriched in mTOR signaling, and down-regulated genes were enriched in glucose metabolism. Remarkably, two of the cellular subpopulations identified were only found in AD, indicated excitatory-neuron and oligodendrocyte cell states, involved in cognitive impairment, amyloid burden, and neuritic plaques. Surprisingly, early AD progression showed highly cell-type-specific transcriptional changes, and later-stage progression of AD showed common transcriptional changes across multiple cell types. Surprisingly, we found that cells isolated from women were overrepresented in both AD-associated cell states identified, consistent with the nearly 2-fold higher prevalence of AD in women, even accounting for age.

1898F

The spliced lncRNA SNORD116HG is essential for the high order chromatin dynamics of the MAGEL2 and NDN locus over long distance. S. Horike¹, S. Akagi¹, G. Okada¹, D.H. Yasui², J.M. LaSalle², M. Meguro-Horike¹. 1) Advanced Science Research Center, Kanazawa University, Kanazawa, Japan; 2) Department of Medical Microbiology and Immunology, UC Davis, USA.

Prader-Willi syndrome (PWS) and Angelman syndrome (AS) are distinct neurodevelopmental disorders that are caused by deletions of a differential parental origin of chromosome 15q11-q13. The PWS/AS imprinted gene cluster on 15q11-q13 spans 4Mb region that includes several protein-coding genes and multiple long non-coding RNAs (*SNORD116HG*, *IPW*, *SNORD115HG*) which are positively regulated by the PWS imprinting center (PWS-IC). Despite the fact that PWS-IC is known to be necessary for the coordinately controlled gene expression of 15q11-q13 imprinted locus, little is known concerning the mechanism by which it act and how it influence to the *MAGEL2* and *NDN* gene expression over long distance. In order to define the precise long-range gene regulation of 15q11-q13 by the imprinted lncRNAs, we have previously investigated the chromosomal architecture and dynamics by DNA-RNA FISH analyses. Our FISH analyses revealed that active paternal *MAGEL2* locus (*MAGEL2* expression is ON) was looping out from its chromosome territory, but inactive maternal locus (*MAGEL2* expression is OFF) was retained in the inside of its chromosome territory. Moreover, we found that the deletion of the PWS-IC causes the inactivation of the *MAGEL2* and *NDN* gene expression, and the re-localization of *MAGEL2* locus in the chromosome 15 territory. Therefore, we hypothesized that some of the imprinted lncRNAs such as *SNORD116HG*, *IPW*, *SNORD115HG* may be involved in the chromosomal architecture and dynamics of *MAGEL2* locus. In this study, we constructed three types of modified human chromosome 15(Δ *SNORD116* locus, Δ *SNORD116* & *IPW* locus, and Δ *SNORD116* & *IPW* & *SNORD115* locus) by CRISPR/Cas9 genome editing system in order to define which lncRNA is involved in this long-range regulation. Then, we found that the spliced lncRNA *SNORD116HG* is essential for appropriate expression of the *NDN* & *MAGEL2* genes over long distance. Moreover, we demonstrated that the *SNORD116HG* is co-localized with the *NDN* & *MAGEL2* genes locus by DNA-RNA FISH analyses. Therefore, we suggested that the *SNORD116HG* have a novel and essential role for the high order chromatin dynamics of the *NDN* & *MAGEL2* genes locus. These novel molecular insights into the high order chromatin dynamics should lead to the understanding of lncRNA roles in complex neurodevelopmental disorders.

1899W

Hypermethylation of the FXN gene as a predictive biomarker of drug response for HDAC inhibitor treatment in Friedreich ataxia. L.N. Rodden, K.M. Gilliam, Y.K. Chutake, S.I. Bidichandani. University of Oklahoma Health Sciences Center, Oklahoma City, OK.

Transcriptional silencing caused by the expanded GAA triplet repeat mutation (GAA-TRE) in intron 1 of the *FXN* gene is the primary molecular defect in Friedreich ataxia (FRDA). Rationally-designed *FXN* reactivation strategies, including HDAC inhibitors, are currently being developed, therefore a clearer understanding of the mechanisms of *FXN* gene silencing and of potential determinants of effective gene reactivation are needed for these therapeutic modalities to succeed. We recently identified a new epigenetic silencing signal in FRDA involving hypermethylation of the *FXN* CpG island (CGI) shore. This was seen in disease-relevant tissues, such as the dorsal root ganglia, cerebellum and the heart, and in iPSC-derived FRDA neurons. Cell lines from FRDA patients with atypically low levels of DNA methylation showed higher efficacy of *FXN* gene reactivation via HDAC inhibition, and chemically-induced demethylation potentiated the gene reactivation ability of HDAC inhibition. A prospective study of people with FRDA, involving testing freshly isolated PBMCs for response to ex vivo HDAC inhibitor treatment and level of *FXN* CGI shore methylation indicated that lower CGI shore methylation correlates with superior response to HDAC inhibitor treatment in FRDA. This supports an epigenetic silencing model in FRDA, wherein DNA hypermethylation overlays repressive chromatin, serving as an *epigenetic lock* that impedes *FXN* gene reactivation. *FXN* CGI shore hypermethylation could serve as a predictive biomarker for drug response in FRDA.

1900T

The correlation between the hydroxymethylome and the transcriptome.

D.W. Sant¹, V. Camarena¹, S. Mustafi¹, T.C. Huff¹, Z.M. Wilkes¹, E.R. Martin¹, G. Wang^{1,2,3}. 1) John P. Hussman Institute for Human Genomics, Dr. John T. Macdonald Foundation Department of Human Genetics, Human Genetics, University of Miami, Miami, FL; 2) Bascom Palmer Eye Institute, University of Miami Miller School of Medicine, Miami, FL; 3) Dr. Nasser Ibrahim Al-Rashid Orbital Vision Research Center; University of Miami Miller School of Medicine, Miami, FL.

Hydroxymethylcytosine (5hmC) is a unique epigenetic mark that has been shown to have bidirectional effects on transcription. Stimulation of the TET enzymes increases 5hmC globally and leads to monumental shifts in the transcriptome with both upregulated and downregulated transcripts. How 5hmC achieves this bidirectional effect is still not well understood. Levels of 5hmC in the genebody have previously been found to be positively correlated with transcription, but in the promoter it has been inconsistent between studies. To investigate the relationship between 5hmC and transcription, we performed 5-hydroxymethylcytosine DNA immunoprecipitation (hMeDIP-seq) and transcriptome sequencing (RNA-seq) on cells before and after stimulation of the TET enzymes: ARPE-19 cells were treated with ascorbate (vitamin C) and Schwann cells were treated with either ascorbate or cAMP. All three datasets showed massive shifts in both the hydroxymethylome and the transcriptome. Transcripts were divided into expression quartiles, and higher expression was found to be correlated with an increase in 5hmC in both the gene body and the promoter, but with a decrease in 5hmC at the transcription start site (TSS). When investigating 5hmC levels in nondifferentially expressed transcripts, it was not surprising that the stimulated samples across experiments had higher levels of 5hmC across the gene bodies and promoters, but the 5hmC at the TSS (500 bp region) was actually decreased in all stimulated samples. This increase in 5hmC at the promoters and gene bodies was even greater in the upregulated transcripts across all experiments, and an increase in 5hmC at the TSS was found in the upregulated transcripts in experiments involving ascorbate. Downregulated transcripts showed a greater decrease in RNA with a greater 5hmC increase in the TSS across all three experiments. Results of these experiments demonstrate that changes in 5hmC in the genomic regions of genes may be responsible for a large portion transcriptional changes observed after the activation of the TET enzymes, but is not sufficient to explain the majority of the change in transcription. Instead, it is likely that much of the transcriptional variation is mediated through changes in 5hmC at regulatory elements such as enhancer regions, through Polycomb-mediated silencing, or through inhibition of transcription factor binding.

1901F

DNA methylation changes associated with Parkinson's disease progression: Outcomes from the first longitudinal genome-wide methylation analysis in blood.

A. Henderson-Smith^{1,2}, K. Fisch³, J. Hua⁴, G. Liu⁵, E. Ricciardelli⁶, K. Jepsen⁶, M. Huentelman², G. Stalberg⁵, S. Edland⁷, C.R. Scherzer⁸, T. Dunckley¹, P. Desplats^{7,8}, Harvard Biomarkers Study. 1) Biodesign Institute, Arizona State University, Tempe, Arizona; 2) Neurogenomics Division, Translational Genomics Research Institute, Phoenix, Arizona; 3) Center for Computational Biology & Bioinformatics, Department of Medicine, University of California San Diego; 4) Center for Bioinformatics and Genomics Systems Engineering, Texas A&M Engineering Experiment Station, Texas A&M University; 5) Precision Neurology Program, Harvard Medical School, Brigham & Women's Hospital; 6) Genomics Center, Institute for Genomics Medicine, University of California San Diego; 7) Department of Neuroscience, University of California San Diego; 8) Department of Pathology, University of California San Diego.

Background: Parkinson's Disease (PD) is a common neurodegenerative disorder currently diagnosed through presentation of characteristic movement symptoms. Patients presenting with these symptoms have already undergone significant dopaminergic neuron cell loss. Earlier diagnosis, aided by molecular biomarkers specific to PD, would improve overall patient care. Epigenetics, which is modified by both environmental cues and disease pathophysiology, is emerging as an important component of neurodegenerative diseases, including PD. Alterations to the PD methylome have been reported in epigenome-wide association studies. However, studies have thus reported neither the extent to which methylation changes correlate with disease progression nor the degree to which they are modified by PD medications. **Methods:** We performed a longitudinal epigenome-wide DNA methylation study surveying 850,000 CpG sites in whole blood samples from 197 well-characterized PD patients and 199 control individuals at baseline and 2 years later, using the Infinium Methylation EPIC BeadChip arrays. **Findings:** We identified distinct patterns of methylation in PD cases versus controls at each time-point. Interestingly, specific methylation sites showed consistent changes regardless of sampling time, while others were changed either only at baseline, or only at follow-up, two years later. Additionally, we observed that levodopa and COMT-inhibitors impact DNA methylation, as PD patients that did not received these drugs showed longitudinal changes in methylation in a larger number of sites. **Interpretation:** These results indicate that DNA methylation is dynamic in PD and changes over time during the progression of the disease. To the best of our knowledge this is the first longitudinal epigenome-wide methylation analysis for Parkinson's disease; and the first appraisal of the effects of dopamine replacement therapy on the blood methylome.

1902W

A comprehensive approach to study the epigenome of monocytes in persons with Parkinson disease and matched healthy controls. J. Zhu, T. Carlile, D. Sangurdekar, K.D. Nguyen, H. McLaughlin, C. Sun, M.G. Weisz, O. Goldstein, A. Mirelman, J.M. Cedarbaum, N. Giladi, A. Orr-Urtreger, MA. Penny, S. John, MD. Lalioti. Biogen, 225 Binney Street, Cambridge, MA, 02142, US.

Understanding disease pathophysiology and progression paired with highly predictive biomarkers are crucial for finding effective treatments for complex diseases such as Parkinson's (PD). There is increasing evidence that chronic inflammation mediated by microglial cells, the resident immune cell in the brain, is a fundamental process contributing to the death of dopamine-producing neurons in PD. Because, brain tissue is inaccessible, monocytes from blood have been considered as an accessible proxy for microglia, and changes in their transcriptome have been described. The dynamics of the epigenome, including methylation and chromatin remodeling have profound effects on the transcriptome. The objective of this research is to perform a comprehensive study of the epigenome of monocytes from PD patients and healthy individuals to identify disease signature using gene expression (RNA-Seq), chromatin accessibility (ATAC-Seq), and methylation (MethylationEPIC BeadChip) assays. By contrasting chromatin accessibility and gene expression, we can obtain insights into non-coding regulatory elements implicated in the disease state. The pilot phase described here included 100 Ashkenazi Jewish individuals either manifesting or genetically at-risk for PD, or healthy controls. We inferred correlation between gene expression, chromatin accessibility and methylation across genes and at promoters. Positive within-subject correlation between gene expression and chromatin accessibility at promoter regions, including 1kb flanking the transcription start sites, was observed in 98% of individual samples. Looking across all the samples, we found that genes showing concordance between gene expression and chromatin accessibility (Spearman's correlation coefficients range from 0.2 to 0.6, $p < 0.05$) are enriched in cell movement of mononuclear leukocytes related function ($p < 0.001$) and associated with inflammatory response, neurological disease, skeletal and muscular disorders. These data are indicative of underlying environmental or pathological variability leading to differences in monocyte activation. In addition, approximately 1% of genes had significant correlation (FDR $< 10\%$) between promoter methylation and gene expression across samples. We suggest that monocyte signature and the analysis pipelines developed as part of this study can be universally applied to other diseases where a change in the monocyte transcriptome and epigenome has been established or suspected.

1903T

Unsupervised clustering of spatially resolved single nucleus chromatin accessibility reveals cell-type and region specific transcriptional control in the mouse primary motor cortex. R. Fang^{1,2}, S. Preiss³, X. Wang³, X. Hou³, J. Gu³, B. Chen¹, J. Lucero⁴, M. Behrens⁴, E. Mukamel⁵, J. Ecker^{4,6}, B. Ren^{1,3}. 1) Ludwig Institute for Cancer Research, La Jolla, CA, USA; 2) Bioinformatics and Systems Biology Graduate Program, University of California San Diego, La Jolla, CA, USA; 3) Center for Epigenomics, Department of Cellular and Molecular Medicine, University of California, San Diego, School of Medicine, La Jolla, CA, USA; 4) Salk Institute for Biological Studies, La Jolla, CA 92037, USA; 5) Department of Cognitive Science, University of California, San Diego, La Jolla, CA 92037, USA; 6) Howard Hughes Medical Institute, The Salk Institute for Biological Studies, La Jolla, CA, 92037, USA.

Analysis of chromatin landscape across single cells can resolve heterogeneity and reveal cell-type specific regulatory programs in a complex tissue, but intrinsic sparsity of the data poses significant challenges in the data analysis. Here we introduce Snap, a R package that facilitates pre-processing, quality control, normalization, clustering and visualization of single-nucleus epigenomics data. Snap clusters cells based on chromatin variability in an unsupervised manner and identifies clusters through a highly robust consensus approach. Snap outperformed previous methods by rapidly identifying cell types while eliminating the experimental bias. As a part of BRAIN Initiative, we generated over 47,000 high-quality single-cell chromatin accessibility profiles from two areas of the primary motor cortex (MOp) of adult mouse brain with spatial resolution. Snap identified over 50 distinctive cell populations corresponding to different cellular functions in the MOp. Although the cell types are shared between different areas, excitatory neurons and oligodendrocyte are variable in a striking manner, suggesting regional specificity of transcriptional control. Finally, Snap successfully clusters multiple single-nucleus assays, providing a general framework for single-cell analysis and a possibility of integrative analysis of multi-omics data.

1904F

Epigenomic convergence of genetic and immune risk factors in autism brain. A. Vogel Ciernia, B. Laufer, K. Dunaway, C. Mordaunt, R. Coulson, D. Yasui, J. LaSalle. University of California, Davis, Davis, CA.

Autism spectrum disorders (ASD) are characterized by impairments in social communication and increased repetitive behaviors. ASD etiology is complex, involving multiple genetic and environmental risks. Epigenetic modifications are poised at the interface between genes and environment and are predicted to reveal insight into the gene networks, cell types, and developmental timing of ASD etiology. Here, whole-genome bisulfite sequencing (WGBS) was used to examine DNA methylation in ASD and control frontal cortex samples. Systems biology approaches were leveraged to integrate methylation differences with relevant genomic datasets, revealing ASD-specific differentially methylated regions (DMRs) are significantly enriched for known neuronal and microglial regulatory elements, including cell-type-specific enhancers and transcription factor binding sites. ASD DMRs were also significantly enriched for known ASD genetic risk factors, including both common inherited and rare de novo variants. Weighted gene co-expression network analysis (WGCNA) revealed enrichment of ASD DMRs within developmental expression modules of brain and isolated microglia. Microglial modules identified dysregulated genes in maternal immune activation models of ASD. Weighted gene body co-methylation network analysis revealed a module characterized by hypomethylation of clustered protocadherin genes. Together, these results demonstrate an epigenomic signature of ASD in frontal cortex shared with known genetic and immune etiological risk. Epigenomic insights into cell types and gene pathways will aid in defining therapeutic targets and early biomarkers at the interface of genetic and environmental ASD risk factors.

1905W

DNA methylation in the hippocampus contributes to anxious temperament in young primates. R.S. Alisch^{1,2}, L.A. Papale^{1,2}, J.A. Oler¹, A.S. Fox^{3,4}, D.A. French¹, P.H. Roseboom¹, N.H. Kalin¹, P. Chopra⁵. 1) Department of Psychiatry, University of Wisconsin, Madison, WI; 2) Department of Neurological Surgery, University of Wisconsin, Madison, WI; 3) Department of Psychology, University of California, Davis, CA; 4) California National Primate Research Center, University of California, Davis, CA; 5) Department of Human Genetics, Emory University School of Medicine, Atlanta, GA.

Considerable evidence demonstrates that children with an anxious temperament (AT) are particularly sensitive to new social experiences, showing increased freezing behavior, decreased communications, and increased pituitary-adrenal and autonomic activity. However, despite substantial knowledge of these behavioral phenotypes, the molecular mechanisms governing them remain unclear; in part due to the difficulty in accessing human brain tissues for molecular analyses. Recently, we validated a nonhuman primate model of AT that demonstrates structural and functional similarities between rhesus monkeys and humans in the neural circuits mediating increased freezing, decreased vocalizations, and increased cortisol levels. We further identified that AT-related brain metabolism in the central nucleus of the amygdala (Ce) and the anterior hippocampus (aHIP) significantly differ in their heritability (aHIP > Ce), suggesting that these structures are differentially governed by genetic and environmental influences. Our initial studies of the Ce revealed AT-related differentially methylated loci linked to altered gene regulation. Here we built upon these findings and used whole genome bisulfite sequencing (WGBS) to examine an average of 25.3 million CpG dinucleotides in genomic DNA from the hippocampus of 71 monkeys (including 23 females) and found 645 differentially methylated regions (DMRs) associated with AT severity (permutation P -value < 0.01). Annotation of these DMRs to genomic structures revealed 515 genes that are enriched for neuronal ontological functions, such as synapse assembly and neuron development. Comparison of these genes to the genes previously found in the Ce revealed a significant overlap (P -value < 0.05), indicating that common AT-related epigenetic disruptions occur in these two brain structures. Finally, a significant overlap (P -value < 0.05) also was found between these differentially methylated genes from the monkey brain and AT-related differentially methylated genes reported in human blood, suggesting that blood may be an accessible tissue of value in the identification of differential methylation associated with the risk to develop trait-like anxiety. Thus, we have generated WGBS data on genomic DNA from the blood of these same 71 monkeys and analysis is underway to determine the utility of blood as a viable surrogate to brain tissue, providing novel diagnostic, prognostic, and modifiable therapeutic targets of human anxiety.

1906T

Cell type specific methylome-wide studies in brain and blood implicate inflammation and neurodegenerative processes in major depressive disorder. R.F. Chan¹, A.A. Shabalina¹, B. Dean^{2,3}, M. Zhao¹, G. van Grootheest⁴, B.W.J.H. Penninx⁵, G. Turecki⁵, K.A. Aberg¹, E.J.C.G. van den Oord¹. 1) Center for Biomarker Research and Precision Medicine, Virginia Commonwealth University, Richmond, VA, USA; 2) The Molecular Psychiatry Laboratory, The Florey Institute of Neuroscience and Mental Health, Parkville, Victoria, Australia; 3) Centre for Mental Health, Swinburne University of Technology, Hawthorne, Victoria, Australia; 4) Department of Psychiatry, Amsterdam Neuroscience, VU University Medical Center, GGZ inGeest, Amsterdam, The Netherlands; 5) Douglas Mental Health University Institute and McGill University, Montréal, Québec, Canada.

We sequenced the CpG methylome in three independent collections of human post-mortem brain samples (N=205) and blood samples (N=1,132) of controls and subjects with MDD. In addition to analyzing bulk tissue, we used cell type specific "reference" methylomes paired with a statistical deconvolution approach to conduct methylome-wide association studies (MWAS) in constituent populations of neurons/glia and granulocytes/T-cells/B-cells/monocytes. Cell type specific MWAS yielded case-control differences that were not readily detectable in heterogeneous bulk tissues. Using a round-robin rotation schedule, where two of the three brain collections were used to discover sites for replication in the third independent collection, top sites from the neuron/glia discovery MWASs consistently replicated as a group with enrichment P-values $4.5 \times 10^{-4}/3.9 \times 10^{-3}$, $6.5 \times 10^{-4}/9.8 \times 10^{-4}$, and $<1 \times 10^{-4}/1.3 \times 10^{-3}$. The top replicating finding in neurons was located within *TRIM1* (P-values = 0.0014/0.0025/0.007; combined = 4.4×10^{-6}), a negative regulator of synaptic plasticity. Results for neurons (384 CpGs) were further overrepresented for pathways related to neuronal cell death and neurodegeneration. Top findings for glia (85 CpGs) were overrepresented for inflammatory pathways linked to MDD such as Toll-like receptor and cytokine signaling. Cell type specific analyses in blood identified 372 differentially methylated loci (FDR 0.1) in monocytes and 18 loci in T-cells. Top findings for monocytes, e.g. *ITPR2*, *SHANK2*, and *SLC1A1* (P-values = 1.50×10^{-9} to 7.31×10^{-7}), were overrepresented for glutamatergic signaling pathways (that are both immunoregulatory and neuromodulatory) and enriched at SNPs associated with allergy, asthma, infection susceptibility, and neurodegenerative disorders. In supplementary analyses overlap was observed between brain and blood cell types. For example, a *SHANK2* CpG was implicated in neurons and monocytes with shared direction of effect in both (combined P-value = 1.05×10^{-6}), suggesting that disease affects different tissues through like processes. In summary, epigenomic deconvolution analysis uncovered cell type specific methylation differences associated with MDD. Although individual findings were largely cell type specific, related biological processes were implicated across cell types and even across tissues. Thus, the involvement of inflammatory and neurotoxic processes in MDD are robust findings that impacts neurons, glia, and monocytes.

1907F

Neural cell adhesion molecule 1 (NCAM-1) is associated to long-term memory in nematodes and humans. P. Mastrandreas¹, V. Vukojevic^{1,3}, T. Elbert¹, I.T. Kolassa⁴, V. Freytag¹, F. Peter¹, P. Demougin¹, J.F. deQuervain^{2,3}, A. Papassotiropoulos^{1,3}, A. Stetak^{1,3}. 1) University of Basel, Department of Psychology, Division of Molecular Neuroscience, Birmannsgasse 8, 4055 Basel, Switzerland; 2) University of Basel, Department of Psychology, Division of Cognitive Neuroscience, Birmannsgasse 8, 4055 Basel, Switzerland; 3) University of Basel, University Psychiatric Clinics, Wilhelm Klein-Strasse 27, 4055 Basel, Switzerland; 4) Clinical and Biological Psychology, Institute for Psychology and Education, University of Ulm, 89069 Ulm, Germany.

The neural cell adhesion molecule 1 (NCAM1) has been implicated in several brain-related biological processes, including neuronal migration, axonal branching, fasciculation, and synaptogenesis, with a pivotal role in synaptic plasticity. Here, using an interdisciplinary approach we investigated the evolutionary conserved role of neural cell adhesion molecule 1 (*ncam-1*) in learning and memory. We conducted a large-scale transcriptomic study examining the temporal expression profile of different genes during *C. elegans* long-term memory, and observed sustained changes in *ncam-1* expression up to 24 hours upon aversive olfactory conditioning. Loss of *ncam-1* function selectively impaired long-term memory in a negative but not in positive olfactory associative memory paradigm, without causing acquisition, sensory, motor or short-term memory deficits. Expression of the *C. elegans* or human NCAM1 in nematodes fully rescued loss of *C. elegans ncam-1* gene function. Tagging of NCAM-1 using CRISPR/Cas9 revealed that the protein expression occurs throughout the developmental life cycle of the worm. Additionally, confocal imaging of an established GLR-1::GFP line, revealed no defects in the post-synaptic structures of *ncam-1* (*lf*) worms, suggesting that the effects on aversive long-term memory can be attributed to deficits in the function of the protein and not due to developmental defects during synapse formation. Considering the conserved role of NCAM1 in higher complex organisms, we next expanded our *C. elegans* findings in human, using methylation data across three independent samples. In two samples of European descent, we showed a negative correlation between DNA methylation of the NCAM1 promoter and aversive memory (i.e., recognition of negative pictures). Moreover, in a population of genocide survivors of African descent, DNA methylation at an alternative promoter of the NCAM1 gene was inversely correlated with symptoms of posttraumatic stress disorder (PTSD), including traumatic aversive memory. Taken together, our results support a conserved role of NCAM1 in aversive associative memory from nematodes to humans and might, ultimately, prove to be helpful in elucidating novel therapies in the treatment of memory-related disorders.

1908W

A comprehensive study of epigenetic signatures associated with PTSD in two independent African populations of severely traumatized individuals. V. Vukojevic^{1,2,3}, A. Milnik^{1,3,4}, T. Elbert⁵, I. Tatjana Kolassa⁶, S. Wilker⁶, D. J.-F. de Quervain^{3,4,7}, A. Papassotiropoulos^{1,2,3,4}. 1) University of Basel, Department of Psychology, Division of Molecular Neuroscience, Birmannsgasse 8, 4055 Basel, Switzerland; 2) University of Basel, Department Biozentrum, Life Sciences Training Facility, Klingelbergstrasse 50/70, 4056 Basel, Switzerland; 3) Transfaculty Research Platform Molecular and Cognitive Neurosciences, University of Basel, CH-4055 Basel, Switzerland; 4) University of Basel, University Psychiatric Clinics, Wilhelm Klein-Strasse 27, 4055 Basel, Switzerland; 5) Clinical Psychology and Neuropsychology, University of Konstanz, D-78464 Konstanz, Germany; 6) Clinical & Biological Psychology, Institute of Psychology and Education, University of Ulm, D-89069 Ulm, Germany; 7) University of Basel, Department of Psychology, Division of Cognitive Neuroscience, Birmannsgasse 8, 4055 Basel, Switzerland.

Post-traumatic stress disorder (PTSD) is a chronic pathological response to life-threatening, i.e., traumatic experiences. Intrusive memories of such traumatic events, avoidance and hyperarousal are the core symptoms of the disorder. Furthermore, a complex interplay between genetic factors, epigenetic mechanisms and environment plays a significant role in the pathogenesis of PTSD. We have previously shown that DNA methylation of the glucocorticoid receptor gene (*NR3C1*) promoter is linked to aversive memories and PTSD risk in male survivors of the Rwandan genocide (N=152). Furthermore, DNA methylation at the same *NR3C1* promoter site was associated with picture recognition and differences in recognition-related brain activity in healthy young men (Vukojevic *et al*, 2014). In the current study, we extended our previous work to a genome-wide methylation analysis in a larger Rwandan sample (N=380) assessed with Illumina Human Methylation 450k array. Additionally, we have examined another population of traumatized individuals from Uganda (N=480) with the Illumina Human Methylation EPIC array. In both samples, we investigated the association between DNA methylation at *NR3C1*, aversive memories and PTSD risk. Next to reassessing our previous findings, we explored the shared epigenetic signatures of lifetime PTSD risk and symptomatology across two cohorts on the epigenome-wide level. Interestingly, the common genes were enriched in gene pathways related to immune response, regulation of steroid hormones and energy metabolism. The present study suggests the existence of global and gene-specific epigenetic mechanisms associated with PTSD. Observed epigenetic associations with PTSD risk and symptomatology support the observation that this mental disorder results in altered immunological, metabolic and neural regulations. Therefore, the findings of the proposed study may add to the understanding the mechanisms of PTSD and could thereby help to improve the development of targeted clinical interventions.

1909T

Neuron and glia specific methylome-wide association studies of schizophrenia post-mortem brain samples identify unique loci not detected in bulk tissue. K.A. Aberg¹, R.F. Chan¹, M. Zhao¹, L.Y. Xie¹, G. Turecki², A.A. Shabalina^{1,3}, B. Dean^{4,5}, E.J.C.G. van den Oord¹. 1) Center for Biomarker Research and Precision Medicine, Virginia Commonwealth University, Richmond, VA; 2) Douglas Mental Health University Institute and McGill University, Montréal, Québec, Canada; 3) Department of Psychiatry, University of Utah, Salt Lake City, UT, USA; 4) The Molecular Psychiatry Laboratory, Florey Institute of Neuroscience and Mental Health, Parkville, Victoria, Australia; 5) Centre for Mental Health, Swinburne University, Hawthorn, Victoria, Australia.

We used fluorescence-activated cell sorting to isolate NeuN+ (neurons) and the NeuN- (mainly glia) nuclear fractions from five post-mortem prefrontal cortex samples. For each of the isolated cell fractions DNA was extracted and a methylome-wide profile generated. Subsequently, the neuron and glia specific reference methylomes were used to estimate the proportions of neurons and glia in bulk brain (BA10) methylomes from 164 schizophrenia cases and 160 controls. Next, we used the RaMWAS software to perform a methylome-wide association study (MWAS) for schizophrenia for bulk, and applied a statistical deconvolution method where the estimated cell type proportions were used to perform cell type specific MWAS in neurons and glia separately. Results identified three genome-wide significant (q-value<0.1) loci in bulk and glia, and additional suggestive findings in all three MWASs. It should be noted that the cell type specific top findings were not present among the top bulk findings, supporting the notion that cell type specific effects cancel out in the bulk. The lambdas for the MWASs were 1.058/1.007/1.042 for bulk/glia/neurons, respectively. For each analysis we performed 1,000 MWASs after permuting case-control status. The average and the median lambdas from the permuted MWASs were 1.00 (95% confidence interval: 0.98-1.04) in all instances, suggesting the tests behaved as expected under the null hypothesis, and that the observed lambdas likely reflect true associations. Furthermore, to explore if many markers together contribute more to the disease association than any single marker individually we used a machine learning algorithm to combine associated sites into a single risk score. When screening the top 1,000 markers from the bulk MWAS, we detected a Pearson correlation of 0.26 ($P=8.4 \times 10^{-7}$) between methylation-predicted and actual disease status. Thus, the variation in methylation levels of these markers explained 6.9% of the schizophrenia disease status. The explained variance did not further increase when additional markers were added, suggesting that the majority of independent effects are located among the top 1,000 MWAS findings. In conclusion, cell type specific MWAS identified associations not detectable in bulk analysis. These findings emphasizes the importance to further refine the reference panels so that relevant biological processes, occurring in unique subsets of cells, can be appropriately assayed.

1910F

LCM-Seq: Single cell-type transcriptomic profiling in post-mortem brain of abused suicides. D. Almeida^{1,2}, G.G. Chen¹, J.F. Theroux¹, Z. Aouabed¹, M.A. Davoli¹, N. Mechawar^{2,3}, G. Turecki^{1,2,3}. 1) McGill Group for Suicide Studies, Montreal, Quebec, Canada; 2) Integrated Program in Neuroscience, McGill University, Montreal, QC, Canada; 3) Department of Psychiatry, McGill University, Montreal, QC, Canada.

Background: The transcriptome of a cell constitutes an essential piece of cellular identity and accounts for the multifaceted complexity and heterogeneity of cell types within the mammalian brain. Each discrete cellular population is differentially influenced by extrinsic signals from their local environments and neighbouring cells. Thus, while some studies have investigated transcriptomic alterations underlying the neurobiology of childhood maltreatment and suicide, the use of bulk-tissue homogenates have masked their ability to determine cell-type specific molecular dysfunctions. Here we employ a cell type specific investigation of transcriptomic alterations, in prefrontal layer V pyramidal neurons, associated specifically with a history of childhood abuse (CA). **Methods:** Laser captured microdissection (LCM) was used to isolate prefrontal (BA 10) layer V pyramidal neurons from post-mortem human brain. Subject groups included individuals who died by suicide with and without a history of severe CA and non-psychiatric controls. RNA sequencing libraries were constructed using the SMARTseq v4 kit, to generate full length cDNA libraries, followed by Illumina's Nextera XT DNA library preparation kit. The abused suicide and control groups were used for sequencing. Fifty libraries were sequenced on the HiSeq 4000 using PE 100bp sequencing at a depth of ~40M reads per subject. **Results:** We achieved a mapping efficiency of ~80% and captured a wide distribution of transcripts. Differential gene expression analysis revealed significant (FDR <0.10) dysregulation of a variety of protein coding transcripts in abused suicides compared to controls. Gene Set Enrichment Analysis (GSEA) showed significant enrichment of genes with high CpG density promoters in human brain. Weighted Gene Co-Expression Network Analysis (WGCNA) identified two modules associated specifically with CA. Gene ontology (GO) analysis on these modules revealed that they include genes involved in neuronal structure and signalling. **Conclusions:** By employing LCM followed by RNA sequencing we were able to uncover cell-type specific changes associated with a history of CA. Furthermore, these results open up an interesting avenue to explore whether cell-type specific epigenetic regulation, through DNA methylation, might explain our expression findings. .

1911W

Ethnic and sex-specific induction of G-protein coupled receptor 15 expression in smokers. A. Andersen, R. Philibert. Psychiatry, University of Iowa, Iowa City, IA.

Smoking is responsible for nearly half a million deaths annually in the United States, in part due to an increased risk of cardiovascular disease, diabetes, and chronic obstructive pulmonary disease. Each of these disease states is characterized by systemic inflammation, but the links between smoking and inflammation remain unclear. This may be because serological examinations of smoking-associated inflammation to date have failed to consider cell-based changes in immune functioning simultaneously. Incorporation of epigenetic methods into serological approaches may provide a method to systematically assess these variables. Recent work has demonstrated that demethylation of the CpG site cg19859270 reflects an increase in a specific T cell subtype expressing *G protein-coupled receptor 15* (GPR15), a chemokine receptor expressed primarily in the gut that has been linked with ulcerative colitis, a smoking-associated disease. Although the specific role of the GPR15 in the gut is not fully understood, increased homing of this T cell population to the gut may produce subtle changes in the gut microbiome that in turn promote inflammation. To translate this finding into a tool that can be employed systematically, our laboratory has developed a digital PCR assay for cg19859270 and validated it against flow cytometrically measured expression of GPR15 in a cohort of 62 young adults of mixed ethnicity with a high rate of smoking. GPR15 expression in CD3+/CD4+ cells was strongly associated with both cg19859270 methylation (R-squared = 0.216, P<0.001) and cumulative smoke exposure, as indicated by methylation of cg05575921 (R-squared = 0.443, P=2.19E-09), an established epigenetic biomarker for smoking. Multiple linear regression models including interactions between sex, ethnicity, cg05575921 methylation, and cg19859270 methylation produced significant improvements in predicting GPR15 expression (R-squared = 0.650, P = 3.32E-12). No correlation was found between GPR15 expression or smoke exposure and serum cytokine levels in this young cohort, suggesting that changes in GPR15 expression may precede the development of inflammation and complex medical illness in smokers. Going forward, the ability to robustly predict GPR15+ expression in a genetically contextual manner using digital PCR methods will allow for further explorations of the impact of this specific T cell population on inflammation and disease progression in older cohorts.

1912T

DNA methylation profiles of monozygotic twins discordant for ADHD by Illumina MethylationEPIC BeadChip microarray. T.V.M.M. Costa^{1,5}, J.G. Damasceno¹, F.A. Marchi², F.A.R. Madia¹, E.A. Zanardo¹, G.M. Novo-Filho¹, M.M. Montenegro^{1,3}, A.T. Dias¹, A.M. Nascimento^{1,3}, M. Rocha¹, C.A. Kim³, V. Schuch⁴, C.B. Mello⁴, M. Muszkat⁴, L.D. Kulikowski^{1,3}. 1) Laboratório de Citogenômica, Departamento de Patologia, Faculdade de Medicina da USP, São Paulo, Brasil; 2) Laboratório NeoGene, AC Camargo Cancer Center, São Paulo, Brasil; 3) Laboratório de Citogenômica, Unidade de Genética, Departamento de Pediatria, Instituto da Criança, HC/FMUSP, São Paulo, Brasil; 4) Núcleo de Atendimento Neuropsicológico Infantil Interdisciplinar (NANI), Departamento de Psicobiologia da Escola Paulista de Medicina, UNIFESP, São Paulo, Brasil; 5) Instituto Federal de Educação, Ciência e Tecnologia do Maranhão (IFMA), Campus Barreirinhas, Barreirinhas, Brasil.

Attention deficit hyperactivity disorder (ADHD) is one of the most common neurobehavioral disorder diagnosed in childhood and is characterized by inadequate levels of hyperactivity/impulsivity and/or inattention. ADHD begins in childhood and affects around 5% of school children worldwide and, in most cases (50-80%), the disorder persists during adolescence and adulthood. Its high heritability (estimated at 76%) suggests that genetic factors strongly influence the etiology of ADHD. However, the genetic architecture of ADHD is not clear. Behavioral genetic studies have demonstrated that genetic influences play a role in the etiology of ADHD. In addition, some environmental factors are epidemiologically associated with ADHD. Therefore, it is very important to study whether epigenetic factors are related to the etiology of ADHD. Thus, performing the DNA methylation profiles of monozygotic twins are an important tool for studying the role of epigenetic factors in ADHD. Diagnosis was performed according to Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR). DNA was extracted from blood lymphocytes of one ADHD patient and his twin brother with neurotypical development (two boys, ages 12). Then, DNA methylation profiles were performed using Illumina MethylationEPIC BeadChip. Arrays data were treated and analyzed using specific packages in the R environment. We calculate the difference between the β values ($\Delta\beta$) of all probes of both twins (with and without ADHD). Limma package was used to obtain significant probes considering $p < 0.05$. The probes were annotated according to the data provided by Illumina using the genome hg19 reference. We selected the probes with $\Delta\beta > 0.1$ that were differentially methylated when comparing the patient with ADHD to the twin. Thus, our results suggest that DNA methylation profiles may present significant differences between monozygotic twins discordant for ADHD. Therefore, we emphasize the importance of investigating methylation profiles in more cases of ADHD discordant twins, in order to verify if, in fact, there are differences in the methylation between profiles and whether these differences are relevant for the development of ADHD.

1913F

Neural-derived biomarkers for antidepressant response from plasma exosomes. S. Saeedi^{1,2}, C. Nagy², J.F. Theroux², M. Wakid^{2,3}, N. Mechawar^{2,3}, G. Turecki^{2,4}. 1) Department of Human Genetics, McGill University, Montreal, QC, Canada; 2) McGill Group for Suicide Studies, Douglas Mental Health University Institute, Montreal, Quebec, Canada; 3) Integrated Program in Neuroscience, McGill University, Montreal, Quebec, Canada; 4) Department of Psychiatry, McGill University, Montreal, Quebec, Canada.

Background: Major depressive disorder (MDD) is a common psychiatric disorder affecting millions of people worldwide. The most common treatment for MDD is antidepressant drug therapy (ADT), with many options available for treatment. Decisions on treatment for MDD are up to the discretion of the physician, and are often highly subjective. Furthermore, 60% of patients do not respond to first trials with antidepressants. As a result, biomarker discovery has become a prominent aspect of ADT. However, the majority of the research in this field is limited to peripheral tissue, leaving questions concerning the relevance of these findings in the context of psychiatric disease. Exosomes may provide a solution to the disconnect between the source of material and disease. Given that exosomes can freely cross the blood brain barrier, neural-derived exosomes (NDE) found in plasma can provide information regarding central changes resulting from ADT response. MicroRNA (miRNA) represent exosomal cargo which are presumed to exert an influence on recipient cells. Differential miRNA profiles can potentially act as predictive biomarkers, and provide mechanistic insight into changes occurring as a result of ADT response. **Methods:** This pilot study uses plasma from 10 controls and 10 patients with MDD (5 ADT responders, and 5 non-responders). Exosomes were isolated using a size exclusion column from Izon Science (Christchurch, New Zealand). Each isolation was divided to produce a "whole exosome" fraction and an immunoprecipitated "(NDE)" fraction using neural marker L1CAM. Quantitation and size determination was done using Tunable Resistive Pulse Sensing (TRPS) on the qNano gold. RNA was extracted from L1CAM+ fraction and it's depleted supernatant. Libraries were made following the 4N-small RNA-Seq (Galas) protocol and sequenced on the Illumina platform. **Results:** Results suggest that NDE exosomes are smaller than the full pool of exosomes. Also, exosomes from depressed patients (both responders and non-responders) are significantly smaller than controls in both full and neural enriched fractions. We have also identified a group of miRNAs that are highly enriched in NDE and that overlap with miRNAs present in brain. Differential analysis shows many hits for follow-up. **Conclusions:** Isolating NDE from plasma provides a valuable resource for biomarker discovery in MDD. Our ongoing work aims to provide a neural miRNA profile in MDD, and profiles of drug response and non-response.

1914W

Genome-wide profiling of DNA methylome and transcriptome in peripheral blood monocytes for major depression: A monozygotic discordant twin study. Y. Zhu¹, E. Strachan², E. Fowler², T. Bacus³, P. Roy-Byrne², J. Zhao¹. 1) Department of Epidemiology, University of Florida, Gainesville, FL; 2) Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA; 3) Department of Pediatrics, University of Washington, Seattle, WA.

Major depressive disorder (MDD) is one of the most prevalent mental disorders. Epigenetic factors play an important role in MDD, but the specific genes or genomic regions associated with MDD remain largely unknown. Here we report findings from a genome-wide profiling of DNA methylation using the Infinium MethylationEPIC BeadChip in peripheral blood monocytes from 79 monozygotic (MZ) discordant twin pairs. Gene expression was measured by RNA-seq in monocytes of the same twins. We identified 39 differentially methylated regions (DMRs) harboring putative genes associated with lifetime history of MDD. Integrative DNA methylome and transcriptome analysis revealed both negative and positive correlations between DNA methylation and gene expression, although there appears to be a trend that promoter DNA methylation is largely negatively correlated with gene expression. Network analysis revealed distinct modules enriched in signaling pathways related to neuron apoptosis, stress-activated protein kinase signaling, insulin regulation, and mTOR signaling. Together, our results revealed a critical role of altered DNA methylation in depression and identified novel genes and biological pathways underlying MDD pathogenesis. Moreover, our study highlights the utility of using peripheral blood epigenetic markers, and demonstrates that an MZ discordant co-twin control design could aid the discovery of novel DNA methylation signals, even in a relatively small sample size. If validated, the newly identified genes/pathways may serve as novel therapeutic targets for MDD and related disorders.

1915T

Molecular changes that drive abnormal cortical development following mid-gestational Poly(I:C)-mediated maternal immune activation. C.P. Canales¹, M.L. Estes¹, I. Zdilari¹, L. Su-Feher¹, S. Cameron¹, J.P. Aboubechara¹, R. Catta-Preta¹, K. Lim¹, T. Stradleigh¹, D. Van Der List¹, L. Haapanen², K. Farrelly¹, J. Van De Water², A.K. McAllister¹, A.S. Nord¹. 1) Center for Neuroscience, University of California Davis, Davis, CA; 2) Div. of Rheumatology/Allergy and Clin. Immunology, University of California Davis, Davis, CA.

Maternal immune activation (MIA) has emerged as risk factor for neurodevelopmental disorders (NDDs), including autism and schizophrenia. MIA is among the strongest environmental risk factors and investigating the molecular mechanisms underlying pathology in offspring following MIA is crucial for understanding gene-environment interactions associated with neurodevelopmental disorders. Here, we applied transcriptional and epigenomic profiling to identify changes in fetal mouse cortex across a time course following mid-gestational MIA via maternal poly(I:C) injection at E12.5. To reduce variability in the experimental model, we established the effective dosage of Poly(I:C) to ensure induction of reproducible levels of MIA and disease-relevant changes in offspring. We then performed RNA-seq in cortical extracts from E12.5, E14.5 and E17.5 from MIA and control offspring and identified strong transient transcriptional signatures and developmental changes in fetal cortex. These changes included an initial acute signature that correlates with activation of genes associated with stress response pathways in the fetal brain, followed by alterations in proliferation, neuronal differentiation, astrogenesis, synaptogenesis and cortical lamination that emerged at E14.5 and peaked at E17.5. We also looked at epigenetic signatures at E14.5 by ChIP-seq analysis and found changes in histone modifications linking transcriptomic and epigenomic changes in fetal brain. To validate transcriptional signatures, we used immunohistochemistry at E17.5, examining several candidates identified by RNA-seq. Anatomical and protein expression changes in E17.5 cerebral cortex were concordant with signatures present in the RNA data, including reduced proliferative populations (i.e. Ki67, Pax6), altered cortical lamination patterns (Tbr1, Ctip2) and precocious astrogenesis (GFAP), with no evidence of previously reported cortical focal dysplasia (Satb2, Cux1). In summary, these transcriptomic signatures validated at the protein level provide novel insight into the molecular and developmental brain pathologies linking MIA and neurodevelopmental sequelae.

1916F

Sex hormones regulate *SHANK* gene expression. A. Eltokhi^{1,2}, S. Berkel¹, H. Fröhlich¹, D. Gonzalez¹, R. Rafiullah¹, R. Sprengel², G. Rappold¹. 1) Institute of Human Genetics, Heidelberg, Germany; 2) Max Planck Institute for Medical Research, Heidelberg, Germany.

Autism spectrum disorders (ASD) have a higher prevalence in male individuals compared to females, with a ratio of affected boys compared to girls of 4 : 1 for ASD and 11 : 1 for Asperger syndrome. Mutations in the *SHANK* genes (comprising *SHANK1*, *SHANK2* and *SHANK3*), coding for postsynaptic scaffolding proteins, have been tightly associated with ASD. As early brain development is strongly influenced by sex hormones, we investigated the effect of dihydrotestosterone and 17 β -estradiol on *SHANK* expression in the human neuroblastoma cell model, SH-SY5Y. Both sex hormones increased the expression of all three *SHANK* genes, which could be effectively blocked by androgen and estrogen receptor antagonists. In neuron-specific androgen receptor knock-out mice (ArNesCre), we found a nominal significant reduction of all *SHANK* genes at postnatal day 7.5 in the cortex. In the developing cortex of wild-type mice, a significant sex-differential expression was found for *Shank1* and *Shank3* at embryonic day 17.5 and for *Shank1* at postnatal day 1.5. This is of special interest as heterozygous *SHANK1* deletions showed a reduced penetrance in female individuals and only segregated in male carriers with high functioning autism. **Together, we could show that *SHANK* expression is influenced by sex hormones and is different between developing cortices of male and female mice, providing a possible mechanism underlying the sex bias in *SHANK*-associated ASD.**

1917W

Parallel enhancer analysis in mouse brain to characterize regulatory variants in development and disease. J.L. Haigh¹, J. Lambert¹, L. Su-Fehler¹, K.J. Lim¹, S.J. Morse¹, I. Zdilar¹, L.C. Byrne², A.S. Nord¹. 1) University of California, Davis, CA; 2) University of Pittsburgh, Pittsburgh, PA.

Enhancers are cis-regulatory elements with the capacity to promote gene expression. Enhancer activity is believed to underlie the differential expression patterns across neurodevelopment, between cell types, and in response to activity-dependent and neuroimmune signaling in the brain. Mapping the activity of enhancers with spatiotemporal and condition-dependent specificity will be critical to understanding gene regulatory networks and their role in human disease. Recent advances in massively parallel reporter assays (MPRAs) have enabled the functional characterization of enhancers via comparing RNA-based activity readout with genomic DNA delivery control. We implemented an MPRA strategy for in vivo delivery into the mouse brain. A pilot library of genomic enhancer candidate sequences containing common non-coding sequence variants associated with epilepsy and schizophrenia was developed. The enhancer library was delivered to neonatal mouse brain via adeno-associated virus (AAV), and genomic DNA and total RNA collected one week later at P7. Results from preliminary studies suggest this method is able to identify sequences capable of acting as enhancers in in vivo mouse brain based on transcriptional activity. We are currently validating the in vivo activity of individual disease-relevant enhancers by AAV delivery to the brain followed by immunohistochemistry and confocal microscopy to assess expression patterns in cerebral cortex. Having established the utility of the MPRA approach in vivo in brain, we are additionally testing this approach to assay allele-specific changes in enhancer activity linked with disease-associated sequence variation. These methods will allow rapid screening of libraries of DNA sequences for enhancer activity in vivo, with the potential to identify whether disease-associated variants contribute to altered gene expression with regional, temporal, and cell-type specificity in the brain. Such functional examination of enhancers will be critical toward understanding how non-coding sequence variation in human populations contributes to brain development and neurological and psychiatric disorders.

1918T

Associations between KIBRA (*WWC1*) gene methylation and cognitive abilities in a childhood monozygotic twin difference design. C.R. Lewis^{1,2}, A. Henderson-Smith², R.S. Breitenstein¹, K. Lemery-Chalfant¹, L.D. Doane¹, M.J. Huettelman². 1) TGen, Phoenix, AZ; 2) Arizona State University, Tempe, AZ.

Research suggests genetic variation drives individual differences in cognitive abilities. For example, 12 years ago others and we established an association between KIBRA gene (*WWC1*) variation and episodic memory in young, middle, and elderly aged participants (Papassotiropoulos et al., 2006; Schaper et al., 2008). However, research suggests environmental factors are also influencing complex phenotypes via the epigenome. We used a monozygotic (MZ) twin difference design to examine cotwin variation in *WWC1* methylation in relation to cotwin differences in a battery of cognitive functions. Comparing MZ twins allowed us to assess if environmentally driven differences in methylation affect phenotypes while controlling for genotype on methylation and performance. The sample included 48 MZ twin pairs (51% male; 50% Non-HispanicWhite, 15% Latinx, 8% AfricanAmerican, 4% AsianAmerican), mean age = 8.5 years, drawn from the Arizona Twin Project (Lemery-Chalfant et al., 2013). DNA methylation was quantified using the Infinium MethylationEPIC BeadChip. Executive attention and inhibition was assessed with the Flanker Task (Linear Integration Speed Accuracy Score [LISAS] for congruent and incongruent trials); short-term and working memory (STM and WM) with Digit Span (Total Forward and Backward); picture vocabulary, problem solving, and reading comprehension with the Woodcock Johnson. We extracted the first principal component from % methylation of 82 CpG sites near *WWC1* after removing all sites with a < .3 loading (45 sites remained; variance explained = 43.29%) for use in linear mixed models. Next, we computed MZ difference scores in % methylation and task performance for linear regression analysis. *WWC1* methylation was negatively associated with reading comprehension ($b = -2.91$, $p = 0.007$) and no other performance. MZ differences in *WWC1* methylation predicted executive attention performance (LISA Congruent: $b = 0.367$, $p = 0.022$) and no other performance. Our results indicate methylation levels of *WWC1* may influence reading comprehension, whereas environmentally driven methylation levels influence executive attention. Clearly, DNA methylation and complex cognitive functions employ an intricate relationship. As we unravel the distinction between environmentally and genetically driven methylation, we may be able to develop novel targetable interventions for neurodevelopmental, psychiatric, and neurodegenerative disorders that share a common cognitive dysfunction phenotype.

1919F

An integrated genetic and epigenetic approach identified new candidate genetic loci for narcolepsy. M. Shimada^{1,2}, T. Miyagawa^{1,2}, H. Toyoda², K. Tokunaga², M. Honda^{1,3}. 1) Department of Psychiatry and Behavioral Sciences, Tokyo Metropolitan Institute of Medical Science, Tokyo, Japan; 2) Department of Human Genetics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; 3) Seiwa Hospital, Neuropsychiatric Research Institute, Tokyo, Japan.

Narcolepsy with cataplexy is a sleep disorder characterized by excessive daytime sleepiness and cataplexy. One proposed cause of narcolepsy is loss of hypocretin-producing cells in the lateral hypothalamus. Several genetic factors of narcolepsy including *HLA-DQB1*06:02* have been identified, however, a detailed etiology still remains to be elucidated. Epigenetic modifications such as DNA methylation have recently been suggested to play a certain role in the pathogenesis of complex diseases. In this study, we first performed a microarray-based epigenome-wide association study (EWAS) of DNA methylation for narcolepsy using blood samples and examined differentially methylated positions (DMPs) linked to the disease. As a result, CpG sites with top associations were found to be significantly abundant in non-CpG islands ($P = 7.32 \times 10^{-10}$) and more than 95% of such sites were hypomethylated in the narcolepsy patients. A pathway analysis identified that genes annotated to the top-ranked CpG sites had the functional relevance to hormone secretion and monocarboxylic acid metabolic processes. DMPs that showed an association with narcolepsy can be a cause of the disease, however, they also have the possibility to be a result of the disease. In order to search methylation loci which contribute to the development of the disease, we next conducted a new integrative approach of the EWAS and a genome-wide association study (GWAS) for narcolepsy. To achieve this, we utilized the concept of methylation quantitative trait loci (meQTL), i.e., methylation levels of meQTLs are related to genotypes of genetic variants. We searched meQTLs of which both methylation and genetic variant were associated with narcolepsy. We detected 19 meQTLs, all of which were *cis*-meQTLs located in the *CCR3* region. Further, now we are performing an EWAS using DNA samples extracted from lateral hypothalamus in the brain. We then examined whether the 100 top-ranked CpG sites found in the blood analysis also showed associations in the brain analysis. We found that associations of three CpG sites, including one annotated to *MAGEL2*, were consistent between blood and brain. *MAGEL2* is located in the region associated with Prader-Willi syndrome, which reportedly presents with symptoms of hypersomnia. Although future replication studies are necessary to confirm the results, the different levels of DNA methylation might have a certain role in the pathophysiology of narcolepsy.

1920W

Positive and negative co-expression with *COMT* in four brain areas. J.L. Dannemiller. Rice University, Houston, TX.

Background: The membrane-bound form of Catechol-O-Methyltransferase (MB-COMT) catabolizes catecholamines such as the neurotransmitter dopamine and is expressed most strongly in human Prefrontal Cortex. Variants of *COMT* (22q11.21) have been associated with various normal and abnormal cognitive phenotypes, but the literature is often conflicting on their effects. It is possible that some of the discrepancies in the literature arise because of differences in the genes co-expressed with *COMT*. **Methods:** Publicly available, post-mortem mRNA expression data (Illumina humanRef-8 v2.0 expression beadchip GPL6104) from four brain areas (Prefrontal Cortex, Cerebellum, Temporal Cortical Wall and Pons; hereafter PfCtx, CBell, TmCtx, Pons) in 138 subjects (GeneNetwork GN Accession: GN482, GEO Series: GSE15745) were used to determine characteristics of genes co-expressed with *COMT*. Bioconductor software for R (v3.5.0) was used as the primary analysis platform with gene as the data-analytic unit. The 500 genes (TOP500) whose expression levels were most strongly correlated (ranked by *p* value) with *COMT* expression were selected for further analysis in each brain area. A form of differential co-expression analysis was implemented by separating the TOP500 genes in each brain area into two sets: those with expression positively (*r*-pos) vs. negatively (*r*-neg) correlated with *COMT*. **Results:** 1) While the proportion of neg-*r* genes in the full sample was approximately 66% in all brain areas, this proportion flipped among the four sets of TOP500 genes. 2) When all length-6 motifs were examined in the promoter regions of these genes (-1000 to +600 bp from the TSS), the proportion of neg-*r* genes containing at least one instance of a given motif significantly exceeded the proportion of pos-*r* genes in 14 of 16 tests surviving *Bonferroni* correction. 3) The proportion of genes in PfCtx with at least one CpG island within 200 bp of the TSS differed significantly: 52% vs. 63% (neg-*r* vs. pos-*r*, respectively). 4) GO analysis (BP) produced a clear difference between the two sets of genes in PfCtx; 9 of the top 10 GO terms for the neg-*r* genes were related to "vesicle" processes. For pos-*r* genes, terms related to "catabolic" processes were enriched. **Implications:** The "vesicular" enrichment results could imply that when *COMT* is downregulated, some of the neg-*r* genes are upregulated possibly because of the increased demand for vesicular packaging of newly endocytic dopamine in presynaptic neurons.

1921T

Small nucleolar RNAs in major depression and antidepressant response.

R. Lin¹, J.P. Lopez¹, L.M. Fiori¹, J.F. Theroux¹, H. Peng¹, C. Ernst¹, C. Cruceanu¹, R. Belzeaux¹, CANBIND Working Group³, J. Foster², S. Kennedy², G. Turecki¹. 1) McGill Group for Suicide Studies, Douglas Mental Health University Institute, McGill University, Montreal, Quebec, Canada; 2) Department of Psychiatry, University Health Network, University of Toronto, Toronto, Ontario, Canada; 3) The Canadian Biomarker Integration Network in Depression (www.canbind.ca).

Background: Major depressive disorder (MDD) is a prevalent global disorder treated primarily by antidepressants (AD). However, 30%-40% of subjects fail to respond after multiple attempts. Thus, discovering biomarkers for MDD and AD-response would greatly improve diagnostics and treatment. Small non-coding RNAs (sncRNAs) have recently been identified as feasible biomarkers associated with disease states, as they are stably circulated in peripheral blood and are involved in several post-transcriptional modification processes. I specifically profile small nucleolar RNAs (snoRNAs), since they have been disregarded in MDD pathology, leaving room for novel discoveries. To holistically evaluate snoRNAs as candidate biomarkers, I am using a unique combinatorial approach of profiling snoRNAs expression in both post-mortem human brain tissue and living human peripheral blood. **Methods:** Post-mortem brain tissue was collected from individuals who died by suicide during an MDD episode, and psychiatrically normal individuals who died suddenly. Peripheral blood was collected from 3 independent cohorts consisting of MDD subjects treated with AD for 8 weeks. Subjects were assessed for MDD severity using the Montgomery-Asberg Depression Rating Scale (MADRS) before treatment (Week 0) and afterwards (Week 8). Post-AD treatment subjects were separated into responders ($\geq 50\%$ decrease in MADRS score; RES) or non-responders (nRES). Small RNA-sequencing was used to generate whole sncRNA-transcriptome expression data from post-mortem brain and peripheral blood. **Results:** SNORD90 and SNORD43 were found to be upregulated in AD-responders across time; this was replicated in all 3 cohorts. Next looking at our post-mortem brain samples, only SNORD90 was found to be differentially expressed between MDD-suicides and controls. All sequencing results have been validated via qPCR. Using in-silico analysis Neuregulin 3 (NRG3) has been identified as a potential target of SNORD90. This predicted SNORD90-NRG3 interaction will be interrogated via knock-down experiments in cell culture. **Conclusions:** To my knowledge, this is the first study focusing on snoRNAs in the context of MDD and AD-response. This novel information will not only further our understanding of MDD pathology, but has potential clinical impact as it also shows characteristics of a biomarker. Overall, my project has the potential of having a major impact on the diagnosis, treatment, and understanding of MDD.

1922F

DNA methylation and three dimensional chromatin structure of the *IRXA* gene cluster in human cocaine dependence. K. Vaillancourt^{1,2}, J. Yang¹, G.G. Chen¹, L. Fiori¹, C. Ernst^{1,3}, E. Calipari⁴, B. Labonté⁴, E. Nestler⁴, T. Forné⁵, D.C. Mash⁶, G. Turecki^{1,2,3}. 1) McGill Group for Suicide Studies, Douglas Hospital Research Center, Verdun, QC, Canada; 2) Integrated Program in Neuroscience, McGill University, Montreal, QC, Canada; 3) Department of Psychiatry, McGill University, Montreal, QC, Canada; 4) Icahn School of Medicine, Mount Sinai Hospital, New York, NY, USA; 5) Institut de Génétique Moléculaire de Montpellier, Montpellier, France; 6) Department of Neurology, University of Miami Miller School of Medicine, Miami, FL, USA.

Background: Multiple, interacting biological mechanisms are likely to contribute to the development and maintenance of cocaine use disorders. Epigenetic mechanisms are of particular interest as they may mediate the long-term effects of chronic cocaine abuse on brain cell functioning in addiction-related brain circuits. Although animal research has built a strong case for the role of DNA methylation and chromatin remodelling in cocaine dependence, little is known about the relationship between these phenomena in human patients. **Methods:** We used reduced representation bisulfite sequencing (RRBS) to identify genome-wide differential methylation, in post mortem caudate nucleus and nucleus accumbens tissue, from 25 individuals with cocaine dependence and 25 drug-naïve controls. We validated and replicated our findings using targeted bisulfite sequencing, and used fluorescence activated nuclei sorting (FACS) to investigate cell-type specificity. We also used transcriptome sequencing, and chromatin conformation capture (3C) based techniques to explore the genomic regulation of the *IRXA* gene cluster. Our *in vitro* studies use immortalized fetal midbrain cells (RenCells) and dCas9-DNMT3a epigenome editing to investigate causal relationships between cocaine-related methylation changes and chromatin functioning at this locus. **Results:** We found altered methylation within two genes of the neurodevelopmental gene cluster, *IRXA*, in the caudate nucleus. We replicated our finding of decreased methylation within a cluster of 25 CpGs in the gene body of *IRX2* in an independent cohort, and found this effect to be specific to neuronal nuclei. In addition, cocaine dependence is associated with increased expression of the *IRX2* gene and altered co-expression of *IRX1* and *IRX2* transcripts. Moreover, we identified a large chromatin loop in human cells, which is related to *IRX2* methylation levels and is likely involved in *IRXA* regulation. **Conclusions:** Chronic cocaine dependence is associated with widespread alternations in DNA methylation in the human striatum, including decreased methylation within the gene body of *IRX2* in the caudate nucleus. This appears to be associated with dysregulated gene expression, and may be driven by altered chromatin looping. Further studies will continue to investigate the relationship between multiple epigenetic mechanisms in human cocaine use disorders.

1923W

Co-expression patterns define epigenetic regulators associated with neurological dysfunction. L. Boukas¹, J.M. Havrilla^{2,3,4}, A.R. Quinlan^{2,3,4}, H.T. Björnsson^{1,5,6}, K.D. Hansen^{1,7}. 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Department of Human Genetics, University of Utah, Salt Lake City, UT; 3) Department of Biomedical Informatics, University of Utah, Salt Lake City, UT; 4) USTAR Center for Genetic Discovery, University of Utah, Salt Lake City, UT; 5) Faculty of Medicine, University of Iceland; 6) Landspítali University Hospital, Reykjavik, Iceland; 7) Department of Biostatistics, Johns Hopkins School of Public Health, Baltimore, MD.

Coding variants in genes encoding for epigenetic regulators are an emerging cause of neurological dysfunction and cancer. However, a systematic effort to identify and prioritize disease candidates within the human epigenetic machinery (EM) has not been performed. Additionally, it is uncertain whether dysregulation of the normal chromatin state is the most likely cause of disease, as studies in model organisms have shown that the protein domains mediating the epigenetic function of these factors can at times be dispensable. Finally, it is unclear whether features exist that can distinguish between variation-intolerant and variation-tolerant EM genes, and between EM genes associated with neurological dysfunction versus cancer. Here, we first rigorously define a set of 295 human genes with a direct role in epigenetic regulation (writers, erasers, remodelers, readers). Using pLI as a metric, we find that the majority of EM genes are very intolerant to loss-of-function variation ($p < 0.001$, Odds Ratio = 7.7 for enrichment in the group of genes with pLI > 0.9), even when compared to the dosage sensitive class of transcription factors ($p < 0.001$, OR = 4.4 for enrichment in the pLI > 0.9 group). Using this strategy, we identify 103 novel EM disease candidates. We then perform a local, protein domain-specific mutational constraint analysis, and show that this intolerance to loss-of-function variation is driven by the protein domains encoding the epigenetic function ($p = 0.009$), strongly suggesting that the disease phenotypes are in most cases caused by a perturbed chromatin state. To search for functional characteristics specific to dosage-sensitive EM genes, we construct tissue-specific co-expression networks, using large-scale RNA-seq data from the GTEx consortium. Unexpectedly, we describe 74 EM genes that are co-expressed in multiple tissues ($p < 0.001$). This subset is almost exclusively populated by extremely variation-intolerant EM genes (72 of the 74 have a pLI > 0.9), and shows very high enrichment for a unique class of EM genes with dual enzymatic and reading function ($p < 0.001$, OR = 13.3). It is also enriched for genes associated with neurological dysfunction ($p = 0.006$, OR = 3.3), even when accounting for dosage sensitivity ($p = 0.003$, OR = 3), but not for cancer-associated EM genes ($p = 0.41$). These findings prioritize disease candidate EM genes, and suggest that the co-expression itself may play a functional role in normal neurological homeostasis.

1924T

Whole-genome bisulfite sequencing of Down syndrome cortex reveals regional DNA hypermethylation of active chromatin states and novel disorder insights. B.I. Laufer¹, J.A. Gomez¹, D.H. Yasui¹, C.E. Mordaunt¹, A.C. Vogel Ciernia¹, H. Hwang¹, J. Lu², V.L. Sheen³, J.M. LaSalle^{4,5}. 1) Department of Medical Microbiology and Immunology, University of California, Davis, CA; 2) Department of Anatomy, China Medical University, Shenyang, China; 3) Department of Neurology, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA; 4) Genome Center, University of California, Davis, CA; 5) MIND Institute, University of California, Davis, CA.

Down Syndrome (DS) is the most common genetic cause of intellectual disability, resulting from trisomy of human chromosome 21 (HSA21). In addition to copy number and expression differences in HSA21 genes, genome-wide differences in DNA methylation have been observed in a variety of DS tissues and cells using reduced representation assays of individual CpGs. To obtain a more complete epigenomic profile of DS brain, whole-genome bisulfite sequencing (WGBS) of DS and matched control brains (male frontal cortex, BA9) was performed. Across the entire genome, 2047 significant (areaStat >20 or <-20) differentially methylated regions (DMRs) that distinguished DS from control brain were identified. A larger number of DS-specific DMRs were hypermethylated (1459) than hypomethylated (588). Notably, HSA21 showed the lowest abundance of autosomal DMRs as well as a uniquely even balance of hypermethylated and hypomethylated DMRs. The hypermethylated DMRs were significantly (FDR<0.05) enriched for active promoters and bivalent chromatin marked by H3K4me3, CpG islands and shores, and CTCF binding sites. Functionally, DMRs were significantly (FDR<0.05) associated with genes related to metabolism, neurodevelopment, transcription, and cell cycle. The DMRs also significantly ($p=0.0003$) overlap with previously identified pan-tissue DS genes with differential methylation, which are primarily hypermethylated (24/25 genes). While predominantly novel, the DMRs identified from WGBS are significantly (FDR<0.05) enriched at the sequence level for CpG sites identified in previous DNA methylation analyses of 12 diverse DS adult and fetal cell-types/tissues. Integration of the DMRs with roadmap epigenomics histone modification data uncovered a brain specific profile of the hypomethylated, but not hypermethylated, DMRs. We hypothesize that the wide-scale DS hypermethylation is related to DNMT3L, which is located on HSA21, and its interactions with H3K4me3 and lamina associated domains, while the hypomethylated DMRs represent a tissue specific response to the hypermethylation. Currently, we are testing this hypothesis by utilizing DS and matched control fetal frontal cortex samples from both sexes and two transgenic human neuronal cell lines (LUHMES and SH-SY5Y) overexpressing DNMT3L. Overall, these results demonstrate that WGBS provides a more complete profiling of DNA methylation differences in DS that offers novel insights into genetic pathways that may be targeted by therapeutics.

1925F

5-Aminolevulinic acid may ameliorate the cognitive function of patients of ATR-X syndrome. T. Wada¹, S. Suzuki², N. Shioda³. 1) Kyoto University Graduate School of Medicine, Kyoto, Japan; 2) Department of Endocrinology and Metabolism, Fukuoka Children's Hospital, Fukuoka, Japan; 3) Department of Biofunctional Analysis Laboratory of Molecular Biology, Gifu Pharmaceutical University, Gifu, Japan.

BACKGROUND; ATR-X syndrome (OMIM #301040) is one of the X-linked intellectual disability syndromes caused by mutations of ATRX gene, encoding a chromatin remodeling protein, ATRX. We have no treatment for it up to now. It has been reported that ATRX protein targets tandem repeats, forming G-quadruplex (G4) structures, and regulates genes nearby. Recently we have published an article by Shioda N. in Nature Medicine, reporting that 5-Aminolevulinic acid (5-ALA) can be a potential therapeutic strategy to target G-quadruplexes and improve the cognitive function of ATR-X syndrome on the study of ATR-X model mice. This presentation is a case report of a boy of ATR-X syndrome, whose language ability improved with taking 5-ALA. **CASE;** The case is a 4-year-old boy, whose diagnosis of ATR-X syndrome was confirmed molecular genetically, c.7192 C>T [p.Gln2398*] in ATRX gene. He showed typical clinical features of ATR-X syndrome with severe intellectual disability. At his age of 3 years and 4 months, he could not speak any meaningful words. He could only babble. His father gave 5-ALA to his son, hoping to improve his health for 9 months. At two month after starting taking 5-ALA at 5mg every day orally, he got several meaningful words, including "Ocha (referring to "tea" in English)", "Ka-san (Mom)", "Ji-chan (Grandpa)" and so on. After stopping taking 5-ALA for a month, he had lost his meaningful words. But he restarted taking 5-ALA at 30mg every day, and he could recover his language ability and acquired more new words, including "Tentei (teacher)", "Ko-ki (airplane)", "Ju-su (juice)", "Gyu-gyu (Milk)", and so on. He had to discontinue to take 5-ALA due to financial reasons, and he had lost his words again. No side effects were observed. **DISCUSSION;** This is the first case report suggesting that 5-ALA can ameliorate the cognitive function for ATR-X patients. 5-ALA is an endogenous non-protein amino acid, and is available commercially as an amino acid supplement in Japan. In general most patients of ATR-X patients cannot acquire any meaningful words, except some mild cases with a few specific variants in ATRX gene. This report can support our recent data that 5-ALA can ameliorate the cognitive function of ATR-X model mice. We are preparing to start the clinical trial study for administering 5-ALA to patients of ATR-X syndrome in order to confirm that 5-ALA is a good treatment for cognitive disability of ATR-X patients.

1926W

MiRNAs involvement in the Fragile X-associated Disorders development.

A.A. Dolskii, N.A. Lemskaya, D.V. Yudkin. State Research Center of Virology and Biotechnology VECTOR, Koltsovo, Russian Federation.

Fragile X-associated Disorders (FXD) include Fragile X-associated tremor/ataxia syndrome (FXTAS), Fragile X-associated Primary Ovarian Insufficiency (FXPOI) and Fragile X syndrome. All of them are related to the CGG repeat expansion in the 5'-UTR of *FMR1* gene and changing in synthesis of FMRP. The allele of the *FMR1* gene is called premutation, when the repeat expansion in the range of 55 to 200 triplets. Its frequency is 1:250 in women and 1:350 in men. In this case, expression of *FMR1* mRNA can be increased, but the amount of protein, usually decreases or stays at normal level. In this case, the FXTAS and FXPOI are developed. Fragile X syndrome emerges with the repeat expansion, more than 200 units and promoter region methylation (full mutation allele). It should be noted that the symptoms of full mutation and premutation alleles are different, indicating different mechanisms of pathogenesis in both cases. In the first case, this is the absence of the FMRP protein, whereas the pathogenesis mechanisms of the FXTAS and FXPOI is unclear. The alteration of the *FMR1* gene expression during repeat expansion indicates the possible miRNAs involvement in the development of this pathology. It is assumed that a decrease of FMRP protein in premutation is associated with RNA interference. Also, the *FMR1* gene activity can be influenced by miRNAs which arise from the GCC repeat in the *ASFMR1* transcript. The aim of this work is to study the of miRNAs involvement in the regulation of the *FMR1* gene expression in FXD. To achieve the goal, the level of 8 microRNAs was studied by Taq-man PCR in B-lymphocytes cell lines with different repeat size and different *FMR1* gene activity. Also, to study the possible effect of microRNAs containing the GCC repeat, the level of expression of *ASFMR1* was determined. An increase in some miRNAs expression in full mutation cell lines was shown, suggesting the possible involvement in the Fragile X syndrome pathogenesis. For other ones there was no correlation between their expression levels and the ratio of the expression values of the mRNA gene of the *FMR1* gene and the amount of the FMRP protein, which indicates that there is no participation in the FXD development. The work was supported by the Russian Science Foundation project 18-15-00099.

1927T

(CGG)n repeat instability and active transcription in human tissue

culture. I.V. Grishchenko, Y.V. Purvinsh, N.A. Lemskaya, D.V. Yudkin. State Research Center of Virology and Biotechnology VECTOR, Koltsovo, Russian Federation.

The expansion of repeats is a type of mutation, which is an increase of the repeats number in the hundreds of units. This phenomenon leads to a certain human genetic diseases, such as Huntington's disease, Fragile X syndrome. More than 30 diseases are caused by expansions of repeats dispersed throughout the human genome. Fragile X syndrome is the most common heritable intellectual disability associated with the repeat instability. It is caused by the expansion of (CGG)_n repeat in the promoter region of the gene *FMR1*. The number of repeats in the normal ranges is from 5 to 54 units. With the repeat expansion in the range of 55 to 200 triplets, the allele of the *FMR1* gene is called premutation. Fragile X syndrome develops when the number of repeated triplets becomes over 200. Today the cause of this mutation is unknown. According to theory the expansion is due to the formation of alternative stable secondary DNA structures. At sites with these structures there is a failure of the repair and replication systems, which leads to an increase the size of the repeat. A feature of the *FMR1* gene promoter region is coding strand contains (CGG)_n repeat, and the template strand contains CCG triplets. Thus, on the coding strand increased repeat form a G-quadruplex, and the ones on template strand form an i-motif. Since expansion take place in non-dividing cells like oocytes or neurons it can be assumed an expansion mechanism involves aberrant DNA repair, particularly transcription-coupled repair. Moreover, in the premutant alleles of the *FMR1* gene which rapidly accumulating the repeated units a dramatically increasing of the transcription level were observed. These findings indicate the possible involvement of active transcription in the of (CGG)_n repeats expansion mechanism. The project aims to investigate (CGG)_n repeat instability associated with the Fragile X syndrome during transcription in human tissue culture. Work was carried out on HEK293A culture cells, cultures of immortalized B-lymphocytes of patients and control cell lines. Recombinant plasmids with induced promoter TRE containing the premutant and normal alleles downstream of TRE were constructed. Further electroporation of immortalized B-lymphocytes and liposome transfection of HEK293A cells were carried out. In this study we investigated the effect of active transcription on the (CGG)_n repeat length in the plasmid over time. The reported study was funded by Russian Science Foundation project 18-15-00099.

1928F

FASD as an experimental model of an epigenetic neurodevelopmental disorder. B. Alberry¹, E.J. Chater-Diehl¹, B.I. Laufer², S.M. Singh¹. 1) Biology, Western University, London, Ontario, Canada; 2) SickKids Research Institute, Program in Genetics and Genome Biology, Toronto Ontario, Canada; 3) Department of Medical Microbiology and Immunology, School of Medicine, University of California, Davis, CA.

Most neurodevelopmental disorders involve genes, environment and poorly understood epigenetic interactions that are multifaceted, dynamic and complex. Clarifying these interactions is critical but difficult, requiring a suitable model. We propose Fetal Alcohol Spectrum Disorder (FASD), a common neurodevelopmental disorder resulting from prenatal alcohol exposure modeled in C57BL/6J, is highly suited for this insight. Here we demonstrate that following continuous mild prenatal alcohol exposure in mice during neurodevelopment, mice have behavior phenotypes seen in FASD. Also, they have many life-long hippocampal molecular changes including genome-wide gene and microRNA expression (RNA-Seq), and DNA methylation (meDIP-Seq). The relationship between these features is not understood and may account for initiation as well as persistence of behavioral abnormalities in FASD. Interestingly, changes in DNA methylation in response to prenatal alcohol exposure in mice have also been demonstrated in buccal swabs from children with FASD. More importantly, postnatal manipulations or treatment paradigms potentially applicable to children can be applied and tested in young mice during development. In fact, it is possible to improve or worsen behavioral outcomes by favorable and unfavorable postnatal environments in mice subject to prenatal ethanol exposure, in this case early life stress via maternal separation. Additionally, behavioral outcomes are reflected in newly established molecular features in response to postnatal environment. Critical genes and pathways identified provide a foundation for the development of FASD diagnosis using DNA methylation and favorable interventions for children born with FASD. Finally, pathways and ontologies associated with these features implicate a role for epigenetic regulation of gene expression common in neurodevelopmental disorders.

1929W

Common mitochondrial DNA variants implicated in variations in mitochondrial oxidative activity. N. Cai^{1,2}, J. Stephenson², E. Yonova-Doing³, M.J. Bonder², B. Mirauta², J. Danesh², P. Chinnery^{4,5}, J. Howson³, O. Stegle², N. Soranzo¹. 1) Wellcome Sanger Institute, Cambridge, Cambridgeshire, United Kingdom; 2) European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Cambridge, United Kingdom; 3) British Heart Foundation Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom; 4) MRC Mitochondrial Biology Unit, Cambridge Biomedical Campus, Cambridge, United Kingdom; 5) Department of Clinical Neurosciences, Cambridge Biomedical Campus, University of Cambridge, Cambridge, United Kingdom.

Mitochondrial DNA (mtDNA) codes for 13 proteins and 22 RNAs involved in the synthesis of key respiratory chain components, and hence an attractive candidate for the "missing heritability" in complex diseases through influencing metabolic activity. In this study, we examine the relationship between inherited mtDNA variations and 995 metabolites quantified from nearly 8,000 healthy blood donors in the INTERVAL study. Using high-resolution whole-exome (50x) and whole-genome (15x) sequencing, we are able to capture the near totality of mtDNA variation (of which 511 are high quality biallelic SNPs) and interrogate all potential causal SNPs. We show that sites defining common European haplogroups U and K are significantly associated ($P < 10^{-14}$) with the levels of N-formylmethionine (fMet), the amino acid that initiates translation in the mitochondria, independent of nuclear variations and accounting for 1% of its heritability ($h^2=18\%$). This finding implicates mtDNA variants in regulation of mitochondrial gene expression and, in turn, oxidative phosphorylation activity. We therefore investigated the functional impact on respiratory chain complexes associated with potentially causal SNPs by examining their effects on gene expression and protein levels in 44 human primary tissues from GTEx iPSCs from HipSci Consortiums. We find that MT1811G (16s rRNA), the top SNP associated with fMet affects protein levels of mtDNA encoded genes ($p < 0.05$) in iPSCs. In addition, MT1189C (12s rRNA) shows the strongest and most consistent effects on expression of the mtDNA encoded genes across tissues, including significant negative associations with levels of ND3 transcripts (Complex I) in 7 tissues (5% FDR, $p < 10^{-5}$), and positive effects on levels of COX3 transcripts (Complex IV) in 36 tissues. In contrast, we only detect enrichment of its effects on nuclear genes encoding subunits of respiratory complexes, in particular tissues such as heart left ventricle and several brain regions. This raises the possibility of cis-transcriptional control by mtDNA variants, with opposing effects on different complexes determining metabolic strategies, likely with tissue specific impacts. Our results, together with previous evidence that mtDNA variants have functional effects on oxidative phosphorylation, ATP levels, and generation of reactive oxygen species, shed light on potential new mechanisms where inherited mtDNA variation may influence oxidative, cellular and organismal phenotypes.

1930T

Conserved roles for CHD7, the chromatin remodeler mutated in CHARGE syndrome, in transcriptional elongation of genes involved in neural, neural crest, and inner ear development. E. Ritter¹, H. Yao², D.F. Hannum², G.J. Sanchez³, E.D. Sperry^{3,4}, A. Saiakhova⁵, D.R. Fuentes⁶, M.E. Bowen⁷, A.N. Niederriter-Shamir^{3,4}, J.M. Skidmore¹, A. Srivastava³, L.D. Attardi⁷, T. Swigut⁶, J. Wysocka^{6,8,9}, P.C. Scacheri⁵, D.M. Martin^{1,3,4}. 1) Dept. of Pediatrics & Communicable Diseases, University of Michigan, Ann Arbor, MI; 2) Dept. of Biostatistics, University of Michigan, Ann Arbor, MI; 3) Dept. of Human Genetics, University of Michigan, Ann Arbor, MI; 4) Medical Scientist Training Program, University of Michigan, Ann Arbor, MI; 5) Dept. of Genetics, Case Western Reserve University, Cleveland, OH; 6) Dept. of Chemical and Systems Biology, Stanford University, Stanford, CA; 7) Dept. of Radiation and Cancer Biology, Stanford University, Stanford, CA; 8) Dept. of Developmental Biology, Stanford University, Stanford, CA; 9) Howard Hughes Medical Institute, Stanford University, Stanford, CA.

CHD7 is an ATP-dependent helicase that is critical for normal mammalian development. Heterozygous pathogenic variants in *CHD7* cause multiple organ system malformations and CHARGE syndrome, which includes ocular Coloboma, Heart defects, Atresia of the choanae, Retardation of growth and development, Genital hypoplasia and pubertal delay, and Ear abnormalities, including deafness and vestibular disorders. Central and peripheral nervous system impairment are common in CHARGE, yet the underlying mechanisms of CHD7 action in early neural and neural crest development are not well understood. *Chd7*^{+/−} mice exhibit dysplastic semicircular canals and mild mixed conductive-sensorineural hearing loss and are an excellent model for CHARGE syndrome. Reduced *Chd7* function in the ear disrupts expression of genes involved in developmental patterning and neurogenesis. Here we explored potential roles for CHD7 in global regulation of gene transcription in the nervous system, including differentiation of stem cells toward neural epithelial progenitors, neural crest cells, and inner ear tissues. We used mouse embryonic stem cells (ESCs) null for *Chd7* to study CHD7 in early neuronal development, during the transition of embryonic stem cells to neural epithelial progenitors (NEC), and found that loss of *Chd7* disrupts neuronal and glial differentiation without affecting proliferation or apoptosis. Transcriptome analysis via RNA-Seq on E10.5 and E12.5 otocysts from *Chd7*^{+/−} and *Chd7*^{0/0} mutant mouse embryos showed a length-dependent effect of *Chd7* on gene transcription, with preferential effects on longer (>100 kb) genes. Notably, a similar length dependent effect was observed in heterozygous *CHD7* mutation-positive human induced pluripotent stem cells (iPSCs) differentiated into neural crest cells (NCCs) when compared to isogenic patient derived iPSCs in which the variant was corrected via CRISPR/Cas9 gene editing. We also observed that genes with high CHD7 occupancy display low RNA polymerase II (Pol II) pausing and higher RNAPII throughout the gene body. Collectively, our findings support roles for CHD7 in chromatin events that regulate transcriptional elongation. Interestingly, length-dependent gene dysregulation has been described for Rett syndrome, Autism Spectrum Disorder, and Fragile X syndrome, suggesting possible links between the molecular etiologies of these disorders and CHARGE syndrome.

1931F

Submicroscopic deletions underlying congenital microcoria modify the regulatory landscape of chromosome 13q32.1. L. Fares Taie¹, B. Nedelec¹, P. David², S. Gerber¹, S. Crippa³, B. Passet⁴, J.L. Vilotte⁴, N. Chassaing⁵, J. Kaplan¹, C. Kostic³, P. Calvas³, J.M. Rozet¹. 1) Laboratory of Genetics in Ophthalmology (LGO). INSERM UMR1163. Imagine – Institute of Genetic Diseases, Paris Descartes University, 75015 Paris, France; 2) LEAT Imagine – Institute of Genetic Diseases, Paris Descartes University, 75015 Paris, France; 3) University of Lausanne · Ophthalmology. Jules-Gonin Eye Hospital. Fondation Asile des Aveugles. CP 5143 1002, Lausanne, Switzerland; 4) UMR1313 Génétique Animale et Biologie Intégrative, Institut National de la Recherche Agronomique, 78352 Jouy-en-Josas, France; 5) Service de Génétique Clinique, Hôpital Purpan, 31300 Toulouse, France.

Congenital microcoria (MCOR) is a rare autosomal-dominant disorder characterized by inability of the iris to dilate owing to absence of dilator pupillae muscle. In addition, the miotic iris is thin and displays abnormal stroma and iridocorneal angle. Iris thinning is consistent with transillumination of miotic irises and high sensitivity to light. In addition, consistent with abnormal iridocorneal angle development, glaucoma is frequently associated with this condition. Previously, we reported heterozygote 13q32.1 deletions as responsible for MCOR in six families. The deletions varied from 35 kbp to 80 kbp in size, but invariably encompassed or interrupted only two genes: *TGDS* encoding the TDP-glucose 4,6-dehydratase involved in Catel-Manzke syndrome and *GPR180* encoding the G protein-coupled receptor 180 which heterozygous loss-of-function caused iridocorneal angle dysgenesis but no microcoria, suggesting the contribution of 13q32.1 regulating elements in the full MCOR phenotype. We generated transgenic mouse lines harboring variable deletions of the syntenic region using CRISPR/Cas9 methodology. Homozygous animal were not viable. Heterozygous animals presented with minor iris dilation dysfunction but molecular analysis including RNAseq, qPCR and Western Blot disclosed strong ectopic expression of neighboring genes during iris development, suggesting that MCOR deletions modify the 13q32.1 regulatory landscape and the fate of some embryonic cells involved in iris formation.

1932W

Induced pluripotent stem cell-derived retinal pigment epithelium is an effective model system to study regulatory genetic variation associated with age-related macular degeneration. *E.N. Smith¹, A. D'Antonio-Chronowska², M. D'Antonio³, W.W. Greenwald², V. Borja², L. Aguiar², J. Sohmer², H. Matsui², S. Boroah⁴, R. Ayyagari⁴, K.A. Frazer^{1,2}.* 1) Pediatrics and Rady Children's Hospital, University of California, San Diego, La Jolla, CA; 2) Institute of Genomic Medicine, University of California San Diego, 9500 Gilman Dr, La Jolla, CA 92093, USA; 3) Bioinformatics and Systems Biology Graduate Program, University of California, San Diego, La Jolla, CA, USA; 4) Shiley Eye Institute, University of California San Diego, La Jolla, USA.

Over 1.6 million people ages 50 or older in the US have visual impairment resulting from age-related macular degeneration (AMD) – and current therapeutic options are quite limited. AMD has a well characterized genetic basis from GWAS, but little is known about the causal variants at associated loci. Induced pluripotent stem cell-derived retinal pigment epithelium (iPSC-RPE) are a highly promising system to functionally characterize the epigenetic and transcriptomic effects of these genetic variants as AMD pathogenesis involves progressive degeneration RPE cells. To overcome the challenges associated with directly obtaining primary RPE, we have begun the differentiation of iPSCs from the iPSCORE resource into iPSC-RPEs. Here we examine whether iPSC-RPEs are an effective model for studying and fine-mapping AMD associated genetic variation. We first differentiated iPSCs from 6 subjects and derived high-quality iPSC-RPEs displaying many morphological, molecular, and functional characteristics of naïve RPE including high ZO-1 and MITF staining. For all 6 iPSC-RPEs, as well as 1 fetal RPE sample, we generated RNA-seq and ATAC-seq. We then examined whether iPSC-RPEs showed similar gene expression profiles to fetal RPEs via principle component analysis (PCA) of the RNA-seq data, and observed clustering of the iPSC-RPEs with fetal RPE. Next, through functional enrichment analysis, we confirmed that the associated PCs were enriched for RPE associated functions, including sensory organ morphogenesis ($P < 10^{-30}$) and retina development ($P = 9.0 \times 10^{-10}$). We then analyzed the ATAC-seq data to determine if accessible chromatin peaks in iPSC-RPEs were: 1) reflective of RPE biology, 2) similar to adult RPE peaks, and 3) enriched for AMD GWAS loci. We observed motifs for important RPE TFs enriched in iPSC-RPE peak regions, including OTX2 ($P < 10^{-2884}$), LHX2 ($P < 10^{-518}$), and MITF ($P < 10^{-215}$). Additionally, there was strong overlap between iPSC-RPE peak regions and peak regions from adult RPEs with and without AMD. Finally, we measured the enrichment for AMD GWAS association signal using FGWAS and found that both iPSC-RPE and fetal RPE peak regions had higher enrichment for AMD risk than adult RPE peak regions, suggesting that iPSC-RPE chromatin accessibility profiles could effectively prioritize regulatory AMD risk variants. Together, these findings support the use of iPSC-RPE as a model system to identify functional regulatory variants associated with AMD.

1933T

ClinTAD: A tool for copy number variant interpretation in the context of topologically associated domains. *J. Spector, A. Wiita.* University of California, San Francisco, San Francisco, CA.

Copy number variants (CNVs) have been associated with a variety of diseases in humans, including autism, intellectual disability, and congenital malformations. For this reason, the American College of Medical Genetics guidelines recommend genomic microarrays for CNVs as "first-tier" tests for the postnatal evaluation of individuals with these conditions. Current guidelines for interpreting CNVs mainly focus on examining genes within the CNV. However, emerging data suggests that alterations in one genomic region may affect gene expression distally via changes in regulatory elements. One important feature of chromatin architecture and gene regulation is topologically associated domains (TADs), with one recent study suggested that as many as 11.8% of pathogenic deletions may exert an effect through disruption of TADs (Ibn-Salem, et al., 2014). We developed ClinTAD, a browser-based tool that can be used to aid in determining the clinical significance of a CNV in the context of TADs. ClinTAD allows a user to input a chromosome number, coordinates, and phenotypic information as a Human Phenotype Ontology (HPO) ID number for either a single case or multiple cases. The single case view returns a visualization showing the CNV, TAD boundaries, and nearby genes, with genes that have phenotypic matches being highlighted. The multiple case view allows the user to enter information about multiple cases at once, and returns a text document with the HPO phenotype matches for each case. We include examples where ClinTAD could be of use in clinical practice, with a particular focus on CNVs of uncertain clinical significance. We specifically discuss three case vignettes where the CNV is predicted to disrupt TAD architecture and genes within the same or neighboring TAD may potentially relate to the patient phenotypes. Alterations in expression of these genes could potentially be used in a clinical interpretation, and these associations would not be suggested by examining genes with the CNV region alone. We anticipate that this decision support tool, freely available at www.clintad.com, will be of widespread utility in clinical practice. We further suggest that it may be time to move beyond the current "CNV-region-only" perspective of array interpretation.

1934F

Disrupted chondrocyte Sox9 expression underlies growth retardation in a mouse model of Kabuki syndrome. J.A. Fahrner^{1,2}, R.C. Riddle³, W.Y. Lin⁴, V.B. DeLeon⁵, S. Chopra¹, S.E. Lad⁴, K.D. Hansen^{1,5}, H.T. Bjornsson^{1,2,6,7}. 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD; 2) Department of Pediatrics, Johns Hopkins University, Baltimore, MD; 3) Department of Orthopaedic Surgery, Johns Hopkins University, Baltimore, MD; 4) Department of Anthropology, University of Florida, Gainesville, FL; 5) Department of Biostatistics and Center for Computational Biology, Johns Hopkins University, Baltimore, MD; 6) Landspítali University Hospital, 101 Reykjavik, Iceland; 7) Faculty of Medicine, University of Iceland, Reykjavik, Iceland.

Mendelian disorders of the epigenetic machinery disrupt fundamental biological processes, most commonly neurological development and growth. One such disorder is Kabuki syndrome 1 (KS1), characterized by intellectual disability, postnatal growth retardation, and distinct dysmorphisms. KS1 is caused by de novo heterozygous loss of function mutations in *KMT2D*, encoding an epigenetic writer that methylates lysine 4 on histone H3 (H3K4me). Prior work demonstrated that a mouse model of KS1 (*Kmt2d*^{fl/geo}) exhibits features of the human disorder, including neurological dysfunction, growth retardation, and facial flattening. Little is known, however, about the mechanistic basis of skeletal growth retardation and craniofacial development in KS1. Here we show that *Kmt2d*^{fl/geo} mice indeed exhibit generalized growth retardation and flattening of the facial profile. High-resolution micro-computed tomography (micro-CT) revealed shortened, thinned long bones with decreased trabecular bone, and morphometric analysis showed ventral bowing of skulls in *Kmt2d*^{fl/geo} mice. Furthermore, a unifying mechanism involving disrupted enchondral ossification at both sites is suggested by in vivo expansion of growth plates at these two distinct regions of skeletal growth. In vitro studies of stable ATDC5 chondrocyte cell lines harboring loss of function and hypomorphic mutations in *Kmt2d* (*Kmt2d*^{-/-} and *Kmt2d*^{ΔR5551/-}), revealed that upon induction of chondrogenesis both exhibited increased proliferation and precocious differentiation in a dose-dependent manner, further supporting the above mechanism. Explicitly, cartilage-specific Alcian blue staining and expression of markers of chondrocyte differentiation are increased in *Kmt2d*^{ΔR5551/-} and *Kmt2d*^{-/-} cells compared to *Kmt2d*^{+/+} cells. We hypothesized a mechanism involving SOX9, which is known to induce chondrogenesis and inhibit osteogenesis. In fact, *Sox9* showed over 10-fold higher expression levels in *Kmt2d*^{-/-} compared to *Kmt2d*^{+/+} cells. SOX9-mediated enhanced chondrogenesis at the expense of osteogenesis provides a potentially novel and plausible explanation for the increased proliferation and precocious differentiation of chondrocytes whilst long bones are shortened with decreased bone formation. Further studies, including genome-wide transcriptome analyses, may determine the basis of *Sox9* dysregulation, implicate additional pathways, and suggest novel therapeutic strategies for growth retardation in KS1 and related disorders.

1935W

Identification of putative epigenetically regulated key genes and pathways related to nonsyndromic cleft lip and palate using genome-wide DNA methylation analysis. N.M. Saiyed¹, S. Vishweswaraiah², S.S. Chettiar³, S. Bakshi⁴, N.K. Jain³, U. Ratnamala⁴, U. Radhakrishna². 1) Biotechnology, Nirma Institute of Science, Nirma University, Ahmedabad, Gujarat, India; 2) Department of Obstetrics and Gynecology, Oakland University William Beaumont School of Medicine, Royal Oak, MI, USA; 3) Department of zoology, School of sciences, Gujarat University, Ahmedabad-9, India; 4) Department of pharmacology, Creighton University, Omaha, NE.

Nonsyndromic cleft lip with or without cleft palate (NSCL/P) is considered the most common, severe craniofacial disorder having both environmental and genetic etiologies. NSCL/P abnormalities are associated with increased morbidity and mortality in childhood, and adulthood. Progress has been made in elucidating the genetic mechanisms of familial NSCL/P; however, an increasing body of evidence implicates epigenetic processes as playing a role in adversely influencing orofacial development. To elucidate the epigenetic mechanism involved in the susceptibility and development of NSCL/P, genome-wide DNA methylation was performed on blood samples from 48 individuals including 24 cases with NSCL/P, and 24 controls. DNA methylation was examined using the Illumina Infinium MethylationEPIC array. Comparison of the methylation profiles between NSCL/P and controls revealed 636 differentially-methylated CpG loci associated with 672 unique genes that had at least 2.0-fold change in methylation (false discovery rate (FDR) $p \leq 0.00001$) or $p < 5 \times 10^{-8}$ with a receiver operating characteristic area under the curve (AUC) ≥ 0.75 . These genes included *COL4A1*, *COL4A2*, *TP63*, *TBX1*, *SUMO1*, *MSX1*, *HOXA2*, *PAX6*, *WNT6*, *NF2*, *NKX2-5*, *DLX5*, *ALX4*, *CLPTM1* etc., which have previously been reported to be associated with NSCL/P. *RPS6KA5/S6K1*, *DDAH2*, *IQGAP3*, *DAB2*, *ALDH9A1*, *KCNN1*, *GLIS3*, *CASP8AP2*, *BRD8*, *MLXIP*, *NIF3L1*, *ITGA3*, *ITGB4*, *WNT7B*, *PPP2R1B*, *DAPK3*, *RAD21*, *RPL23*, *STK36*, *DYRK1B*, *HAND1*, *ACTA1*, *TCF4*, *BACH2* and *RHOA* are the genes identified in the present study in significant association with NSCL/P. Pathway analyses indicated that alterations in the growth hormone, *Cdc42*, GABA receptor, P38 MAPK and protein kinase, axonal guidance, ephrin receptor, Wnt-B catenin, sonic hedgehog, BMP, JAK/Stat, endothelin-1, and epithelial adherens junction signaling pathways are associated with risk of NSCL/P. These results strongly supports proof-of-concept that NSCL/P is associated with differential DNA methylation in whole blood, which may have utility as biomarkers of NSCL/P risk.

1936T

Therapeutic trial with an HDAC inhibitor in Rubinstein-Taybi syndrome.

D. Lacombe^{1,2}, M.P. Baudier¹, E. Bestaven³, S. Fraisse¹, E. Taupiac¹, E. Grech¹, I. Guillaud¹, P. Fergelot^{1,2}, J. Van-Gils^{1,2}, A. Gimbert⁴, F. Salvo⁴, G. Catheline⁵, C. Bader⁶, N. Moore⁷, P. Perez⁶, J.R. Cazalet⁶, B. Arveiler^{1,2}. 1) Medical Genetics, CHU Bordeaux; 2) INSERM U1211, Université de Bordeaux, Bordeaux, France; 3) Institut de Neurosciences Cognitives et Intégratives d'Aquitaine, Université de Bordeaux; 4) Direction de la recherche clinique, CHU Bordeaux; 5) Département d'Imagerie Médicale, CHU Bordeaux; 6) Service d'information médicale, USMR, CHU Bordeaux; 7) Service de Pharmacologie Médicale, CHU Bordeaux, Bordeaux, France.

The Rubinstein-Taybi syndrome (RTS) is a multiple congenital anomalies – intellectual deficiency syndrome characterized by growth retardation, microcephaly, characteristic facial features, and broad thumbs and first toes. RTS is genetically heterogeneous with mutations in the *CREBBP* or *EP300* genes encoding transcription cofactors involved in histone acetylation. RTS knock-out mouse studies showed an improvement of long-term memory using histone deacetyltransferase (HDAC) inhibitors or inhibitors of phosphodiesterase 4 (PDE4). This suggests therapeutic options for RTS. Among the identified HDAC inhibitors, sodium valproate has the marketing authorization for epilepsy and mood disorders and can go through the blood brain barrier. We conducted an exploratory phase 2 therapeutic trial («proof of concept»), with a primary endpoint to estimate the efficiency of sodium valproate after one year of treatment (30 mg/kg/d) on long term memory in RTS children. Secondary Outcomes included cognitive profiles, brain imaging, motor skills evaluation, and histone acetylation profile. It was a monocentric, double-blind, phase 2 trial, designed as a one-step Fleming with a control group, randomized 2:1 into two parallel groups. 41 RTS cases between ages 6 and 21 participated. Using subtests of neuropsychological battery test especially designed for memory evaluation (point location for CMS, and image recognition for RBMT), results were not significant between the treatment group and the placebo group for the primary outcome. The same essay with the RTS knock-out mouse is currently in process. However regarding motor skills for walking, it was noticed a slight decrease of cadence and an increase of the duration of the double step in the valproate group, showing an amelioration in the treatment group. Posturometry showed more oscillations in the placebo group during the task. This confirms the *CREBBP* and *EP300* roles in motor skill learning, and encourages to continue therapeutic approaches with HDAC (or PDE4) inhibitors in RTS. Grant references: PHRC 2011, Ministry of Health; French Rubinstein Taybi syndrome Fondation. .

1937F

Investigation of the effect of allelic deletion of an enhancer element in the critical region of the 9p deletion syndrome using CRISPR/Cas9 and real-time quantitative PCR. X. Hauge¹, A. Gaede¹, L. Manera¹, D. Forteh-Ngochi¹, J. Sloan², K. Traver³. 1) Kennesaw State University, Kennesaw, GA; 2) EGL Eurofins Clinical Diagnostics, Tucker, GA; 3) Oncology Cytogenetics, Emory University School of Medicine, Atlanta, GA.

Deletions of the terminal region of the short arm of chromosome 9 are associated with trigonocephaly, dysmorphic facial features and mental retardation. The deletion of Cerberus 1 (*CER1*) gene located at 14.7 Mb from the 9p terminus has been postulated as the cause of trigonocephaly. However, several independent studies showed that patients with typical 9p deletion phenotype possess smaller than 12.4 Mb deletions. It raises the possibility that cis-regulatory elements for the *CER1* gene could be located in the first 12.4 Mb region of 9p. We have previously identified 8 conserved non coding sequences in the deletion breakpoint region between 10 Mb and 12.4 Mb of 9p. Of these, we found a strong enhancer element 5008 and a strong repressor 3362. Enhancer element 5008 showed a cell-type specific activity in a dual luciferase assay. It increased the luciferase activity 14 fold in a human mesenchymal progenitor cell line derived from H9 hESCs as compared to a 3.5-fold increase of the luciferase activity in the HEK293T cells. We also detected an increase in the *CER1* gene expression in the mesenchymal progenitor cell line after the treatment with osteogenic differentiating medium. We utilized *S. aureus* Cas9 and 4 guide RNAs to excise the enhancer element 5008. We created 4 different constructs with deletions ranging from 500 bp to 2000 bp, which correspond to the core sequence of enhancer element 5008 and the surrounding region. We established 3 different deletion clones from HEK293T cells. The pre-selected deletion breakpoint in the 12.4 Mb genomic region of 9p was verified for each clone by using Sanger sequencing. The effect of allelic deletion of the enhancer element 5008 on the *CER1* gene expression was investigated in each deleted clone using real-time quantitative PCR. The preliminary data showed that the expression of *CER1* gene was decreased in mesenchymal progenitor cells possessing a deletion of the enhancer element 5008. However, further study is needed to confirm the result.

1938W

CHARGE Syndrome in a 22 year old patient of Bogota-Colombia: Report case with a c.5458C>T mutation of the gene encoding CHD7. *D. Arteaga¹, O. Beltran^{1,4}, C. Restrepo^{2,3}*. 1) Grupo de Biogenética & Bioderecho Facultad de Medicina de la Universidad Militar Nueva Granada, Bogotá D.C, Transversal 3ra. No 49-00; 2) Facultad de Medicina, Universidad del Rosario, Bogotá, Colombia, Calle 12C No. 6-25; 3) Genética Molecular de Colombia, Bogotá, Avenida carrera 68 #67F-16; 4) Departamento de genética Organización Internacional Colsanitas, Colombia, Bogota, Consultorio, Carrera 9 #103A-36.

CHARGE syndrome is a very rare genetic syndrome that results in a combination of multiple organ anomalies because of an epigenetic mutation given by the loss of function of the gene encoding the Chromodomain Helicase DNA binding protein 7 (CHD7). This protein is an ATP dependent chromatin remodeler and hence, it is essential for the chromatin modifications that customize the DNA information that is encoded dependent of the cell stage or in specific conditions. Recent evidence suggest the importance of epigenetic mechanism for the central nervous system development, and in this case, in multiple organ features like the iris, ears, genitals, heart, and bones. In patients with clinically typical CHARGE syndrome, the detection rate of this mutation is over 90%, but in patients suspected of having this syndrome the sequencing of CHD7 reveals a mutation in 32-41% of this patients. In 2004 another mutation arise as a cause of this syndrome in suspected cases, known as the SEMA3E, and there are other genetic alterations that can phenocopy this syndrome, like other chromosomal aberrations (like 22q11.2 deletion that involves the gene encoding T-box 1 – TBX1), teratogen exposure (like retinoic acid), or even maternal diabetes. The purpose of this article is to expose a clinical definite case of CHARGE syndrome in a 22 year old male patient in Bogotá, Colombia, the only son from healthy non-consanguineous parents, which meets three major diagnostic characteristics (VIII disfunction, outer ear malformation and choanal atresia) and four minor characteristics (genital hypoplasia, facial asymmetry, developmental delay and growth deficiency), and was study by molecular sequencing of CHD7 and SEMA3E in blood samples. The study results in a nonsense mutation c. 5458C>T of the gene encoding CHD7, confirming the diagnosis of CHARGE syndrome. .

1939T

DNA methylation signatures of BAF complex disorders demonstrate functional continuum of Coffin-Siris and Nicolaides-Baraitser syndromes. *B. Sadikovic^{1,2}, E. Aref-Eshghi^{1,2}, E.G. Bend³, R.L. Hood⁴, L.C. Schenkel^{1,2}, D.A. Carere², R. Chakrabarti⁵, S.C.S. Nagamani⁶, S.W. Cheung⁵, P.M. Campeau⁷, C. Prasad⁸, V.M. Siu⁵, L. Brady⁸, M.A. Tarnopolsky⁹, D.J. Callen⁸, A.M. Innes⁹, S.M. White¹⁰, W.S. Mechino¹¹, A.W. Shuen⁵, G. Pare¹², D.E. Bulman⁴, P.J. Ainsworth^{1,2}, H. Lin^{1,2}, D.I. Rodenhiser^{5,13}, R.C. Hennekam¹⁴, K.M. Boycott⁴, C.E. Schwartz², Care4Rare Canada Consortium*. 1) Department of Pathology and Laboratory Medicine, Western University, London, ON, Canada; 2) Molecular Genetics Laboratory, Molecular Diagnostics Division, London Health Sciences Centre, London, ON, Canada; 3) Greenwood Genetic Center, Greenwood, SC, USA; 4) Children's Hospital of Eastern Ontario Research Institute, University of Ottawa, Ottawa, ON, Canada; 5) Children's Health Research Institute, London, ON, Canada; 6) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA; 7) Department of Pediatrics, University of Montreal, Montreal, QC, Canada; 8) Department of Pediatrics, McMaster University, Hamilton, ON, Canada; 9) Department of Medical Genetics, Alberta Children's Hospital Research Institute for Child and Maternal Health, University of Calgary, Calgary, AB, Canada; 10) Department of Paediatrics, University of Melbourne, Melbourne, Australia; 11) Genetics Program, North York General Hospital, Toronto, ON, Canada; 12) Department of Pathology and Molecular Medicine, McMaster University, Hamilton, ON, Canada; 13) Department of Pediatrics, Biochemistry and Oncology, Western University, London, ON, Canada; 14) Department of Pediatrics, Academic Medical Center, University of Amsterdam, Amsterdam, Netherlands.

Coffin-Siris and Nicolaides-Baraitser syndromes (CSS and NCBRS) are two Mendelian disorders caused by mutations in subunits of the BAF chromatin remodeling complex. We report overlapping peripheral DNA methylation signatures in individuals with various CSS subtypes (*ARID1B*, *SMARCB1*, and *SMARCA4*) and NCBRS (*SMARCA2*). We demonstrate that the degree of similarity in epi-signatures of some subtypes of CSS and NCBRS can be greater than that within CSS, indicating a link in the functional molecular basis of the two syndromes. We show that chromosome 6q25 microdeletion syndrome, harboring *ARID1B* deletions, also exhibits the same CSS/NCBRS methylation profile. Specificity of this epi-signature was confirmed across a wide range of neurodevelopmental conditions including other chromatin remodeling and epigenetic machinery disorders. We demonstrate that a machine-learning model trained on this profile can resolve ambiguous clinical cases, reclassify those with variants of unknown significance, and identify previously undiagnosed subjects through targeted population screening.

1940F

Identification of a unique DNA methylation signature associated with Nicolaides-Baraitser syndrome. R. Weksberg^{1,2}, R. Ejaz¹, E. Chater-Diehl², S. Choufani², C. Cytrynbaum¹, D. Chitayat^{1,3}. 1) Division of Clinical and Metabolic Genetics, Department of Paediatrics, The Hospital for Sick Children, Toronto, Canada; 2) Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Canada; 3) The Prenatal Diagnosis and Medical Genetics Program, Department of Obstetrics and Gynecology, Mt. Sinai Hospital, Toronto, Canada.

The critical role of epigenetic regulation in human development has recently emerged. Hundreds of epigenes modulate a complex interaction of layered epigenetic marks, including DNA methylation (DNAm). Pathogenic variants in many of these epigenes cause various neurodevelopmental disorders and have been shown to be associated with unique epigenome-specific patterns of genome-wide DNAm. The SWI/SNF chromatin remodeling complex is a key developmental regulator of gene expression. Heterozygous pathogenic variants in *SMARCA2*, a component of the SWI/SNF complex, cause Nicolaides-Baraitser syndrome (NCBRS), characterized by facial dysmorphism, sparse hair, seizures and intellectual disability. Given the role of the SWI/SNF complex in transcriptional regulation, we hypothesized that individuals with NCBRS would have a unique DNAm signature that molecularly reflects *SMARCA2* loss of function and could contribute to the disorder's pathophysiology. DNA from whole-blood of two females and two males with NCBRS and pathogenic variants in *SMARCA2* were subjected to genome-wide DNAm analysis along with controls (n=25) using the Infinium MethylationEPIC Bead-Chip. Raw data were processed in the *Minfi* Bioconductor package, including background correction, Illumina normalization, and quality control. *Limma* regression with sex and age as covariates was used to identify differentially methylated CpG sites, and *bumphunter* to identify differentially methylated regions (DMRs). Gene Ontology analysis characterized genes overlapping DNAm changes. We found 1031 CpG sites with a false discovery rate (FDR) q-value < 0.05 and a beta (DNAm) value difference > 10% between NCBRS cases and controls. Two thirds of the DMRs were hypermethylated in NCBRS cases. DMRs were enriched in pathways pertaining to skin keratinization and elasticity, energy metabolism, histone acetylation and channelopathies, offering unique insight into the molecular changes underpinning NCBRS. Hypermethylated regions were enriched at genes involving transcriptional regulation. In this preliminary study, we identified a robust genome-wide DNAm signature for NCBRS. The results further our understanding of downstream target genes of *SMARCA2*. Identification of epigenetically driven biological pathways via the DNAm signature can be used for precision medicine by providing means of functionally classifying variants of unknown significance in this gene, and identifying novel therapeutic targets.

1941W

Maternal effect variants in subcortical maternal complex members cause imprinting disorders. T. Eggermann, L. Soellner, M. Begemann, I. Kurth, M. Elbracht. RWTH Aachen, Aachen, Germany.

The subcortical membrane complex (SCMC) comprises oocyte-derived proteins which are essential for the progression of the zygote and the early embryo. They are encoded by maternal effect genes as they are only expressed in the oocytes, highly abundant in early embryogenesis and then degraded without compensation by the embryonal genome. Though the precise physiological role of SCMC genes is unknown, animal models indicate their fundamental role in proper embryonal development. In humans, women carrying mutations in SCMC factors, i.e. KHDC3L, PADI6, TLE6 and NLRP7, are prone to reproductive failure, including complete hydatidiform moles and miscarriages. Additionally, maternal effect mutations in some of the SCMC genes are also associated with disturbed imprinting marks in their offspring. These aberrant imprints often affect numerous loci, resulting in a multilocus imprinting disturbance (MLID). MLID has firstly been reported in the imprinting disorders Beckwith-Wiedemann and Silver-Russell syndromes (BWS, SRS). Both diseases are molecularly heterogeneous, but in both the major epigenetic and genetic alterations affect the Imprinting Control Regions 1 (ICR1 in SRS) and ICR2 (in BWS) in 11p15. In BWS, loss of methylation (LOM) of the ICR2 is the major finding (~50% of patients), and in this group 25% of patients show a MLID. In SRS, 10% of the subgroup of ICR1 LOM (40% of SRS patients) exhibit MLID. In fact, the vast majority of MLID carriers reported so far have been clinically ascertained either as BWS or SRS, but cases with unspecific phenotypes have also been identified. The pathomechanism behind MLID is unknown, but it has been postulated that in more than 50% of mother of MLID patients variants in SCMC genes (e.g. NLRP2, NLRP5, NLRP7, PADI6) are present. We will report on genetic, epigenetic and clinical data in 29 MLID families, including mutation data in SCMC genes, epigenotype-phenotype correlations in the patients and reproductive history of their mothers. Based on these data, we will emphasize the relevance to identify MLID carriers among imprinting disorder as the basis for personalized genetic and reproductive counseling.

1942T

Utilizing dCas9 directed chromatin loop reorganization to unsilence the paternal *UBE3A* allele in LUHMES neuronal cell line. O. Gutierrez Fugón^{1,2,4}, D.H. Yasui¹, H. O'Geen^{2,3}, D.J. Segal^{2,3,5}, J.M. LaSalle^{1,2,5}. 1) Department of Medical Microbiology and Immunology, University of California, Davis, CA; 2) Genome Center, University of California, Davis, CA; 3) Department of Biochemistry and Molecular Medicine, University of California, Davis, CA; 4) Integrative Genetics and Genomics, University of California, Davis, CA; 5) UC Davis MIND Institute, Davis, CA.

Angelman syndrome (AS) is a severe neurogenetic disorder caused by functional loss of the *UBE3A* gene. *UBE3A* encodes for ubiquitin-protein ligase E3A that is essential for synaptic development. This gene is biallelically expressed in non-neurons but in neurons it is imprinted on the paternal allele and is exclusively expressed from the maternal allele. Most AS cases are caused by a deletion of the maternal *UBE3A* resulting in no production of UBE3A in the brain of these patients. An antisense transcript of *UBE3A* (*UBE3A-ATS*) has been shown to be responsible for the silencing of the paternal allele. In neurons, transcription of *UBE3A-ATS* initiates at *SNRPN*, extends through *IPW* and continues through *UBE3A*, however in non-neurons it terminates near *IPW*. CTCF is a genomic insulator, which binds DNA in a methylation sensitive manner, to form chromatin loops that have been shown to regulate gene expression. A CTCF binding site is present near *IPW*, therefore we hypothesized that CTCF creates a chromatin loop at this locus in non-neurons that prevents the transcription of *UBE3A-ATS* past *IPW*. Furthermore, CTCF is evicted by DNA hypermethylation in neurons, allowing extension of the antisense transcript past *IPW*. As a neuronal model system, we tested Lund Human Mesencephalic (LUHMES) cells, originally derived from the ventral mesencephalon of an 8-week-old fetus. This cell line was engineered to selectively proliferate, allowing controlled neuronal differentiation. In LUHMES cells, we detected *UBE3A-ATS* by qRT-PCR following 7 days of neuronal differentiation but not in the undifferentiated proliferating state. The results demonstrate that these cells can model neuronal versus non-neuronal transcriptional patterns at *UBE3A*. In addition, DNA methylation analysis by pyrosequencing has demonstrated that the CTCF binding site at *IPW* is hypermethylated in neurons compared to non-neurons. In future planned experiments, to determine if a CTCF chromatin loop is sufficient to inhibit transcription of *UBE3A-ATS* past *IPW* in neurons, we will target a dCas9-CTCF chimeric protein to the *IPW* binding site. To determine if CTCF binding at *IPW* is necessary for the termination of *UBE3A-ATS* in non-neurons we will express a dCas9/DNA methyltransferase DNMT3A fusion protein in undifferentiated LUHMES cells. This research approach may provide insight into how chromatin looping regulates paternal *UBE3A* silencing in neurons and could lead to the development of epigenetic therapeutics for AS.

1943F

Comparisons of chromosome-wide DNA methylation patterns between maternal and paternal UPD16 reveal evidence for additional imprinted regions. K.V. Schulze¹, P. Szafranski¹, H. Lesmana², R.J. Hopkin², A. Hamvas³, J.A. Wambach⁴, C.M.B.C. Fonseca¹, Q. Liu¹, J. Karolak¹, J.R. Lupski^{1,5,6}, N.A. Hanchard^{1,6,7}, P. Stankiewicz¹. 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Division of Human Genetics, Cincinnati Children's Hospital Medical Center, University of Cincinnati College of Medicine, Cincinnati, OH; 3) Pediatrics, Northwestern University Feinberg School of Medicine, Chicago, IL; 4) Division of Newborn Medicine, Edward Mallinckrodt Department of Pediatrics, Washington University School of Medicine, St. Louis, MO; 5) Dept. of Pediatrics, Baylor College of Medicine, Houston, TX; 6) Texas Children's Hospital, Houston, TX; 7) USDA/ARS/Children's Nutrition Research Center, Baylor College of Medicine, Houston, TX.

Trisomy 16 is the most frequent trisomy in pregnancy and is only viable with trisomic rescue, which in 1/3 cases results in maternal uniparental disomy (UPD(16)mat). In contrast to the very rare UPD(16)pat thought to be benign, UPD(16)mat can manifest with a range of congenital malformations, e.g. heart defects and pulmonary hypoplasia. Interestingly, similar anomalies have been observed in newborns with the lethal lung disease alveolar capillary dysplasia with misalignment of pulmonary veins (ACDMPV), caused by point mutations or CNV deletions involving the *FOXF1* locus at 16q24.1. Unlike point mutations, CNV deletions arise de novo exclusively on the maternal allele. Given the parental allelic bias in ACDMPV as well as the phenotypic differences between UPD(16)mat and -pat, we hypothesized that there may be additional imprinted or partially imprinted loci on chromosome 16q24.1 that mediate the influence of structural variants on the development of congenital malformations. Chromosome 16-targeted bisulfite sequencing, capturing >75% of CpG loci, was undertaken on paired peripheral blood samples with either UPD(16)mat, UPD(16)pat, and normal controls. Bioinformatic analyses revealed 18 regions with evidence of parent-of-origin biased DNA methylation (≥40% methylation difference between UPD(16)mat and -pat with intermediate methylation in controls). We confirmed known imprinted loci near *ZNF597* and found evidence of maternal methylation at a CpG island close to *PRR25*, which is reported to be paternally expressed in lymphoblastoid cells. Methylation differences between maternal and paternal alleles were less pronounced in non-imprinted regions; however, we found evidence of hypomethylation in UPD(16)pat at *NARFL*, *PMM2*, and *WWOX* as well as partial imprinting at *CCPED1*, *MGRN1*, and an intergenic region on 16q24.1; the latter contained an additional 138 regions with modest evidence of parent-of-origin methylation biases. Bisulfite sequencing of UPD(16)mat fibroblasts, lung ACDMPV samples, and a lung control indicated that some parent-of-origin biased methylation patterns were consistent across tissues while others appeared tissue-specific. Given that only four of our 18 candidate regions are covered by the commonly used 450K methylation array, the high resolution of our approach allowed us to find evidence of new candidate imprinted loci on chromosome 16 that could contribute to the birth defects seen in UPD(16)mat and ACDMPV.

1944W

Complete lung agenesis in three fetuses harboring a 659kb 7q36

TAD-disrupting tri/tetraplication neighboring *SHH*. L. Van Maldergem¹, B. Fischer-Zirnsak², P. Blanc³, V. Roze⁴, P.A. Hidalgo⁵, E. Landais⁵, V. Guigue⁶, D. Gaillard⁷, V. Kremer⁸, R. Ramanah⁹, D. Riethmuller⁹, C. Cabrol¹, P. Kuentz⁴, F. Arbez-Gindre⁹, F. Harms¹⁰, U. Kornak², K. Kutsche¹⁰, S. Mundlos², J. Piard¹.

1) Ctr Hum Genet, University of Franche-Comté, Besançon, France; 2) Inst Hum Genet, Charité, Univ of Berlin, Berlin, Germany; 3) Dept Med Genet, Hospices Civils de Lyon, Lyon, France; 4) Cytogenetics Lab, University Hosp, Univ Franche-Comté, Besançon, France; 5) Cytogenetics Lab, Univ Hospit, Univ Champagne-Ardenne, Reims, France; 6) Dept Gyn Obst, Univ Hospit, Besançon, France; 7) Dept Pathol, Univ Hospit, Univ Champagne-Ardenne, Reims, France; 8) Cytogenetics Lab, Univ Strasbourg, Strasbourg, France; 9) Dept Pathol, Univ Hospit, Besançon, France; 10) Inst Hum Genet, Univ Hamburg, Hamburg, Germany.

Development of the respiratory system is a complex and tightly controlled process. The canonical Hedgehog pathway acts as a critical piece in the lung development signaling puzzle. The inactivation of *Shh* and *Gli* genes in mice results in foregut malformations ranging from lung hypoplasia and abnormal tracheoesophageal septation to severe tracheal truncation patterns and complete lung agenesis. Malformations such as lung agenesis are rare events usually seen as sporadic cases and the molecular basis remains poorly understood. Here we describe three fetuses born to unrelated French parents with complete lung agenesis. The diagnosis was made by ultrasound. Fetal autopsy after termination of pregnancies at 26, 23 and 14 weeks revealed complete agenesis of both lungs in all three fetuses. Furthermore, two of them had moderate interdigital webbing. The male fetus had a micropenis, and in another the trachea was partially absent. We performed exome sequencing in two of them and found no convincing coding alteration. However, high resolution array-CGH revealed an approximately 659 Kb tri/tetraplication in all fetuses. Interestingly, the same structural variant was detected in the unaffected father, although in a mosaic state. The candidate CNV lies along *SHH* long-range regulatory domain and encompasses a specific enhancer, *MACS1*, driving the spatiotemporal expression of *SHH* in the airway epithelium. We speculate that *MACS1*, becomes disconnected from its gene through the formation of a new chromatin domain (neo-TAD) resulting in the ectopic activation of a neighboring gene, probably *MXN1*.

1945T

The chromatin accessibility signature of aging in human blood leukocytes stems from CD8+ T cells. D. Ucar¹, E. Marquez¹, C. Chun¹, R. Marches¹, R. Rossi¹, A. Uyar¹, T. Wu¹, J. George¹, M. Stitzel¹, K. Palucka¹, G. Kuchel², J. Banchereau¹.

1) Jackson Laboratory, Farmington, CT; 2) University of Connecticut Health Center, Farmington, CT.

Aging is linked to deficiencies in immune responses and increased systemic inflammation. To unravel regulatory programs behind these changes, we applied systems immunology approaches and profiled chromatin accessibility and transcriptome in peripheral blood mononuclear cells (PBMCs), and purified monocytes, B and T cells. Analysis of samples from 77 young and elderly donors revealed a novel and robust aging signature in PBMCs with simultaneous systematic chromatin closing at promoters and enhancers associated with T cell signaling, and, a potentially stochastic, chromatin opening mostly found at quiescent and repressed sites. Combined analyses of chromatin accessibility and transcriptome uncovered immune molecules activated/inactivated with aging, and identified the silencing of *IL7R* gene and the IL-7 signaling pathway genes as potential biomarkers. This signature is borne by memory CD8+ T cells, which exhibited an aging-related loss in binding of NF-κB and STAT factors. Thus, our study provides a unique and comprehensive approach to identify candidate biomarkers as well as mechanistic insights into aging associated immunodeficiency.

1946F

Persistence of differential DNA methylation related to prenatal smoking exposure and the effects on later life health outcomes. V. Karhunen¹, P. Wiklund², M. de Silva³, R. Richmond^{3,4}, F. Rezwan⁵, M. Wielscher⁶, J. Holloway⁶, A. Rodriguez^{1,6}, S. Sebert⁷, C. Relton^{3,4}, M. Järvelin^{1,2,7,8}. 1) Department of Epidemiology and Biostatistics, Imperial College London, London, United Kingdom; 2) Center for Life Course Health Research, University of Oulu, Oulu, Finland; 3) Medical Research Council Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom; 4) School of Social and Community Medicine, University of Bristol, Bristol, United Kingdom; 5) Human Development & Health, Faculty of Medicine, University of Southampton, United Kingdom; 6) School of Psychology, University of Lincoln, Lincoln, United Kingdom; 7) Biocenter Oulu, University of Oulu, Finland; 8) Unit of Primary Care, Oulu University Hospital, Oulu, Finland.

Maternal smoking during pregnancy is known to alter DNA methylation in newborns. Many of these changes appear to persist throughout childhood and into adolescence but whether they endure into adulthood is unclear. Furthermore, whether changes in offspring DNA methylation mediate the effect of smoke exposure in utero on later life health outcomes is unknown. We assessed the association between exposure to maternal smoking in pregnancy and offspring blood DNA methylation in 2,821 individuals aged between 16 and 48 years from five prospective birth cohort studies. We concentrate on 6,073 CpG sites that has been previously reported as differentially methylated in cord blood of newborns exposed to maternal smoking in utero. We also used two-sample Mendelian randomisation to estimate the effect of maternal smoking related CpG sites on disease outcomes. We found evidence of association with exposure to maternal smoking for 70 differentially methylated CpG sites (P value $< 1 \times 10^{-7}$) in 36 genomic regions in adolescents and adults. All of these CpG sites showed similar direction of effects to those reported in previous studies in newborns. We found a dose-dependent effect of smoking intensity during pregnancy on offspring DNA methylation levels. Sensitivity analyses implied that the associations were neither driven by offspring's own smoking nor confounded by paternal smoking behaviour, suggesting intra-uterine mechanisms. Longitudinal analysis showed no evidence for change in direction or magnitude of the associations, indicating sustained alteration in offspring DNA methylation related to intrauterine smoking exposure. Mendelian randomisation analyses provided evidence for a causal role of four maternal smoking related CpG sites on an increased risk of inflammatory bowel diseases or schizophrenia. Our findings indicate that maternal smoking during pregnancy has long-lasting effects on offspring DNA methylation and that some of these methylation changes may act as the mechanism between maternal smoking and later-life diseases in the exposed offspring.

1947W

A decade of epigenetic change in aging twins: Genetic and environmental contributions to longitudinal DNA methylation. C.A. Reynolds¹, Q. Tan², E. Munoz¹, J. Jylhävä³, J. Hjelmborg³, L. Christiansen², S. Hägg³, N.L. Pedersen³. 1) Psychology, University of California Riverside, Riverside, CA; 2) University of Southern Denmark, Odense, Denmark; 3) Karolinska Institutet, Stockholm, Sweden.

Epigenetic changes may result from the interplay of exposures and genetic influences and contribute to differences in aging outcomes. However, the etiologies contributing to stability and change in DNA methylation has been rarely examined longitudinally. We considered DNA methylation across a ten-year span in two samples of same-sex aging twins: the Swedish Adoption Twin Study of Aging (SATSA, N=53 pairs), and the Longitudinal Study of Aging Danish Twins (LSADT, N=43 pairs). The average age was 62.9 and 72.5 years (SD=7.2) at each wave in SATSA and 76.2 and 86.1 years (SD=1.8) in LSADT. Data processing of methylation probes from the Illumina HumanMethylation450 array included adjustments and normalizations for cell counts and batch effects, resulting in a joint analysis set of 359,399 probes. Biometrical analyses of M-values, adjusted for age, sex and country, were fitted. Bivariate ACE and ADE Cholesky models evaluated the degree to which additive genetic (A), dominance or non-additive genetic (D), common environmental (C), and unique non-shared environmental influences (E) contributed to variation and covariation in M-values within and across time. Scalar differences in variances were modeled for the two samples. Across sites, 58.5% showed significant cross-time associations indicative of stability via genetic and/or environmental mechanisms. At $p < 1 \times 10^{-7}$, 1.9% sites showed significant broad genetic (A, D) or familial effects (A, C) within or across time ($df = 6$), and 10.4% met $p < 1 \times 10^{-2}$. Genetic (r_A , r_D) and shared environmental correlations (r_C) were strong across time with median values of .99, .72, and .87, respectively, versus .16 for nonshared environmental influences (r_E). Mean broad heritability estimates were higher at baseline ($M1 = 23.8\%$) than a decade later ($M2 = 18.0\%$). Broad heritability estimates varied by CpG location ($p < 2.2 \times 10^{-16}$) with higher estimates, at both assessments, for open-sea and shores ($M1 = 24.5\%$; $M2 = 18.6\%$) versus islands ($M1 = 22.6\%$; $M2 = 17.2\%$) and shelves ($M1 = 23.6\%$; $M2 = 17.6\%$). Shared environmental contributions (C) tended to be modest ($M1 = 5.7\%$; $M2 = 5.4\%$). Overall, results suggest genetic contributions to DNA methylation tended to be small in effect size and lessen across 10 years but contributed to stability across time. Nonshared environmental influences largely explained the unique influences emergent across the decade suggesting person-specific experiences become more salient to DNA methylation in late life.

1948T

Mitochondrial DNA copy number (mtDNA-CN) influences nuclear DNA methylation at CpGs associated with neuroactive ligand-receptor pathway interactions.

C.A. Castellani¹, R.J. Longchamps¹, J.A. Sumpter¹, C.E. Newcomb¹, J.A. Lane², M.L. Grove³, J. Bressler⁴, J.A. Brody⁴, J.S. Floyd⁴, K.D. Taylor⁵, A. Tin⁶, J. Coresh⁶, J.S. Pankow⁷, M. Fornage³, E. Guallar⁸, B. O'Rourke⁸, N. Pankratz², N. Sotoodehnia⁴, E. Boerwinkle³, D.E. Arking^{1,8}. 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Department of Laboratory Medicine and Pathology, University of Minnesota School of Medicine, Minneapolis, MN; 3) Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX; 4) Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA; 5) Institute for Translational Genomics and Population Sciences, Los Angeles BioMedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA; 6) Department of Epidemiology and the Welch Center for Prevention, Epidemiology and Clinical Research, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; 7) Division of Epidemiology & Community Health, School of Public Health, University of Minnesota, Minneapolis, MN; 8) Division of Cardiology, Johns Hopkins University School of Medicine, Baltimore, MD.

Mitochondrial DNA copy number (mtDNA-CN) has been associated with a variety of aging-related diseases, including all-cause mortality. The mechanism by which mtDNA-CN influences disease may be through regulation of nuclear gene expression via the modification of nDNA methylation. Supporting this, we report the cross-sectional association of mtDNA-CN and nDNA methylation as well as experimental evidence that modulating mtDNA-CN leads to changes in nDNA methylation and expression. nDNA methylation and mtDNA-CN were derived from DNA extracted from buffy coat for 1,567 African Americans (AA) and 940 European Americans (EA) from the Atherosclerosis Risk in Communities (ARIC) study. Twenty-three and fifteen independent CpGs reached epigenome-wide significance in ARIC AA and EA cohorts, respectively. Effect estimates for significant CpGs in ARIC AA and EA were strongly correlated ($r=0.7$). In addition, five significant CpGs validated in an AA cohort ($N=239$) from the Cardiovascular Health Study (CHS) ($p<0.05$) and 19 of the 23 CpGs showed the expected direction of effect, however this was not the case for the CHS EA cohort ($N=294$) where we were not able to validate any CpGs. Mendelian randomization using methylation QTLs for mtDNA-CN associated CpGs as the mediator demonstrated that nDNA methylation at these CpGs does not directly drive alterations in mtDNA-CN ($\beta=0.001, p=0.83$) and there was no enrichment of p-values. In contrast, experimental modification of mtDNA-CN through knockout via CRISPR-Cas9 of *TFAM*, a regulator of mtDNA replication, demonstrated that modulation of mtDNA-CN drives changes in nDNA methylation and gene expression of specific CpGs/transcripts. Strikingly, the 'neuroactive ligand receptor interaction' KEGG pathway was found to be the top overrepresented pathway in the ARIC meta-analysis ($p=1.56E-03$), as well as *TFAM* methylation ($p=1.35E-04$) and expression ($p=4.30E-04$) studies. These results show that changes in mtDNA-CN influence nDNA methylation at specific loci and lead to changes in gene expression of specific genes, including those acting in the 'neuroactive ligand receptor interaction' pathway and may lead to impacts on human health and disease via altered cell signaling.

1949F

Biological aging is associated with stress and gene expression during pregnancy. A.K. Knight, S. Resin, V. Kilaru, D. Cobb, E.J. Corwin, A.L. Dunlop, A.K. Smith. Emory University, Atlanta, GA., USA.

Stress, the environment, and health conditions can increase rates of biological aging. Biological aging can be measured based on combinations of CpG sites or RNA transcripts, which make up epigenetic and transcriptomic clocks, respectively. The difference between biological and chronological age is termed age acceleration. As pregnancy is a time of vast physiological changes and unique stressors, we would expect pregnancy to be reflected in measures of age acceleration. We leverage data from 255 women from the Emory University African American Microbiome in Pregnancy Cohort to determine whether age acceleration changes over pregnancy. Participants provided a blood sample for DNA methylation (HumanMethylation450 or MethylationEpic) and gene expression (HT-12) analyses, cytokine panels, and glucocorticoid resistance testing between 7-14 weeks and 24-32 weeks gestation. Overall, the methylation-based predictors ($.6<r<.63$; $p<2e-16$) correlated more strongly with biological age than the transcriptomic predictor ($r=.16$; $p=5.6e-7$). Of the four biological age predictors tested (Horvath, Hannum, Levine, Peters), only the transcriptomic-based predictor was associated with changes over pregnancy ($t=2.8$, $p=.005$), likely reflecting the increased stability of DNA methylation as compared to gene expression. However, age acceleration calculated from all four prediction methods was associated with concentrations of at least one inflammatory cytokine and/or glucocorticoid resistance ($p<.05$). Additionally, each of these measures of age acceleration was associated with gene expression (1231-9405 transcripts; $FDR<.05$), with 332 transcripts associating with all four prediction methods ($FDR<.05$). Thus, changes in age acceleration based on any method are likely to be biologically relevant. Each prediction method has unique strengths which must be considered in the design studies evaluating factors that contribute to or result from biological aging.

1950W

Effects of miRNAs that are predominantly present in young mouse plasma on myogenic differentiation and muscle regeneration. *M. Fukuoka¹, N. Ito², S. Takeda², H. Hohjoh¹*. 1) Department of Molecular Pharmacology, National Institute of Neuroscience, National Center of Neurology and Psychiatry, Kodaira, Tokyo, Japan; 2) Department of Molecular Therapy, National Institute of Neuroscience, National Center of Neurology and Psychiatry, Kodaira, Tokyo, Japan.

MicroRNAs (miRNAs) that are 21~23-nucleotide-long small non-coding RNAs function as mediators in gene silencing, and play essential roles in gene regulation by inhibiting translation of messenger RNAs (mRNAs) and by digestion of mRNAs. Recently miRNAs have been found in various body fluids including plasma/serum as well as inside cells, and a part of such miRNAs appear to be enclosed in extracellular vesicles such as exosomes and be associated with RNA-binding proteins, e.g., Argonaute2. Such cell-free (extracellular) miRNAs might be implicated in cell-cell communication and be capable of becoming useful biomarkers. Aging is characterized by a progressive loss of physiological integrity, in which functional declines in physiology such as a decrease in hearing, visual acuity, muscle strength and memory skills occur. To see such aging at a molecular level, we investigated plasma-circulating (cell-free) miRNAs prepared from young (6-week-old) and aged (2-year-old) C57BL6/J mice by expression profile analyses in this study. The results indicated that the level of myogenic miRNAs, which are involved in myogenic differentiation, was higher in the plasma of young mice than in that of aged mice. In addition, young mouse serum, but not aged serum, appeared to have the potential for inducing myogenic differentiation of mouse C2C12 cells (a myoblast cell line). We investigated the effects of the miRNAs increased in young mouse serum on myogenic differentiation using C2C12 cells: miRNAs of interest were introduced into C2C12 cells and the expression of the *Myogenin* and *Myh1* genes that were used as indicators for myogenic differentiation was examined. As a result, a certain miRNA appeared to have a strong induction for myogenic differentiation relative to other examined miRNAs, and might be associated with the increase in the level of miR-1, which plays a role in gene regulation during myogenic differentiation. We further examined the effects of the miRNA on muscle regeneration in injured mice and found that cross-sectional areas of regenerating muscle fibers enlarged in the presence of the miRNA mimics, suggesting that the miRNA might be capable of enhancing muscle regeneration. Our findings suggest that the miRNA may play an important role in gene regulation during myogenic differentiation, and may be related to muscle aging.

1951T

Vegetarian diets, circulating miRNAs and healthspan in humans. *T. Liu, C. Wang*. Center for Genomics, Loma Linda University, Loma Linda, CA.

Abstract: Diet has been shown to play an intricate role in modulating lifespan and aging process through nutrient sensing pathways/systems such as insulin/IGF, mTOR and Sirtuin gene family in animal studies. However, the mechanism of how these pathways are regulated by diet remains unclear in humans. Our recent studies suggested that lifetime diet-modulated epigenomic reprogramming (i.e., DNA methylome) be associated with healthspan while it is unknown if a lifetime vegetarian diet may alter the circulating miRNA expression which may regulate healthspan in humans. We hypothesize that long-term vegetarian diet modulates circulating miRNA expression which may affect the healthspan by regulating gene expression involved in nutrient sensing pathways/systems. We exploited a large (N = 96,000) and long-standing cohort, Adventist Health Study-2 (AHS-2), a well-characterized, group of participants with detailed clinical, dietary and health outcomes follow-up data over many years. The AHS-2 study found that a long-term vegetarian diet is a major determinant of both longevity and maintenance healthspan, but the underlying molecular mechanisms are largely unknown. We have collected plasma samples from the AHS-2 cohort for ~100 subjects with different dietary patterns (vegetarian vs non-vegetarian) and have performed miRNA deep sequencing for 33 subjects out of ~100 plasma samples so far. We identified 19 differentially expressed miRNAs between two dietary groups (2 fold-change plus, $P \leq 0.05$), of which several miRNAs such as miR-320d, miR-664a and miR-203a are associated with lifespan regulation. We also found several miRNAs involved in cell senescence and tumor suppression pathways, i.e., miR-127, miR-22 and miR-140 etc., which were differentially expressed between vegetarians and non-vegetarians. Overall, our study showed that certain circulating miRNAs could serve as good candidate biomarkers for lifetime vegetarian diet-modulated healthspan and longevity in humans. .

1952F

scaRNA1 levels alter pseudouridylation in spliceosomal RNA U2 affecting alternative mRNA splicing and development. C.K. Nagasawa¹, N. Kibiryeve², J. Marshall², J.E. O'Brien Jr.², D.C. Bittel^{1,2}. 1) Kansas City University of Medicine and Biosciences, Kansas City, MO. 64106; 2) Children's Mercy Hospital, Kansas City, MO 64108.

Vertebrate embryonic development is complex and dynamically regulated, yet only partially understood. Over the last few years, there has been a growing interest in understanding the role that alternative mRNA splicing has in regulating developmental and pathological processes. Messenger RNA splicing is carried out by the spliceosome, a multimegadalton ribonucleoprotein complex composed of 5 snRNAs (small nuclear RNAs) and multiple proteins. Small noncoding cajal body-specific RNAs (scaRNAs) biochemically modify (i.e. pseudouridylation, methylation) snRNAs and is essential for proper spliceosomal function. Our previous research suggested that biochemical modification directed by scaRNAs altered mRNA splicing and significantly impacted development. One of those scaRNAs, scaRNA1, is responsible for pseudouridylating nucleotide U89 of the U2 snRNA. Pseudouridylated nucleotides are highly conserved across species and located in functionally important areas of the spliceosome, making it necessary for proper mRNA splicing. Here we present evidence that changes in expression of scaRNA1, either in vitro or in vivo, lead to changes in pseudouridylation accompanied by changes in mRNA splicing and altered development. These experiments provide evidence supporting a new model in the regulation of embryonic development that is unappreciated and needs further investigation.

1953W

Gene-environment interactions in asthma: RV and the 17q locus. B.A. Helling¹, B. Laxman¹, K. Luo¹, Y.A. Bochkov², J.E. Gern², S.R. White¹, C. Ober¹. 1) Human Genetics, University of Chicago, Chicago, IL; 2) Pediatrics, University of Wisconsin Madison, Madison, Wisconsin.

Asthma is the most common chronic childhood illness in the United States. Although, GWAS have identified over 30 loci, these loci explain only a small proportion of asthma risk. This is due in part to the significant contribution of environmental factors and gene-environment interactions. For example, rhinovirus (RV)-associated wheezing in early life is predictive of asthma at school age, and this risk is modified by genotype at 17q12-21, the locus most significantly associated with asthma in GWAS. To identify novel pathways involved in asthma risk and specifically RV-associated asthma risk, we developed a cell culture model using primary bronchial epithelial cells (BECs) from donated lungs not used for transplant. BECs were selected from 9 asthmatic and 9 non-asthmatic individuals. BECs were treated with RV or vehicle for 24 hours and then processed for ATAC-seq to identify regions of open chromatin. ATAC-seq peak calls from each sample were pooled to create a 'peak library' and quality reads from each sample were mapped to this peak library. Total counts per peak were tabulated for each sample. After normalization and correction for relevant covariates, linear regression was used to identify differential chromatin accessibility (DC) between asthmatics and non-asthmatics or RV and vehicle treated samples. We identified 3,531 DCs associated with asthma and 4,821 associated with RV treatment genome-wide (FDR <0.05). At the 17q locus (chr17: 37850000 – 38150000), we identified 2 asthma-associated DCs, one in the transcriptional start site of *ERBB2* and one upstream of *GSDMA*. SNPs in both genes have been associated with asthma in GWAS. The DC upstream of *GSDMA* has predictive features of a transcriptional regulator including a CpG island and binding sites for POLR2A and CTCF in a lung epithelial cell line, A549 (ENCODE). In our BEC culture model, both asthma status and RV treatment had genome-wide effects on chromatin accessibility, and we identified two DCs at the asthma-associated 17q locus. We are currently analyzing RNA-seq data on these samples for integrative analyses of gene regulation in this cell model to ultimately identify novel pathways involved in asthma. This work was supported by P01 HL070831 and U19 AI095230. B.H. is supported on T32 HL007605.

1954T

Tissue-specific RNA editing in diversity outbred mice. A. Srivastava¹, V.K. Sarsani¹, D.A. Skelly¹, D.M. Gatti¹, K. Choi¹, N. Raghupathy¹, E.J. Chesler¹, M.P. Keller², A.D. Attie², R. Korstanje¹, G.A. Churchill¹. 1) The Jackson Laboratory, Bar Harbor, ME, USA; 2) The University of Wisconsin, Madison, WI, USA.

RNA editing is defined as a co- or post-transcriptional change in the sequence of RNA transcripts that can lead to increased RNA and protein diversity. In mammals there are two types of RNA editing: A-to-I editing, which occurs by deamination of adenosine converting it to inosine, and C-to-U editing, which occurs by deamination of cytosine converting it to uracil. RNA editing typically alters only a fraction of the targeted transcripts that are present in a tissue. The fraction of edited transcripts can be influenced by polymorphisms or expression variation in the editing enzymes and by polymorphisms in the target RNA. Editing efficiency can also be tissue-specific and may be influenced by extrinsic factors including sex and age. We hypothesized that a better understanding of how genetic variation in editing enzymes and their RNA targets interact with the contexts of tissue, sex, and age will shed new light on RNA editing mechanisms. To understand the tissue-specific genetic architecture of RNA-editing, we obtained RNA-Seq data from liver, kidney, heart, brain, bone, and pancreatic islet tissues from Diversity Outbred (DO) mice, a genetically diverse multiparent mouse population. We identified RNA-editing sites using a bioinformatics workflow that accounts for known genetic variations in DO mice and minimizes false positive detection. We carried out genetic association mapping of the editing ratio as a quantitative trait and identified both common and tissue-specific features of RNA editing. We found that A-to-I editing ratios are controlled primarily by genetic variation in the target RNA whereas C-to-U editing ratios are primarily associated with a genetic variation at the *Apobec1* locus on mouse chromosome 6. We found significant levels of C-to-U editing only in liver, kidney and pancreatic tissues and observed a striking tissue-specific variation in allelic effects at the *Apobec1* locus. In addition, we found that C-to-U editing in kidney is strongly influenced by sex. A-to-I editing was present across all tissues with broad conservation in the target-specific editing ratios. Bone revealed a pattern of change with age that may be explained by age-related changes in the cellular composition of the marrow compartment. In summary, this study revealed distinct genetic architectures for C-to-U versus A-to-I RNA editing; a common pattern of genetic regulation is shared across most tissues but with tissue-specific influences of genotype, sex, and age.

1955F

A first generation map of genome-wide methylation correlation suggests long-range genomic coordination and has implications for disease-mapping. A. Jajoo, O. Hirschi, K. Schulze, Y. Guan, N. Hanchard. Baylor College of Medicine, Select a Country.

Genome-scale assessments of DNA methylation have been used to identify disease-contributing loci in variety of human diseases, cancers, and traits. These studies have largely focused upon identifying differences in methylation level between disease and non-disease states. Methylation sites in close proximity of each other are known to have sustainable correlation structures (CSs). These inter-relational CSs are described in the literature, but their consistency across individuals and tissues, genomic context, and relevance to disease have remained unexplored. Here, we develop a novel Image Clustering Method (ICM) that leverages image conversion and correlation matrices, to assess patterns of correlated CpG loci on the commonly used Illumina 450KBeaArray. We apply our method to 28 experimentally-derived and publically-available datasets, spanning 27 tissues and 11,000 individuals. We identify 207 clusters of contiguous correlation between adjacent sites that extend a median of 709 kb; these clusters are enriched for CpG island promoters ($P < 0.01$), with between 30 and 50% being conserved across tissues. We also observe significant positive and negative correlation between contiguous clusters extending as far as 250kb. These non-contiguous clusters are bound within topology-associated domains (TADs) and suggest coordinated methylation between CpG promoters, particularly at gene-clusters, as well as between histone methylation marks and promoters. Finally, we identify significant differences in cluster correlation between diseased and non-disease states, including cancer, with disease-specific patterns of disruption.

1956W

Understanding human puberty-associated genes using 3'UTR-seq of the mouse hypothalamus and pituitary gland during postnatal development. H. Hou^{1,2}, L. Uusküla-Reimand^{1,3}, M. Hudson^{1,2}, C. Chan^{1,2}, D. Sokolowski^{1,2}, A. Roy¹, K.E. Yuki¹, A. Goldenberg^{1,4}, M. Palmert^{1,5}, M. Wilson^{1,2}. 1) Genetics and Genome Biology Program, Sickkids Research Institute, Toronto, Canada; 2) Department of Molecular Genetics, University of Toronto, Canada; 3) Department of Gene Technology, Tallinn University of Technology, Tallinn, Estonia; 4) Department of Computer Science, University of Toronto, Canada; 5) Department of Paediatrics, Hospital for Sick Children and University of Toronto, Canada.

The timing of human puberty is highly variable, sexually dimorphic, and associated with adverse health outcomes. It is controlled by both genetics and environmental factors. Recent genome-wide association studies (GWAS) have identified more than 300 loci that are associated with age at menarche (AAM) in women or voice breaking in men. These GWAS studies provide valuable insights into the pathways involved in regulation of pubertal onset. However, little is known about the exact function of the majority of the GWAS loci and their associated genes. To explore the role of the increasing list of GWAS-implicated genes, it is critical to understand their expression patterns. We performed transcriptome-level gene expression profiling in male and female mouse hypothalamus and pituitary at five ages during pubertal development, using an automated 3'UTR-seq method. We showed that gene expression levels obtained using 3'UTR-seq correlate well with those obtained with qPCR, supporting that 3'UTR-seq is an efficient method for large-scale profiling. We identified large numbers of developmentally regulated genes, especially between postnatal days 12 and 22. The hypothalamus undergoes a greater level of gene expression change compared to the pituitary. On the other hand, while the hypothalamus transcriptome is largely similar between male and female mice, pituitary shows abundant sex-biased gene expression patterns, especially upon the onset of puberty. Pseudotime analysis supports that the observed sexual dimorphism in timing is not due to developmental delay in female vs. male mice, but divergence in gene expression between sexes. Using time-series analysis, we identified multiple co-expression gene clusters where functions of known genes could facilitate our understanding of candidate GWAS genes. Finally, we performed small RNA-seq in the same pituitary samples and discovered few sex-biased, but abundant developmentally regulated microRNAs, which potentially contribute to the regulation of GWAS-implicated genes. Transcriptome-level gene expression profiling during mouse pubertal development in a sex- and tissue-specific manner has provided valuable insights on the potential function of GWAS-implicated genes and pathways. Moreover, our study demonstrates that large-scale transcriptome profiling is a feasible and valuable post-GWAS candidate gene prioritization approach.

1957T

A fast and streamlined method for methylome insights from various DNA samples. I. Andreou¹, R. Samara², B. Dugan², S. Huebner¹, C. Meerschiff¹, A. Piotrowski¹, E. Lader². 1) QIAGEN GmbH, Hilden, Germany; 2) QIAGEN Sciences Inc, Frederick, MD, USA.

Epigenetic changes play a crucial role in the regulation of important cellular processes such as gene expression and cellular differentiation, and were also identified as key factors in many various diseases. The methylation of cytosines in the genome, an important epigenetic regulatory mechanism, reduces the transcriptional activity of adjacent genes. Whole genome bisulfite sequencing (WGBS), combining bisulfite-mediated conversion of un-methylated cytosine to uracil and next generation sequencing (NGS), allows the genome-wide detection of 5-methylcytosine residues at unprecedented single-base resolution and thus enables the connection between gene activity and the precise localization of a DNA methylation mark. Clinically relevant specimens for next-generation sequencing can be difficult to come by, are limited in amount, or strongly fragmented. In order to safeguard and use samples as efficiently as possible, solutions that push the limits of input amounts and DNA integrity without sacrificing sensitivity and performance are needed. Here we present a fast and streamlined workflow for optimal bisulfite treatment, and NGS library construction, which overcomes these challenges. The combination of dedicated bisulfite conversion protocols using reagents of the EpiTect Fast Bisulfite Conversion Kit and the ultra-efficient DNA repair, end-polishing and adapter ligation chemistries of the QIAseq Methyl Library Kit allows fast and efficient bisulfite conversion and NGS library preparation to accurately interrogate the methylome starting from low input gDNA, ccfDNA, FFPE-DNA, or enriched CpG islands.

1958F

COPD and IPF overlap GWAS SNP rs2076295 regulates desmoplakin expression. S. Bates, F. Du, B. Pham, C. Benway, F. Guo, W. Qiu, M. Cho, X. Zhou. Brigham and Women's Hospital, Boston, MA.

Genome-wide association studies (GWAS) have identified overlapping loci between Chronic Obstructive Pulmonary Disease (COPD) and Idiopathic Pulmonary Fibrosis (IPF), including a strong co-localized expression quantitative trait locus (eQTL) with desmoplakin (*DSP*). *DSP* is a key component of the desmosome, a cell-cell adhesion structure that links desmosomal cadherins to the intermediate filament (IF) cytoskeleton. The desmosome is widely studied in the skin and heart where the tissues are often subjected to mechanical stress. However, little is known about the localization and function of *DSP* in lung diseases. Interestingly, the same top SNP for these two lung diseases, rs2076295, shows divergent effects, with the COPD risk allele protective for IPF and vice versa. To identify the functional variants in this locus and also address the function of *DSP* in the development of COPD and IPF, we utilized CRISPR/Cas9 based techniques to knockout and overexpress *DSP*, edit rs2076295 to homozygous genotypes of its major and minor allele and observe effects, and delete a 55 base pair area containing the variant. From these experiments, we demonstrated that the rs2076295 region and variant modulate *DSP* expression. Through RNA-seq analysis of the cell lines we generated, we have identified critical signaling pathways affected by *DSP* expression involved in pathogenesis of COPD and IPF. We further examined cellular phenotypes of CRISPR-edited cells and found key regulatory functions of *DSP* important for COPD and IPF development. These mechanistic studies illuminate shared but divergent pathways in COPD and IPF.

1959W

Null findings in epigenome-wide association studies: Lack of power or lack of associations? T. Battram, T.R. Gaunt, N.J. Timpson, G. Heman. MRC Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom.

For epigenome-wide association studies (EWAS) researchers have assumed DNA methylation explains some of the variance of their traits of interest. But interpretation of current and design of future studies cannot be made in an informed manner until it is known how much variation attributable to DNA methylation remains to be found. We estimated the proportion of variance related to DNA methylation at CpGs assayed by the 450k array (m^2_{450k}) for 4847 phenotypes within the ALSPAC cohort ($n = 941$), by producing DNA methylation-kinship matrices and using restricted maximum likelihood analysis implemented in LDAK. EWAS were performed on all 4847 traits and we found a relationship between number of differentially methylated probes (DMPs) identified at $P < 1 \times 10^{-5}$ and m^2_{450k} ($P = 3.4 \times 10^{-23}$). The ability for m^2_{450k} to predict the number of DMPs identified was assessed using receiver operating characteristic (ROC) curves. At low P value thresholds ($P < 1 \times 10^{-7}$) m^2_{450k} was no better than chance, but at higher P value thresholds (max threshold: $P < 1 \times 10^{-3}$) the sensitivity and specificity increased dramatically. We then measured the proportion of total variance related (r^2) to DMPs at varying P value thresholds (1×10^{-4} – 1×10^{-7}) amongst 11 traits in an independent cohort ($n = 334$). As P value thresholds increased (more DMPs identified) the r^2 increased. Both these findings provide evidence for a model whereby differential methylation at many CpG sites across the genome each contribute towards a small fraction of the overall association between DNA methylation and complex traits. This study explores the extent to which assayed DNA methylation variation could predict complex traits. This varied largely among traits - just among 125 metabolites measured the interquartile range = 16.3%. Unsurprisingly, it was found the greater the extent of prediction (higher m^2_{450k}) the greater the likelihood of identifying DMPs in EWAS. This suggests that the relevance of studying differential methylation changes measured in blood varies largely, depending on the trait. It was found many DMPs contribute to the total m^2_{450k} , suggesting that for complex traits each CpG may be relevant and associate minimally with a phenotype. Thus, as for genome-wide association studies (GWAS), EWAS will have to increase sample sizes in hopes of discovering DMPs or alternatively produce accurate methods to group correlated sites.

1960T

Up-regulation of SMN transcript is associated with a disease susceptibility in Weaver syndrome? K. Hosoki¹, E. Imagawa², N. Matsumoto², N. Okamoto¹. 1) Department of Molecular Medicine, Osaka Women's and Children's Hospital Research Institute, Osaka, Japan; 2) Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan.

Weaver syndrome (WS) is an overgrowth syndrome characterized by intellectual disability, limb anomalies, and abnormal muscle tone. WS is caused by a heterozygous mutation of *EZH2* gene, however, the pathogenesis in WS is still unclear. The *EZH2* is a member of polycomb repressive protein 2 (PRC2) associated with the regulation of gene transcription because of the function to mediate histone 3 lysine 27 trimethylation (H3K27me3). Recently, it is suggested that one of the transcriptional regulations with PRC2 is associated with long non-coding RNAs (lncRNAs), and there are a few reports that the transcription of disease-related genes is regulated by lncRNA-mediated recruitment of PRC2, however, this regulation mechanism is also unclear. Taken together, we herein evaluate the association between the gene transcription and PRC2 using *EZH2* mutant cells derived from a WS patient. In this study, we focused on lncRNA-related SMN transcript regulation as previously suggested. We used EBV-transformed lymphoblasts isolated from a WS patient with a heterozygous *EZH2* mutation with a decrease of H3K27me3 [Imagawa *et al.* Hum Mutat 2017]. We confirmed an increased expression level of *SMN1/SMN2* by RT-qPCR. We also detected the *SMN1/SMN2* overexpression when we performed *EZH2* knockdown using antisense oligonucleotides in HEK293 cells. Our results suggest that SMN transcriptional regulation is associated with PRC2 and/or H3K27me3, and may influence the pathogenesis in WS patients, considering that SMN deficiency leads to spinal muscular atrophy which is a severe neuromuscular disorder. In previous studies, the *SMN1* duplications were detected in amyotrophic lateral sclerosis (ALS), which is a progressive neurodegenerative disorder, and it is suggested that *SMN1* duplication is related the susceptibility of neuromuscular phenotype in ALS. The *SMN1* duplication in ALS is expected the increased expression level of *SMN1/SMN2*. Therefore, our results suggest that the up-regulation of SMN transcript is associated with the susceptibility of abnormal muscle tone in WS. We would like to examine detailed transcriptional regulation to understand the WS pathogenesis.

1961F

Growth disrupting mutations in epigenetic regulatory genes are associated with abnormalities of epigenetic aging. A.R. Jeffries¹, R. Maroofian², C.G. Salter^{1,3,4}, B.A. Chioza¹, H.E. Cross⁵, M.A. Patton¹, I. Karen Temple^{3,4}, F.I. Rezwan³, K. Tatton-Brown^{2,6,7}, J. Mill¹, A.H. Crosby¹, E. Baple^{1,8}. 1) University of Exeter Medical School, RILD Wellcome Wolfson Centre, Royal Devon & Exeter NHS Foundation Trust, Barrack Road, Exeter, EX2 5DW, UK; 2) Genetics Research Centre, Molecular and Clinical Sciences Institute, St George's, University of London, Cranmer Terrace, London SW17 0RE, UK; 3) Human Genetics and Genomic Medicine, Faculty of Medicine, University of Southampton, Southampton, SO16 6YD, UK; 4) Wessex Clinical Genetics Service, Princess Anne Hospital, Coxford Road, Southampton, SO16 5YA, UK; 5) Department of Ophthalmology and Vision Science, University of Arizona School of Medicine, 655 N. Alvernon Way, Tucson, Arizona, USA; 6) Division of Genetics and Epidemiology, Institute of Cancer Research, 15 Cotswold Road, London SM2 5NG, UK; 7) South West Thames Regional Genetics Service, St George's University Hospitals NHS Foundation Trust, London SW17 0QT, UK; 8) Peninsula Clinical Genetics Service, Royal Devon & Exeter Hospital, Gladstone Road, Exeter, EX1 2ED, UK.

Germline mutations in epigenetic regulatory genes such as *DNMT3A* are commonly associated with overgrowth disorders, whereas somatic mutations are typically associated with malignancy. In order to elucidate the molecular effects of germline mutations in *DNMT3A*, we profiled genome-wide patterns of DNA methylation in *DNMT3A* p.(Arg771Gln) mutation carriers in a large sibship with Tatton-Brown-Rahman syndrome (TBRS), together with their mosaic father. We also profiled 15 other previously documented TBRS patients with distinct pathogenic *de novo* *DNMT3A* mutations. All TBRS patients displayed widespread reduction of DNA methylation at discrete sites across the genome, with significant overrepresentation at sites annotated to molecular pathways related to morphogenesis, development and differentiation, and pathways implicated in a predisposition to malignancy. Additionally, we performed a biological age measure using the DNA methylation age calculator. Individuals harbouring the *DNMT3A* p.(Arg771Gln) mutation, together with all other TBRS patients profiled, displayed significantly accelerated DNA methylation aging. Further to this, our profiling studies of another overgrowth disorder affecting the histone methyltransferase NSD1 (Sotos Syndrome) revealed similar accelerated epigenetic aging. Notably our studies of Kabuki syndrome, a disorder associated with poor growth resulting from pathogenic mutations affecting the KMT2D histone methyltransferase gene, revealed lower epigenetic age compared to their chronological age. Together, our findings provide important new insights into the role of aberrant epigenetic mechanisms and altered epigenetic aging in growth disorders such as TBRS, Sotos and Kabuki syndromes.

1962W

Finding the missing heritability of complex disease: Optimization of the Assay for Transposase-Accessible Chromatin using Sequencing (ATAC-seq) for genome-wide examination studies. L.R. Main^{1,2,3}, K.L. Miskimen^{2,3}, E.R. Chan², H. Bai^{3,4}, W.S. Bush^{2,3,4}, J.L. Haines^{1,2,3}. 1) Department of Genetics and Genome Sciences, Case Western Reserve University School of Medicine, Cleveland, Ohio 44106, USA; 2) Department of Population and Quantitative Health Sciences, Institute for Computational Biology, Case Western Reserve University School of Medicine, Cleveland, Ohio 44106, USA; 3) Institute for Computational Biology, Case Western Reserve University, Cleveland, Ohio 44106, USA; 4) Institute for Systems Biology and Bioinformatics, Case Western Reserve University, Cleveland, Ohio 44106, USA.

Although thousands of loci have been implicated in the inheritance of complex traits, in most cases a significant portion of the calculated heritability remains unexplained. One possible explanation for this missing heritability is the inheritance of epigenetic states. Studies in multiple model systems demonstrate that methylation states and histone modifications can be passed down to progeny, but these phenomena have been difficult to observe reproducibly in humans. Two significant hurdles in examining the phenomena are identifying an appropriate study population and applying a reproducible genome-wide epigenetic assay. To address the first hurdle, we can focus on an isolated founder population such as the Amish. The Amish present an ideal opportunity to study heritability because they have a relatively homogeneous environment, large and available multi-generational families, and their endogamy leads to reduced genetic variation. To address the second hurdle, we can examine chromatin accessibility using ATAC-Seq. Blood samples were obtained from an Amish nuclear family and were examined using ATAC-Seq to look at open chromatin structure. Because ATAC-Seq has been largely used to focus on specific segments of DNA, we first tested for genome-wide reproducibility. A large amount of variation was observed within this initial study, as low as 3.8% of called peaks were overlapping in technical replicates. Two more sets of experiments were then conducted to better understand the sources of the variability. The first experiment utilized multiple technical replicates of three unrelated individuals and another was performed on a synchronized cell line. We observed a high level of variability in all the ATAC-Seq experiments. Publicly available data showed similar variability to our results, with peak overlap ranging from 0.48-10.1%. This large range of variability demonstrates that ATAC-Seq is not highly reproducible with current bioinformatics techniques. Currently we are focused on optimizing the peak-calling step in the data analysis pipeline to improve the signal-to-noise ratio. Developing and refining a complete QC and analysis pipeline specifically for ATAC-Seq genome-wide analysis is important since there are very few standards on QC and downstream analysis of this method. Fully understanding the limitations of the ATAC-seq assay are critical as we begin to design studies that utilize ATAC-Seq data to examine genome-wide epigenetic inheritance in humans.

1963T

Site-directed manipulation of DNA methylation in cells *In Vivo* and the effect on alternative splicing. R. Shayevitch, D. Askayo, I. Keydar, G. Ast. Tel-Aviv University, Tel-Aviv, Israel.

Alternative splicing contributes to proteome diversity. As splicing occurs co-transcriptionally, epigenetic determinants such as DNA methylation likely play a part in regulation of alternative splicing. Previously, we have shown that DNA methylation marks exons and that a loss of DNA methylation alters splicing patterns in a genome-wide manner. To interrogate the influence of DNA methylation on splicing of individual genes, we developed a method to manipulate DNA methylation *in vivo* in a site-specific manner using the deactivated endonuclease Cas9 fused to enzymes that methylate or demethylate DNA. We used this system to directly change the DNA methylation pattern of selected exons and introns. We demonstrated that changes in the methylation pattern of alternatively spliced exons, but not constitutively spliced exons or introns, altered inclusion levels. This is the first direct demonstration that DNA methylation of exon-encoding regions is directly involved in regulation of alternative splicing.

1964F

DNA methylome and whole transcriptome analyses identify genes potentially contributing to phenotypic differences in a set of monozygotic twins with Gaucher disease. *N. Tayebi¹, S. Hassan¹, F. Donovan², A. Elkahouloun², M. Langeveld³, C. Hollak³, E. Sidransky¹.* 1) NIH/ NHGRI/MGB, Bethesda, MD; 2) Cancer Genetics and Comparative Genomics core, NHGRI/ NIH, Bethesda, MD; 3) Academic Medical Center, Amsterdam, The Netherlands.

Gaucher disease (GD) is an autosomal recessive disorder resulting from mutations in *GBA1*, encoding glucocerebrosidase, which leads to lysosomal accumulation of glucocerebroside. Vast phenotypic heterogeneity is observed among patients, even between monozygotic twins. Epigenetic marks such as DNA methylation may contribute to this variation. We previously described twin sisters with GD, confirmed to be monozygotic by DNA fingerprinting, who had divergent phenotypes. Mutational analysis revealed homozygosity for *GBA1* mutation N188S, c.670C>T. One sister exhibited splenomegaly, anemia and thrombocytopenia, severe bone pain, altered saccadic eye movements and complex seizures, while the second had no GD manifestations. We investigated whether epigenetic differences might underlie this phenotypic discrepancy, looking at methylation sites in the genomic DNA from each twin. Bisulfite- modification of DNA was used to discriminate between cytosine and methylated cytosine in DNA from peripheral blood. Treated DNAs were run on an Illumina HumanMethylation450 BeadChip. Data were analyzed using GenomeStudio. In the more affected twin, we identified five methylated and three demethylated genes at CpG islands (P-value= 0.01-0.001), and two differentially methylated genes in the enhancer region (P-value= 0.05-0.01). One demethylated gene, *CACNA1A*, encoding a neuronal alpha-1a subunit of a voltage dependent calcium channel, and one methylated gene, *MEST*, a member of the alpha/beta hydrolase superfamily, were validated by real time PCR using RNA extracted from both leukocytes and fibroblasts. Whole exome sequencing, performed to exclude somatic mosaicism, revealed no significant differences. A ClariomTM S transcriptome array, performed on RNA samples from the twins and three aged-matched female controls, confirmed the up- or down- regulation of genes identified on the methylation array. Differential expression of many other genes was seen, some related to *MEST* and *CACNA1A*. Both GeneMANIA pathway and Protein-Protein network (IMEX protein data base) analyses confirmed potential interactions between the proteins encoded by *GBA1*, *MEST*, *CACNA1A*, either directly or through intermediate proteins. Thus, epigenetic analyses can be used to help identify genes that influence disease manifestations and contribute to the complexity of genotype-phenotype relationships in Gaucher disease.

1965W

1.1% of the global DNA is methylated in human amniotic fluid stem cells.

P. Upadhyaya¹, I. Antonucci^{1,2}, L. Stuppia^{1,2}. 1) Department of Psychological, Health and Territorial Sciences, Laboratory of Molecular Genetics, School of Medicine and Health Sciences; University of Chieti-Pescara, Chieti, Abruzzo, Italy; 2) Stem Tech Group, Center for Aging Studies (CESI); University of Chieti-Pescara, Chieti, Abruzzo, Italy.

Every phenotype of cell has their characteristic properties. These properties allow the cells to carry out their specific functions. The different lineages of the cells functionally do not match with each other. This difference is due to the different in gene expressions among the cells. It has been cleared till now that epigenetics largely controls the phenomenon by which control over the expression of genes takes place, without modifying the sequence of the DNA. There are several epigenetic mechanisms that come into play to determine the fate and behaviour of a cell. These mechanisms include DNA methylation, histones modification and regulation by micro RNA. In our research, we are focussed on quantification of methylated DNA (5-methylcytosine; 5-mC) within the total DNA of human amniotic fluid stem cells (hAFSCs), which contribute around 1% of the cell population in amniotic fluid and show multipotent characteristics with the ability to differentiate into cells of all three germ layers. The global DNA methylation quantification was performed on the nuclear DNA extracted from hAFSCs using MethylFlash Methylated DNA Quantification kit-Colorimetric (Epigenetek; USA) by following manufacturer's protocol. The kit uses optimised antibody and enhancer solutions to allow high specificity to 5-mC, with no cross-reactivity to unmethylated cytosine and no or negligible cross-reactivity to hydroxymethylcytosine. Our results found consistent 5-mC in all the five hAFSC lines. The mean 5-mC was found to be 1.11 ± 0.03 %. Statical analysis showed the data to be significant (two-tailed P value= < 0.0001, coefficient of variation = 2.83%). Interestingly, our results also suggest that difference in passage number does not change the amount of 5-mC, which implies that the methylation remains almost equal among the cells with the different number of cell divisions. In conclusion, global DNA methylation quantification may be used as the marker to identify the lineage and functional ability of any cell type. More research is necessary to quantify the global methylation of the DNA in all the human normal and cancerous cell types to identify epigenetic differences in normal and pathological cell types.

1966T

Incorporate measurement errors in epigenome-wide association studies.

W. Guan¹, Y. Bai¹, J.S. Pankow¹, E.W. Demerath¹, J. Bressler², M.L. Grove², M. Fornage². 1) University of Minnesota, Minneapolis, MN; 2) The University of Texas Health Science Center at Houston, Houston, TX 77030.

Genome-wide DNA methylation measures are now routinely used to investigate their association with various outcomes of interest. Inherent to the array-based DNA methylation measures, e.g., Illumina HumanMethylation450 (HM450) or Infinium MethylationEPIC chips, is the associated technical variation of non-biological interest. One example of such variation is the bead-to-bead variation due to the use of multiple beads on HM450 or EPIC arrays. These technical variations can be viewed as measurement errors, and are commonly ignored in downstream association analysis. We have proposed a novel statistical framework to take into account of these measurement errors. Specifically, we used a mixed effects model to quantify the measurement error, and developed an expectation-maximization (EM) algorithm to estimate the model parameters. We applied our proposed method to the Atherosclerosis Risk in Communities (ARIC) methylation data ($n = 2,843$; HM450 array) in an epigenome-wide association study of smoking status, after accounting for the bead-to-bead variation. We identified 14 additional CpG sites associated with smoking, at sites with high technical variation (intraclass correlation coefficient < 0.4). We expect that our new method can improve statistical power of association tests and accuracy of parameter estimates in future epigenome-wide association studies (EWASs) using methylation arrays.

1967F

Estimating cell type abundance in GTEx enables insights into cellular mechanisms and origins of eQTLs.

F. Aguet¹, S. Kim-Hellmuth², M. Oliva³, M. Muñoz-Aguirre⁴, J. Quan⁵, V. Wucher⁵, H.S. Xi⁵, B.E. Stranger³, T. Lappalainen³, R. Guigó⁴, K.G. Ardlie¹, GTEx Consortium. 1) Broad Institute, Cambridge, MA; 2) New York Genome Center, NY, NY; 3) U Chicago, Chicago, IL; 4) CRG, Barcelona, Spain; 5) Pfizer Inc, Cambridge, MA.

The Genotype-Tissue Expression (GTEx) project has identified expression quantitative trait loci (cis-eQTLs) for the majority of genes across a range of human tissues. However, the interpretation of these eQTLs has been limited by the complex cell type composition of GTEx tissue samples. Here, we applied gene expression-based deconvolution approaches to estimate the cellular composition of GTEx tissues, based on gene expression profiles of 17,382 RNA-seq samples from 838 individuals across 49 tissues (v8 data release). We found that highly abundant cell types were robustly estimated across multiple methods and were consistent with cell types expected to be present in different tissues. Cell type composition explained a large part of inter-individual variation in gene expression, and provided insights into cellular events associated to pathological changes in GTEx individuals. To demonstrate that computationally inferred cell type abundances can be used as proxies to identify cell-specific eQTLs, we used neutrophil estimates from GTEx whole blood RNA-seq data (670 individuals) to identify cell type-interacting eQTLs (i-eQTLs). We found 1258 genome-wide significant (FDR 5%) i-eQTLs interacting with neutrophil abundance; these showed good replication of eQTLs identified in purified neutrophils. Notably, 14% of neutrophil i-eQTLs had a nominal whole blood cis-eQTL p-value > 0.05 , implying that a significant fraction of cell type-specific eQTLs is undetectable when studying heterogeneous bulk tissue expression. Finally, colocalization analysis of neutrophil i-eQTLs with multiple immune-related diseases not only pinpointed the cellular origin of known whole blood eQTLs, but identified novel colocalized loci that were masked in bulk tissue. Taken together, our results emphasize the importance of identifying genetic effects on gene expression at the cellular level, and demonstrate that cell type-specific effects can be uncovered in bulk tissues by integrating increasingly available single-cell and cell type-specific transcriptomic data. The GTEx tissues provide a unique opportunity to expand this approach to immune and stromal cell types in tissues where cell-specific eQTLs have not yet been characterized.

1968W

Validation of causal variants underlying eQTLs in LCLs and iPSCs by CRISPR/Cas9 genome editing. *KA. Barr¹, Y. Gilad².* 1) Section of Genetic Medicine, University of Chicago, Chicago, IL, USA; 2) Department of Human Genetics, University of Chicago, Chicago, IL, USA.

Expression quantitative trait loci (eQTLs) connect natural variation to functional changes in gene expression and provide an essential link between genotype and disease risk. To date, hundreds of thousands of eQTLs have been identified in diverse human tissues. In the majority of cases, linkage disequilibrium precludes the identification of the true causal variant underlying these eQTLs. While fine mapping, hierarchical models and reporter assays have been used to prioritize variants, these methods cannot prove causality. It is now possible to directly assess the consequence of individual variants using precise genome editing. Currently, only a handful of causal variants have been confirmed with these methods. To address this need, we are directly testing the effects of several genetic variants underlying tissue-specific eQTLs. These eQTLs are active in either lymphoblastoid cell lines or induced pluripotent stem cells. Each variant is likely to be causal according to a Bayesian hierarchical model that incorporates chromatin state and transcription factor footprints. Using CRISPR/Cas9 and template-directed genome repair, we are generating cell lines that differ at only the variants in question. We will then assess the consequences of these variants on gene expression using qPCR. These results will provide validation and serve as a benchmark for past and future eQTL studies. In addition, we expect that the identification of causal variants will open the door to future inquiries on molecular mechanisms that underlie differences in gene expression between individuals and tissues types.

1969T

Validation of cis-regulatory transcript variants using fine-mapping and CRISPR/Cas9 genome editing. *M. Brandt^{1,2}, F. Hormozdizadeh^{3,4}, A. Vasileva^{1,2}, T. Lappalainen^{1,2}.* 1) Department of Systems Biology, Columbia University, New York, NY; 2) New York Genome Center, New York, NY; 3) Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA; 4) Program in Medical and Population Genetics, Broad Institute of Massachusetts Institute of Technology (MIT) and Harvard, Cambridge, MA.

Expression quantitative trait loci (eQTL) studies in humans have identified thousands of common genetic variants associated with gene expression levels. However, due to linkage disequilibrium, the true causal variants and therefore the molecular mechanisms remain mostly unknown. Furthermore, functional effects of rare variants are often challenging to detect in population studies. In this study, we have developed a scalable CRISPR-based assay for experimental assessment of regulatory effects of genetic variants in transcribed regions. We have applied this method to rare protein-truncating variants and eQTLs from the GTEx project. First, we utilized CRISPR/Cas9 to introduce nine rare stop-gained SNPs from GTEx into human cells. These variants are expected to trigger nonsense mediated decay, resulting in a reduction in transcripts which contain the alternative variant. We transfected Cas9-inducible 293T cells with gRNAs and single stranded homologous templates specific to each variant. After editing, we extracted mRNA and gDNA from the pool of edited cells and performed targeted sequencing. A difference in the representation of the edited variant in gDNA versus mRNA confirms its effect on transcript abundance. Compared with matched controls, the stop-gained variants show a significant allele-specific reduction in transcript abundance ($p = 7.8e-3$). This demonstrates our ability to detect the effects of causal regulatory variants on gene expression, which can be applied in the future to testing whether rare variants of interest trigger nonsense-mediated decay. Next, we applied the same method and principles to validate common eQTLs found in the general population. Starting with 15,219 fibroblast eGenes from 483 individuals in the GTEx v8 data, we selected 189 eGenes for experimental validation of the causal variant. We focused on genes with good statistical fine-mapping, demonstrated as a CAVIAR fine-mapping posterior probability of greater than 0.8 for at least one SNP inside the transcript. Additionally, genes were selected for relatively high effect size and gene expression in HEK293 cells. The putative causal variants from these eGenes are being experimentally tested, followed by analysis of statistical fine-mapping accuracy and regulatory annotation overlap of causal variants. Altogether, our method allows scalable experimental confirmation of putative causal regulatory variants, and improved interpretation of regulatory variation in humans.

1970F

Human cytomegalovirus UL23 protein regulates transcription of interferon- γ -stimulated genes. H. Li¹, L. Feng¹, C. Foo¹, S. Wu¹, Y. Ran¹, X. Yang¹, T. Zhou¹, Z. Deng^{1,2}. 1) Dept. of Biotechnology, Jinan University, Guangzhou, Guangdong, China; 2) Keck School of Medicine, University of Southern California, Los Angeles, CA.

Human cytomegalovirus (CMV), a member of the human herpesvirus family, is a common opportunistic virus causing severe ailments and deaths in people with immature or compromised immune systems. HCMV expresses viral proteins to modulate the host immune responses at every step of its life cycle, which play a crucial role in viral pathogenesis. Interferons (IFNs) are part of the innate immune response to viral infections and confer potent antiviral effects. These cytokines also have important roles in immune-surveillance for malignant cells. Interferon- γ (IFN- γ) responses are vital for a host to combat infections of many human viruses including human herpesviruses. Upon treatment of IFN- γ , transcription of many genes responsible for IFN- γ immune responses is activated primarily by the signal transducer and activator of transcription (STAT) proteins such as STAT1 protein. UL23 is a virion protein found in the tegument and is expressed in the cytoplasm in HCMV-infected cells. However, UL23 is dispensable for viral replication in cultured cells and little is currently known about its function. In this study, we observed (a) UL23 significantly reduced levels of STAT1 in the nuclei, the sites where these proteins act to induce transcription of IFN- γ stimulated genes, and (b) decreased levels of the induction of the transcription of IFN- γ stimulated genes. UL23-deficient HCMV mutants induced higher transcription of IFN- γ stimulated genes and exhibited lower titers than parental and control revertant viruses expressing functional UL23 in IFN- γ treated cells. Our results highlight the roles of UL23 in facilitating viral immune escape from IFN- γ responses and enhancing viral resistance to IFN antiviral effects.

1971W

Novel regulators of *CFTR* gene expression in airway epithelial cells. M. NandyMazumdar, S. Yin, M. Mutolo, S-H. Leir, A. Harris. Department of Genetics and Genome Sciences, Case Western Reserve University School of Medicine, Cleveland, OH.

The cystic fibrosis transmembrane regulator (*CFTR*) gene is subject to complex regulation by multiple mechanisms including the recruitment of both activating and repressive transcription factors (TFs), which coordinate cell-type-specific chromatin looping at the locus. Within a topologically associated domain (TAD), multiple *cis*-elements regulate *CFTR* gene expression by looping to the promoter and/or to sites of CTCF occupancy at the TAD boundaries. Though lung disease is key feature of Cystic Fibrosis (CF), *CFTR* expression in the airway is substantially lower than in the pancreas or the intestine. Moreover, although the TAD boundaries are conserved among diverse cell-types; different cell lineages utilize specific regulatory elements for optimal gene expression. Our previous work established several airway-specific DNase hypersensitive sites (DHS) which encompass enhancers. Of these, 2 major enhancers identified at the -35 and -44 kb DHS with respect to the promoter, may serve important regulatory roles; thereby maintaining the required cell-type selective *CFTR* expression profiles. In addition to cell-specificity; these enhancers demonstrated cooperativity and binding of several important TFs including Interferon regulatory factors 1 and 2 (IRF1, IRF2) and nuclear factor (erythroid-derived 2)-like 2 (Nrf2) at the -35 and -44 kb sites respectively. In order to determine their function, we used CRISPR/Cas9 technology to delete these sites in the 16HBE14o human bronchial epithelial cell line. Here we successfully excised these 2 airway-specific enhancers from multiple cell clones. Interestingly, the enhancer-deleted clones demonstrated almost complete loss of *CFTR* expression as measured by RT-qPCR assays, suggesting that these *cis*-regulatory elements may be critical in airway *CFTR* transcription. Moreover, we are examining the impact of loss of each element on higher order locus architecture by 4C-seq. Concurrently new candidate transcription factors that activate *CFTR* expression in airway epithelial cells were identified in an siRNA screen targeting 1500 TFs and chromatin modifying factors. Current experiments are combining ChIP-seq with RNA-seq following TF depletion for candidate factors, to determine direct and indirect mechanisms of *CFTR* regulation. These data will also enhance our understanding of the transcriptional networks in human airway epithelial cells, which are pivotal to several common lung diseases. .

1972T

Regulatory role of chromosome 11p13 in cystic fibrosis lung disease severity. H. Swahn, K-M. Lamar, J. Sabith Ebron, S. Yin, M. NandyMazumdar, S-H. Leir, A. Harris. Department of Genetics and Genome Sciences, Case Western Reserve University School of Medicine, Cleveland, OH.

Mutations in the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene cannot fully explain lung disease phenotype in cystic fibrosis (CF). Previous genome-wide association studies identified genomic regions associating with CF lung disease severity; among these the highest p-value single nucleotide polymorphisms mapped to an intergenic region at chromosome 11p13. This interval lies between the E74-like factor 5 (*ELF5*) and Ets-homologous factor (*EHF*) genes, and Apaf-1 interacting protein (*APIP*) gene. *ELF5* is thought to confer asthma susceptibility in certain populations, and *EHF* regulates response to injury and inflammation in the lung. *APIP* is a methionine salvage enzyme involved in apoptosis. Our goal is to determine the functional significance of this region, which is predicted to be regulatory. Our previous work established that the 11p13 modifier region contains specific sites of open and active chromatin and encompasses airway selective *cis*-regulatory elements. Two of these elements, DHS11.2516 and 11.2521, encompass enhancers of the *ELF5* and *EHF* promoters, but not *APIP*. Furthermore, we showed that 11p13 is contained within a topologically associated domain (TAD), and the *APIP* promoter lies outside of this TAD. Our results focus attention on the likely role of *ELF5* and/or *EHF* in modifying CF lung disease severity. We hypothesize that these epithelial selective genes are regulated by *cis*-elements found at 11p13, and that dysregulation of these genes contributes to CF lung disease severity. Here we used CRISPR/Cas9 technology to remove the two enhancer elements from the endogenous region in airway epithelial cells. We found that deletion of 11.2516 increased expression of several genes at 11p13 and altered locus architecture, but deletion of 11.2521 had no apparent impact. Moreover, we are using nuclease dead Cas9 (dCas9) fused either to VP64, p65 and Rta, or to KRAB and SID, to activate or repress, respectively, promoters and *cis*-regulatory elements across 11p13 in human airway epithelial cells. These experiments will connect critical *cis*-regulatory elements with the individual genes at 11p13 and may reveal the molecular mechanism underlying the role of this region in modifying lung disease severity. Key processes such as epithelial cell barrier function, migration, proliferation and apoptosis will be assayed. These data will show how non-coding variants at 11p13 impact CF lung pathology.

1973F

Epigenetic marks at major histocompatibility complex affect male fertility. S. Sarkar, S. Rajender. Endocrinology Division, CSIR-Central Drug Research Institute, Sector 10, Jankipuram Extension, Sitapur Road, Lucknow, Uttar Pradesh 226031.

Human leucocyte antigens are membranous glycoproteins playing salient role in immune recognition and response. The expression pattern of HLA class antigens is dubious in nature. However, quite a number of studies have reported the expression of HLA class molecules. Trophoblastic *HLA-C* gene variants have been shown to be associated with preeclampsia in African and European women. Maternal *HLA-C* genetic variants are also seen to have effect on high fetal and maternal mortality and morbidity. HLA class 1 molecules were found to be expressed in spermatozoa of infertile patients but showed no significant difference against fertile groups. On the other hand, it was found that mRNA levels of *HLA-G* in normal spermatocytes were significantly higher than the testicular samples with only Sertoli cells and/or spermatogonia, which suggests its involvement in spermatogenesis. A study observed that the predisposition to antisperm antigen can increase the levels of some HLA class1 antigen in vasectomized men. A whole genome epigenetic study on adolescent chemotherapy patients causing infertility also showed differential methylation pattern in *HLA-C* gene. We investigated differential methylation regions present in different classes of *HLA* in patients with different sperm count in order to correlate epigenetic patterns with sperm production ability. To achieve this objective, our experimental method included isolation of sperm DNA from individuals of various sperm count groups followed by Infinium 450K DNA methylation array. Among these groups we have identified six different classes of *HLA* genes to be differentially methylated. *HLA* class I genes showed higher level of methylation in oligospermic group as compared to normospermics, suggesting the potential role of these genes in spermatogenesis. Similar methylation pattern was also seen in DNA from PBMCs. Out of all immune response gene only *HLA-DRB1* showed disparate methylation pattern in oligospermics (bvalue=0.345) than normal men (bvalue=0.679). This DNA methylation pattern suggests an important role of these immune responsive genes in sperm production. These observations opened a valuable insight into the cause of infertility which directs the focus to select best sperm for fertilization success. However, more epigenetically focused experimental designs are warranted to draw conclusive insights and theories about the role of immune response genes and reproductive health.

1974W

Epigenetics of cyanogenic toxicity of dietary origin: The model of konzo.

K. Kocher^{1,2}, M. Bramble^{1,2}, N. Vashist^{1,2}, S. Bhattacharya^{1,2}, K. Panchapakesan^{1,2}, S. Knoblach^{1,2}, D. Tshala-Katumbay^{3,4}, E. Vilain^{1,2}. 1) The George Washington University, Washington, DC; 2) Children's National Research Institute, Washington, DC; 3) Oregon Health & Science University, Portland, OR; 4) University of Kinshasa, Kinshasa, DRC.

The consumption of poorly processed cyanogenic bitter cassava, a vital staple in the diet of over 600 million people around the globe, is associated with konzo, an irreversible, non-progressive motor neuron disease. Dietary cyanide toxicity results in a considerable burden on Public Health in sub-Saharan Africa, affecting up to 5% of the population in certain regions, with numerous individuals affected with chronic motor and cognitive disabilities. Yet the underlying biological mechanisms of this toxicity remain poorly understood. While certain risk factors such as malnutrition and young age are associated with konzo onset, there has been no comprehensive mechanistic approaches to understanding the etiology or susceptibility factors associated with konzo. Not all individuals who consume cyanogenic cassava develop konzo, suggesting that certain epigenetic factors may be associated with disease onset. We hypothesize that the exposure to cyanogenic factors present in the cassava result in durable epigenetic changes affecting long term neurologic development. To test this hypothesis, we investigated the global DNA methylation differences associated with konzo-affected youth in comparison to age-matched healthy controls from the same outbreak zones of the Democratic Republic of Congo. We noted that there were approximately 60,000 differentially methylated regions among the genomes of konzo-affected youth and healthy controls, many of which are enriched for biological processes of interest in the context of konzo etiology, including metabolic and neurological pathways that may be influenced by cyanogenic cassava consumption. Additional investigation of transcriptomic level changes that are involved with the susceptibility to and pathobiology of konzo through dietary cyanogen exposure is underway. Understanding the genomic and epigenomic landscape of konzo will provide invaluable insights into dietary cyanogen metabolism, and will have a high impact from a global health perspective, as the metabolism and associated neurodegenerative and cognitive effects of dietary cyanogen could affect millions of individuals around the world.

1975T

Integrative genomic analysis of microRNA transcriptional responses to malaria infection.

A. Diawara¹, M.M. Dieng¹, V. Manikandan¹, H. Tamin El Jarkass¹, R. Eteer¹, A.B. Tiono², S.B. Sirima², I. Soulama², Y. Idaghdour¹. 1) New York University, Abu Dhabi, United Arab Emirates; 2) Centre National de Recherche et Formation sur le Paludisme, Ouagadougou, Burkina Faso.

Joint analysis of clinical phenotypes, genotypic and gene expression data emerged as a promising approach to map complex diseases. We used this approach to address the sources of variability of global miRNA gene expression in 121 whole blood samples collected from West African children sampled before and during *Plasmodium falciparum* infection and after treatment. Genome-wide expression profiling revealed 95 miRNAs responsive to infection and 72 miRNAs associated with parasite load. Cross-correlation of infection-responsive miRNAs with host total RNA from the same individuals pinpoints miRNAs and target mRNAs involved in key immune response pathways. Furthermore, using whole-genome sequencing data we have conducted a genome scan for miRNA eQTLs and identified genetic variants associated with infection-responsive miRNAs and discovered several cases of genotype-by-infection interactions. The novel insight gained from this study elucidates the roles of miRNA and regulatory variation as mediators of host response to malaria infection and highlights the power of systems genetics approaches in infectious disease. .

1976F

Regulatory network of transcription factors determining CYP3A4 expression: Identified with mathematical modeling and CRISPR-mediated genome editing. D. Wang^{1,4}, R. Lu², G. Rempala³, W. Sadee⁴. 1) Department of Pharmacotherapy and Translational Research, College of Pharmacy, University of Florida, PO Box 100486, 1345 Center Drive MSB PG-05B, Gainesville, FL32610; 2) Department of Clinical Sciences, Bioinformatics Core Facility, University of Texas Southwestern Medical Center, Dallas, Texas 75235; 3) Mathematical Bioscience Institute, The Ohio State University, Columbus, OH43210; 4) Center for Pharmacogenomics, Department of Cancer Biology and Genetics, The Ohio State University, Columbus, OH43210.

CYP3A4 is the predominant drug metabolizing enzyme in the liver. CYP3A4 expression and enzyme activity displays large inter-person variability with unknown causes, substantially affecting individual drug exposure and therapeutic outcomes. The expression of CYP3A4 is controlled by several liver enriched transcription factors (TFs), but how these TFs interact with each other to regulate CYP3A4 expression remains unknown. We hypothesize that several key TFs are critical to CYP3A4 expression through a regulatory network with nonlinear, dynamic, and hierarchical interactions. To resolve this complex interactive system, we use existing liver gene expression microarray data (GSE9588) and a novel mathematical modeling approach, global sensitivity analysis or Sobol indices (SI). This approach successfully converges on a network of TFs critical to CYP3A4 expression and identified interacting TF pairs, triplets and quadruplets. Some of the key factors are known, while others are novel as CYP3A4 regulators. ESR1 emerges as the key TF with the strongest main effects and the most frequently included interacting partner. Consistent with this result, CRISPR-mediated gene knockout or siRNA/shRNA gene knockdown of *ESR1* decreased CYP3A4 expression, while CRISPR-mediated transcriptional activation of *ESR1* increased CYP3A4 expression. Moreover, ESR1 also regulate the expression of other TFs known to regulate CYP3A4 expression. These results validate ESR1 playing critical roles in CYP3A4 expression. As the role of ESR1 was detectable in both male and female livers and no estrogen was added, the results demonstrate a role for unliganded ESR1 in CYP3A4 expression, consistent with the known transcriptional role of unliganded ESR1 in other cells.

1977W

Characterization of the nuclear and cytosolic transcriptomes in human brain tissue reveals new insights into the subcellular distribution of RNA transcripts. L. Feuk^{1,2}, A. Zaghlool^{1,2}, A. Niazi^{1,2}, A.K. Björklund³, J. Orzechowski⁴, A. Ameur^{1,2}. 1) Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden; 2) Science for Life Laboratory in Uppsala, Uppsala University, Sweden; 3) Department of Cell and Molecular Biology, National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 4) Department of Biochemistry and Biophysics, National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Stockholm University, Stockholm, Sweden.

Transcriptome analysis has mainly relied on analyzing RNA sequencing data from whole cells, overlooking the impact of subcellular RNA localization and its influence on our understanding of gene function, and interpretation of gene expression patterns in cells. Here, we performed the first comprehensive analysis of cytosolic and nuclear transcriptomes in human fetal and adult brain samples. We show significant differences in RNA expression for protein-coding and lncRNA genes between cytosol and nucleus. Transcripts displaying differential subcellular localization belong to particular functional categories and display tissue-specific localization patterns. We also show that transcripts encoding the nuclear-encoded mitochondrial proteins are significantly enriched in the cytosol compared to the rest of protein-coding genes. Further investigation of the reliability of using the cytosol or the nucleus for differential gene expression analysis in single cell experiments as a proxy for whole transcriptome indicates important differences in results depending on the cellular compartment used for analysis. These differences were manifested at the level of transcript category and number of differentially expressed genes. Our data provides the first resource of RNA subcellular localization of RNA in human brain and highlight the influence of using the cytosol and the nucleus in interpreting and comparing results from single cell studies.

1978T

The accessible chromatin landscape of the hippocampus at single-cell resolution. K.A. Torkenczy¹, J. Sinnamoni², M. Linhoff³, S. Vitak⁴, H. Pliner⁵, C. Trapnell³, F. Steemer⁴, G. Mandel³, A.C. Adey^{1,5}. 1) Molecular and Medical Genetics, Oregon Health and Science University, Portland, OR; 2) Vollum Institute for Neuroscience, Portland, OR, USA; 3) University of Washington, Department of Genome Sciences, Seattle, WA, USA; 4) Illumina, Inc., San Diego, CA, USA; 5) Knight Cardiovascular Institute, Portland, OR, USA.

The hippocampus is a major component of the vertebrate brain. It is made up of anatomically and functionally distinct regions composed of multiple cell types that play a critical role in the formation and retrieval of episodic and spatial memory. However, the complex nature of this brain region poses a challenge in its accurate characterization. Cell type specific differences are lost due to the averaging of properties of cells comprising the tissue. Single-cell studies address this problem by identifying individual cells based on their relative occupied place in the landscape laid out by types and cell states within the assessed tissue. Recent single-cell transcriptional profiling in the hippocampus has demonstrated this by recapitulating cell types previously defined based on morphology. While RNA profiling provides valuable information on transcript abundance, it misses other regulatory factors that define the epigenetic state of cells, such as enhancers. There are up to an order more of such highly tissue and cell-type specific regulatory elements than transcripts in a cell. Single-cell Combinatorial Indexed Assay-for-Transposase-Accessible-Chromatin (sci-ATAC-seq) allows for the accurate description of the single-cell regulatory landscape through the mapping of promoter and enhancer accessibility and the identification of putative cell-state specific DNA binding transcription factor proteins. We applied sci-ATAC-seq to produce 2,346 high-quality single-cell chromatin accessibility maps with a mean unique read count per cell of 29,201 from both fresh and frozen hippocampi; observing little difference between the preparations. Using this dataset, we identified eight distinct major clusters of cells representing both neuronal and non-neuronal cell types and characterized the driving regulatory factors and differentially accessible loci that define each cluster. We then applied a recently described co-accessibility framework to identify co-regulated elements, which produced 487,156 links in a cell-type specific context, including 146,818 distal-promoter peak associations. Lastly, we carried out an additional sci-ATAC-seq preparation from cultured hippocampal neurons (899 high-quality cells, 43,532 mean unique reads) and revealed substantial alterations to their epigenetic landscape when compared with primary tissue.

1979F

WBSCR22: A genetic basis for lateralization of the magnocellular visual system? D.L. Mills¹, L. Dai², J.R. Korenberg². 1) School of Psychology, Bangor University, Bangor, Wales, United Kingdom, United Kingdom; 2) The Brain Institute Department of Pediatrics University of Utah 36 South Wasatch Drive, SMBB 4205 Salt Lake City, UT 84112-5001.

Investigating the role of genes in cognition and brain activity is of considerable interest across disciplines. However, correlational approaches can present a variety of methodological problems including false-positives. Here we propose a basic multi-level model for approaching genetic studies of human cognition. Research in Williams syndrome (WS), a neurodevelopmental disorder caused by a deletion of approximately 25 genes on chromosome 7, has identified a small set of candidate genes associated with specific social and cognitive functions. Recent findings suggest specific genes are linked with socio-cognitive brain activity as indexed by event-related potentials (ERPs) via cases with full and partial deletions in the WS region. Many neurodevelopmental disorders, including WS, are associated with selective vulnerabilities in magnocellular/dorsal stream visual processing. The role of specific genes in this process are examined in four lines of converging evidence using WS as a model. Study 1 examined differential neural activation of the magnocellular and parvocellular pathways as indexed by event-related potentials (ERPs) in a large sample of adults with the full deletion for WS, including FKBP6 through GTF2IRD1. The results suggested that individuals with WS have increased sensitivity of the left and decreases sensitivity to the right hemisphere to direction of motion, whereas activation of the parvocellular pathway did not differ from controls. Study 2 examined the role of specific genes, in ERP amplitudes to direction of motion in individuals with smaller WS deletions. The results suggested that sparing of genes from WBSCR22 and left towards FKBP6 were associated with typical ERP right hemisphere asymmetries to direction of motion. Study 3 correlated gene expression and methylation with ERP amplitudes linked to lateralization of magnocellular activity in individuals with the full deletion for WS using gene expression and methylation techniques. The results showed that expression and methylation of WBSCR22 were correlated with the right hemisphere ERP asymmetry to direction of motion. Study 4 examined expression of WBSCR22 in the visual cortex in the non-human primate brain. Taken together the findings showed converging evidence across levels that WBSCR22 is linked with lateralization of function for the magnocellular pathway. This is the first study to link specific genes with activation of separate pathways in visual stream processing.

1980W

Genetic effects on promoter usage are highly context-specific and contribute to complex traits.

K. Alasoo^{1,2}, J. Rodriguez², J. Danesh^{2,3,4,5}, D.F. Freitag^{2,3,4}, D.S. Paul^{2,3,4}, D.J. Gaffney². 1) Institute of Computer Science, University of Tartu, Tartu, Estonia; 2) Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA, UK; 3) BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK; 4) British Heart Foundation Centre of Excellence, Division of Cardiovascular Medicine, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK; 5) National Institute for Health Research Blood and Transplant Unit (NIHR BTRU) in Donor Health and Genomics, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK.

Genetic variants regulating RNA splicing and transcript usage have been implicated in both common and rare diseases. However, identifying specific transcriptional effects of these variants remains challenging, partly because reference transcriptomes are incomplete and contain many truncated transcripts that lack annotated 3' or 5' ends. We developed a novel analytical approach to overcome these limitations by stratifying transcript annotations into three separate components: promoters, internal exons and 3' ends. We apply our method to genotype and RNA-seq data from human macrophages exposed to a range of inflammatory stimuli (IFN γ , Salmonella, IFN γ + Salmonella) and a metabolic stimulus (acetylated LDL), obtained from up to 84 individuals. We found that over half of the quantitative trait loci (QTLs) colocalising with complex traits were identified only at the transcript level with no detectable effect on total gene expression. Furthermore, 55% of the transcript-level associations regulated either promoter or 3' end usage, many of which are missed by methods that only quantify exon-exon junctions. Finally, we demonstrate that promoter-usage QTLs have distinct genetic architecture and are 50% more likely to be context-specific than alternatively spliced internal exons. In summary, we highlight how different RNA-seq quantification approaches capture distinct aspects of transcription and that characterizing the full spectrum of transcriptional consequences of genetic variation requires a combination of analytical strategies.

1981T

Accurate prediction of chromatin conformation status using deep neural network model.

H. Uryu¹, K. Hata². 1) Medical Genome Center, National Center for Child Health and Development, Japan, Tokyo, Tokyo, Japan; 2) Department of Maternal, Fetal Biology, National Center for Child Health and Development, Japan, Tokyo, Tokyo, Japan.

Chromatin conformations in nuclei play an important role in regulating gene expression, and many reports have demonstrated that disruptions of these three-dimensional (3D) structures lead to disease status. Recent advances of high-throughput sequencing have enabled it to display 3D proximities among each genomic locus on the chromatin contact matrix using Hi-C analysis method. However, large-scale computational resources are indispensable in current protocols of Hi-C analysis for high-resolution contact map, and the repeatability of those contact matrix are still uncertain in the current practical procedures. In addition, it is difficult to understand the mechanism of chromatin conformation disruptions that lead to disease status. For understanding the conformation of chromatin 3D structures, we established deep neural network models to predict the existence of chromatin looping with the data set derived from a human lymphoblastoid cell line GM12878. For supervised learning of neural network model, we prepared the combination of genomic loci tagged as chromatin-loop-positive (N = 8526) using HiCCUPs protocol as previously reported (Rao et al 2014), and also prepared randomly selected genomic loci combinations other than positively-tagged loci as chromatin-loop-negative (N = 131365). The established model displayed its high accuracy (96.25%), sensitivity (99.35%), and specificity (95.22%). The AUC of the model was 0.997 against a test data set. These results suggest that our established deep neural network model is an effective method for analyzing chromatin 3D structures substituted for Hi-C analysis.

1982F

X chromosome aneuploidy alters the epigenetic landscape and gene expression genome-wide. G.N. Filippova¹, J.B. Berletch¹, W. Ma², G. Bonora², X. Chen⁴, H. Hrnčir³, D.L. VanDyke⁵, A.P. Arnold⁴, X. Deng¹, C.M. Disteche¹. 1) Department of Pathology, University of Washington, Seattle, WA; 2) Department of Genome Sciences, University of Washington, Seattle, WA; 3) Department of Statistics, University of California, Riverside, CA; 4) Department of Integrative Biology & Physiology, University of California, Los Angeles, CA; 5) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN.

Two major problems in studying sex chromosome aneuploidy (Klinefelter, XXX and Turner syndromes) in human are: (1) there is considerable variability in the background genotype of individuals, and (2) relevant cell types are often inaccessible. To address these issues we derived human iPSC lines from individuals with sex chromosome aneuploidy and constructed pairs of isogenic lines with a different number of X chromosomes. These cells were differentiated into neuronal cells and cardiomyocytes to address cognitive deficits and heart defects in sex chromosome aneuploidy. In parallel, frontal cortex, cerebellum, heart and spleen tissue samples from mice with sex chromosome aneuploidy on the same strain background were collected. DNA methylation and transcriptome studies showed significant differences in autosomal CpG methylation and gene expression in cells and tissues with a different number of X chromosomes. Notably, human primary cells with a specific type of X aneuploidy (XXY, XXX, or X) clustered in terms of differentially methylated regions on autosomes, with significant alterations in the DNA methylation landscape being maintained after reprogramming to iPSCs. These results indicate that having an abnormal number of sex chromosomes influences the epigenetic landscape across the genome. We also found dysregulated autosomal genes in mouse tissues with sex chromosome aneuploidy. Interestingly, brain consistently showed fewer dysregulated genes compared to heart and spleen across the different types of aneuploidy, suggesting genome-wide effects of X aneuploidy in a tissue/cell-type specific manner.

1983W

Characterization of large familial copy number variation in regions of low gene density and chromatin accessibility. W.A. Khan^{1,2}, R. Singh^{1,2}, Y. Mito^{1,2}, L. Shi^{1,2}, L. Edelmann^{1,2}, S.A. Scott^{1,2}. 1) Sema4, a Mount Sinai venture, Stamford, CT; 2) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY.

It is estimated that ~5-10% of the human genome contains structural copy number variants (CNVs), yet only aberrations that are large and/or rare typically lead to syndromic disorders. The majority of CNVs are small (<500 kb), common, intergenic, and inherited. Interestingly, some large euchromatic CNVs do not result in a penetrant clinical phenotype; however, it is currently unclear how these large aberrations are tolerated and if this is related to gene density and/or other chromatin packaging factors. To determine if large inherited CNVs in otherwise healthy individuals are distinct from pathogenic structural aberrations with respect to content and accessibility, deidentified clinical chromosomal microarray (CMA) data was interrogated from over 11,000 patients tested in a clinical cytogenomics laboratory. All identified CNVs that were 1.5-10 Mb were assessed, which revealed 90 large CNVs in the cohort. Of these, 70 were either coincident with known syndromic regions, a recurrent benign pericentromeric variant, or were transmitted to the proband as the consequence of a balanced structural rearrangement. The remaining 20 CNVs ranged in size from 1.5-10.4 Mb, were not implicated in any recurrent syndrome, and were typically inherited and classified as variants of uncertain significance – likely benign. Additionally, they were distributed over fourteen chromosomes and did not exhibit any parent of origin bias. Chromatin accessibility as determined by deoxyribonuclease I hypersensitivity (DNase HS) from the ENCODE project and CpG dinucleotide content were compared between these 20 CNVs and 20 well-characterized pathogenic syndromic CNVs (2.4-9.9 Mb) as defined by ClinGen. Notably, normalized DNase HS cluster scores, morbid gene content, and CpG island number and length were an order of magnitude greater in the pathogenic CNVs compared to the inherited likely benign CNVs ($F=21.7$, $p<0.0001$). The inherited CNVs also broadly corresponded to GC poor isochores at an 850 haploid band length. Taken together, these results indicate that CNV gene density is a better predictor of clinical significance than CNV size. In addition to informing on CNV interpretation, these data also suggest that a paucity of morbid genes and a closed chromatin environment permits certain regions of the genome to tolerate a change in copy number, which likely influences their population frequencies by allowing transmission in subsequent meioses.

1984T

Using humanized mouse models to understand the role of human accelerated regions in human evolution. E.V. Dutrow¹, D. Emera¹, S. Uebbing¹, T.P. Nottoli², J.P. Noonan^{1,3}. 1) Department of Genetics, Yale School of Medicine, New Haven, CT; 2) Department of Comparative Medicine, Yale School of Medicine, New Haven, CT; 3) Kavli Institute for Neuroscience, Yale School of Medicine, New Haven, CT.

The evolution of uniquely human morphology required genetic changes that altered embryonic development. Human Accelerated Regions (HARs), which include conserved developmental regulatory elements whose sequences have substantially diverged in human, are prime candidates for driving human-specific developmental modifications. Several HARs have been shown to encode developmental enhancers with human-specific patterns of activity using transgenic reporter assays or direct comparisons of enhancer activity in human and nonhuman tissues using epigenetic marks. However, the potential human-specific biological functions of HARs have not yet been examined using genetic models. To address this question, we used genome editing to generate a humanized mouse model for the most prominent HAR experimentally studied to date, *HACNS1*. In transgenic reporter assays, *HACNS1* shows increased activity in the developing limb and pharyngeal arch compared to its nonhuman primate orthologs, and it shows increased levels of epigenetic marks associated with active enhancers in the human limb. The human-specific *in vivo* functions of *HACNS1* remain unknown; however, humanized mouse models provide a novel tool for deciphering the function of *HACNS1* in a living system. We precisely replaced the orthologous 1.2kb region of the mouse genome with *HACNS1*. In this mouse model, we find that *HACNS1* recapitulates its human-specific increase in enhancer activity observed in the developing human limb. *HACNS1* also shows increased activity in the mouse pharyngeal arch. To establish the human-specificity of *HACNS1* activity in our genetic model, we generated a second mouse model for the chimpanzee ortholog of *HACNS1*. The chimpanzee ortholog showed consistently weaker activity in both limb and pharyngeal arch compared to *HACNS1*. Furthermore, we find that *HACNS1* drives increased expression of a critical developmental transcription factor, *Gbx2*, in embryonic limb and pharyngeal arch. We are carrying out both molecular and tissue-level functional characterization of development in humanized mouse tissues to identify phenotypic correlates of human biological traits. Our results show that HARs can directly alter gene expression in genetic models, establishing humanized mice as a platform for experimental studies of human evolution.

1985F

Evaluation of a targeted custom capture bisulfite sequencing approach. D. Daley¹, R. Munthali¹, A. Eslami¹, M. Wan¹, A. Sandford¹, K. Oros², C. Greenwood^{2,3}. 1) Medicine, University of British Columbia, Vancouver, BC, Canada; 2) Lady Davis Institute for Medical Research, Montreal, Quebec, Canada; 3) Departments of Oncology, Epidemiology, Biostatistics & Occupational Health, and Human Genetics, McGill University, Quebec Canada.

CpG methylation can be modified by environmental exposures, such as smoking and medication. Differential methylation of CpG sites likely contributes to the etiology of many common complex diseases. As such, there is an increased interest in evaluating CpG methylation and conducting epigenome-wide association studies. CpG methylation can be assessed using a variety of methods including whole genome bisulfite sequencing which can be prohibitively expensive, pyrosequencing of small regions, or the relatively inexpensive, and commonly utilized chip based approaches such as Illumina's 27K, 450K and Epic arrays. While cost effective, there are significant limitations of the array based approaches including the limited number of CpG sites that can be assessed and the number of CpG sites that demonstrate variable methylation. An emerging alternative to array based approaches, targeted capture bisulfite sequencing can offer several advantages including greater number of CpGs that can be assessed, greater percentage of CpG sites with variable methylation, joint sequencing of both SNPs and CpG sites facilitates evaluation of both haplotype and SNP effects. We recently piloted the Illumina TruSeq Methyl Capture Epic bisulfite sequencing approach using 48 samples selected from asthma studies. We compared the library content (TruSeq) to Illumina's arrays (27K, 450K and Epic). We found that the majority of CpG sites on the arrays (27K (100%), 450K (98%) and Epic (97.5%)) overlapped with the targeted capture library. However, the proportion of CpG sites with variable methylation is considerably higher 50% TruSeq vs. 9% 450K. A full analysis of the pilot of 48 samples is currently underway but our initial findings suggest that bisulfite targeted capture sequencing offers a cost effective alternative to whole genome bisulfite sequencing.

1986W

Systematic mapping of chromatin state and accessibility during mouse fetal development. D.U. Gorkin^{1,14}, I. Barozzi⁴, Y. Zhao^{2,9}, Y. Zhang², H. Huang^{2,12}, A.Y. Lee², B. Li², J. Chiou^{7,12}, A. Wildberg³, B. Ding³, M. Wang³, J.S. Stratton¹¹, J.M. Davidson¹¹, Y. He^{9,10}, V. Afzal⁴, J.A. Akiyama⁴, I. Plajzer-Frick⁴, C.S. Pickle⁴, M. Kato⁴, T.H. Garvin⁴, S. Chee², J.M. Cherry¹¹, Y. Shen⁸, J.R. Ecker¹⁰, W. Wang³, K. Gaulton⁷, D.E. Dickel⁴, A. Visel^{4,5,6}, L.A. Pennacchio^{4,5,13}, B. Ren^{2,3}, ENCODE Consortium. 1) University of California San Diego, San Diego, CA; 2) Ludwig Institute for Cancer Research, 9500 Gilman Drive, La Jolla, CA 92093-0653, USA; 3) University of California San Diego School of Medicine, Department of Cellular and Molecular Medicine, Institute of Genomic Medicine, and Moores Cancer Center, 9500 Gilman Drive, La Jolla, CA 92093-0653, USA; 4) Functional Genomics Department, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California 94720, USA; 5) U.S. Department of Energy Joint Genome Institute, Walnut Creek, California 94598, USA; 6) School of Natural Sciences, University of California, Merced, Merced, California 95343, USA; 7) University of California, San Diego School of Medicine, Department of Pediatrics, 9500 Gilman Drive, La Jolla, CA 92093-0653, USA; 8) Institute for Human Genetics and Department of Neurology, University of California, San Francisco, San Francisco, California 94143, USA; 9) Bioinformatics and Systems Biology Graduate Program, University of California, San Diego, La Jolla, California 92093; 10) Howard Hughes Medical Institute, Salk Institute for Biological Studies, La Jolla, California 92037; 11) Stanford University School of Medicine, Department of Genetics, Stanford, California, USA; 12) Biomedical Sciences Graduate Program, University of California, San Diego, La Jolla, California 92093; 13) Comparative Biochemistry Program, University of California, Berkeley, CA 94720, USA; 14) Center for Epigenomics, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0767, USA.

Embryogenesis requires epigenetic information that allows each cell to respond appropriately to developmental cues. The state and accessibility of chromatin are core components of a cell's epigenome, modulating the function of the underlying genome sequence. However, a comprehensive survey of chromatin dynamics during mammalian embryogenesis has been lacking except for very early stages. We will report unpublished results of a systematic effort by the ENCODE consortium to profile histone modifications and chromatin accessibility in a diverse panel of mouse tissues at 8 time points from 10.5 days post conception until birth. In total we performed 1,128 ChIP-seq assays and 132 ATAC-seq assays across 72 distinct tissue-stages. We will highlight results of several integrative analyses that we think will be of particular interest to human geneticists: First, we find strong enrichment of human trait-associated SNPs ("GWAS SNPs") in the orthologs of mouse fetal cis-regulatory elements (cREs). Many of these enrichments are tissue-dependent, linking traits to specific fetal tissues and cREs therein. In forebrain where we have single-cell ATAC-seq data, we further map disease enrichments to specific cell types within the tissue. Second, we leveraged the breadth of our data to generate genome-wide predictions of cRE target genes. We find that cREs linked to Mendelian disease genes (as cataloged by OMIM) are significantly more likely to harbor GWAS SNPs, highlighting the potential of non-coding variation at Mendelian disease genes to contribute risk for non-mendelian phenotypes. Third, we find that Transcription Factor (TF) genes underlying Mendelian disease are strongly and significantly overrepresented among targets of Polycomb-mediated repression during development (comprising 66% of all Mendelian disease TFs). These results point to incomplete or failed repression as a potential disease mechanism for further exploration. Beyond these highlights, we will describe a multi-tiered compendium of functional annotations for the mouse genome including chromatin state maps, an extensive catalog of cREs with dynamic developmental activity, and a collection of 150 in vivo transgenic reporter assays that reveal a quantitative relationship between epigenomic signals and validation rate. Given the uniquely critical role of the mouse as a model system in biomedical research, we believe that these resources and findings will be of high interest to the human genetics community.

1987T

Collecting roadmap and ENCODE data into reference epigenomes. J.A. Hilton, ENCODE Data Coordination Center. Genetics, Stanford University, Palo Alto, CA.

The Encyclopedia of DNA elements (ENCODE) project has produced data from more than 9,000 experiments using a variety of techniques to study the structure, regulation, and transcription profiles of human and mouse genomes. The data from these experiments first pass through the ENCODE Data Coordination Center (DCC) for basic validation and meta-data standardization before they are openly available at the ENCODE site (<https://www.encodeproject.org/>). The ENCODE portal also hosts data from external projects, including the NIH Roadmap Epigenomics Mapping Consortium. In order to align with the existing structure of the ENCODE portal, Roadmap experimental metadata were carefully curated, and raw data were collected from production labs and public repositories. The import of these experiments allows for Roadmap data to be searched, downloaded, and analyzed alongside ENCODE data. Furthermore, the ENCODE DCC has processed Roadmap data using the uniform processing pipelines designed by the ENCODE consortium. This provides high-quality and consistent data with cross-project compatibility. One of the aims of the Roadmap project was to provide to the scientific community a set of reference epigenomes for human cells and tissues. Using data produced from the Roadmap project, 111 reference epigenomes were assembled, plus an additional 16 reference epigenomes from ENCODE data, each of which have been modeled on the ENCODE portal. The ENCODE DCC has also curated data produced by the ENCODE consortium to compile reference epigenomes for human and mouse biosamples, following guidelines set forth by the International Human Epigenome Consortium. These reference epigenomes, combined with the open access to ENCODE's uniform processing pipeline code, provide the framework for ease of comparison and integration within future studies from the scientific community.

1988F

Rare genetic variation at transcription factor binding site modulates local DNA methylation profiles. A. Martin Trujillo¹, N. Patel¹, F. Richter¹, B. Jadhav¹, P. Garg¹, B.D. Gelb^{1,2}, A.J. Sharp^{1,3}. 1) Department of Genetics & Genomic Sciences and The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Department of Pediatrics, Icahn School of Medicine at Mount Sinai, New York, NY; 3) Graduate School of Biomedical Sciences, Icahn School of Medicine at Mount Sinai, New York, NY.

Background: DNA methylation is an important epigenetic regulator, playing a critical role in many cellular processes such as transcriptional regulation, X-chromosome inactivation, genomic imprinting or maintenance of genome stability. Although DNA methylation has been extensively characterized, the mechanism by which this epigenetic mark is targeted to specific regions in the genome remains unclear. Interestingly, recent studies have revealed that local DNA methylation profiles might be dictated by *cis*-regulatory DNA sequences so called methylation-determining regions, which activity relies on DNA-binding factors. In agreement with this finding, we have recently shown that disruption of CTCF-binding sites by rare single nucleotide variants (SNVs) can underlie *cis*-linked DNA methylation changes in patients with congenital anomalies (PMID: 29802345). These data raise the hypothesis that rare genetic variation at regulatory sequences, such as transcription factor binding sites (TFBS), might contribute to local DNA methylation patterning and represent a global mechanism of genome regulation. **Results:** To investigate the association of rare SNVs within TFBS with altered DNA methylation, we analyzed genome-wide DNA methylation profiles (Illumina 850k), SNVs derived from whole genome sequencing data from 247 unrelated individuals, and 133 active TFBS motifs derived from ENCODE ChIP-Seq data. We observed an enrichment for local outlier DNA methylation values in individuals carrying rare variants at TFBS when compared to non-carriers. Interestingly, this enrichment is greater for rare variants at TFBS than those outside these motifs, highlighting the role of regulatory sequences on modulating proximal DNA methylation profiles. Furthermore, greater difference in position weight matrix scores between alternate and reference alleles is associated with larger methylation changes at nearby CpG sites. This finding suggests that rare SNPs at TFBS might negatively influence the TF-DNA binding that can lead to an altered local DNA methylation profile. **Conclusions:** Our findings provide firstly, insights into the effect of rare variation at TFBS on shaping local DNA methylation affecting genome regulation and secondly, a rationale to incorporate DNA methylation data to interpret the functional role of rare variants.

1989W

An atlas of epigenetic signatures across the human phenome associated with polygenic burden of disease. T.G. Richardson, M. Suderman, T.R. Gaunt, G. Davey Smith, C.L. Relton. MRC Integrative Epidemiology Unit (IEU), Population Health Sciences Institute, University of Bristol, Bristol, United Kingdom.

DNA methylation holds considerable potential as a biomarker for disease diagnosis and prognosis. However, to date epigenome-wide association studies (EWAS) have been confined to a small proportion of complex traits across the human phenome. Polygenic risk scores (PRS) derived from large populations can provide a feasible alternative to help elucidate and prioritise CpG sites which may be implicated in disease based on genetic liability. Such findings can be utilised in conjunction with PRS to improve prediction of later life disease susceptibility. We curated PRS for 212 complex traits from large-scale consortia and the UK Biobank study. Using blood-derived data from the Accessible Resource for Integrated Epigenomics Studies (ARIES) project, we conducted EWAS analyses at 5 different stages across the life course (birth, childhood, adolescence, pregnancy and middle age) for each PRS. In total, 916 associations survived conventional EWAS corrections ($P < 1.0 \times 10^{-7}$) across all time points (at 375 unique CpG sites with 115 complex traits). These associations between genetic liability for a given trait and DNA methylation may indicate that this CpG site is mechanistically involved in disease pathogenesis. For example, concerning cardiovascular disease, *SORT1/CELSR2* methylation was associated with the polygenic burden of coronary heart disease and high cholesterol, whereas DNA methylation at *LIPA* was associated with risk of myocardial infarction. There were also associations detected at loci previously linked with cancer susceptibility, such as *MSMB* methylation and risk of prostate cancer. Despite using methylation data derived from blood, we identified associations where expression data from GTEx implicates genes that appear to be expressed predominantly in other tissue types. For example, there were several associations with neurological traits where CpG-mapped genes appear to be expressed specifically in brain tissue. These included associations between *LINGO1* and *DOC2A* methylation with facets of neuroticism, as well as between *LHFPL3* methylation and a measure of general cognition. These findings support the utility of whole blood as a proxy for tagging epigenetic signatures in tissue types which may be more pertinent to the disease of interest. We have developed a web application to disseminate our findings to the scientific community. This resource should prove valuable for future studies undertaking in-depth evaluations of epigenetic mechanisms in disease.

1990T

Exposure to polybrominated biphenyl (PBB) associates with DNA methylation differences across the genome. S.W. Curtis¹, V. Kilaru², D.O. Cobb², M.L. Terrell³, C.J. Marsit⁴, M. Marcus^{3,4,5}, K.N. Conneely⁶, A.K. Smith^{2,7}. 1) Genetics and Molecular Biology, Emory University, Atlanta, GA; 2) Department of Gynecology and Obstetrics, School of Medicine, Emory University, Atlanta, GA; 3) Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA; 4) Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, GA; 5) Department of Pediatrics, School of Medicine, Emory University, Atlanta, GA; 6) Department of Human Genetics, School of Medicine, Emory University, Atlanta, GA; 7) Department of Psychiatry and Behavioral Sciences, School of Medicine, Emory University, Atlanta, GA.

In 1973, approximately 6.5 million Michigan residents were exposed to an industrial mixture of polybrominated biphenyl (PBB), an endocrine-disrupting compound, when it was added to farm animal feed during a factory accident. Highly exposed individuals and their children have numerous health problems, including genitourinary conditions, spontaneous abortions, and early puberty, though the underlying mechanism behind these health problems remains unknown. Other endocrine-disrupting compounds have been linked to epigenetic differences, but no epigenetic studies have been done for PBB. In this study, DNA from the blood of individuals with PBB exposure was interrogated with the MethylationEPIC BeadChip (N = 659). Associations between each of the ~850,000 CpG sites and current, serum PBB levels were tested with a linear regression that controlled for age, sex, and cell type proportions. After multiple test correction (FDR <0.05), 1758 CpG sites associated with total PBB levels. These CpGs were not enriched in any particular biological pathway, but were enriched in enhancer and insulator regions, and depleted in regions near the transcription start site or in CpG islands ($p < 0.05$). In order to better understand the function of the enhancers that were enriched, PBB-associated CpGs were also interrogated for transcription factor binding sites, correlations with hormones, and links to gene expression. They were also more likely to be in ARNT and ESR2 transcription factor binding sites ($p = 2.32 \times 10^{-29}$ and $p = 1.01 \times 10^{-18}$, respectively), and there was significant overlap between CpGs associated with PBB and CpGs associated with estrogen ($p < 2.2 \times 10^{-16}$). PBB-associated CpGs were also enriched for CpGs known to be associated with gene expression in blood (eQTM) ($p < 0.05$). These eQTMs were associated with the expression of genes that were enriched for pathways related to immune function and endocrine-related autoimmune disease (FDR <0.05) and the signal from PBB correlated with the methylation signal from low-grade inflammation ($R = 0.34$, $p\text{-value} = 4.82 \times 10^{-7}$). Taken together, these results indicate that exposure to PBB is associated with differences in epigenetic marks that suggest that it is acting similarly to estrogen and is associated with dysregulated immune system pathways.

1991F

Trans-eQTL analysis informs tissue-specific regulation of gene expression. A. Battle¹, B. Jo², A. Saha¹, B.J. Strober¹, A. He¹, P. Parsana¹, B.E. Engelhardt², The GTEx Consortium. 1) Johns Hopkins University, Baltimore, MD; 2) Princeton University, Princeton, NJ.

Recent work suggests that regulation of gene expression by distal genetic loci across the genome contributes substantially to the heritability of both gene expression and complex human traits. These distal effects on gene expression can be captured by analysis of trans-eQTLs (distal expression quantitative trait loci) and are thought to act through complex cellular networks specific to particular tissues and cell types. Identifying trans-eQTLs in human data has been challenging due to power, tissue cell-type heterogeneity, confounding factors, and tissue-specificity. We have identified over 400 trans-eQTLs across 30 tissues in the latest (v8) release of the Genotype Tissue Expression (GTEx) project data, which includes 838 postmortem donors with whole genome and RNA sequencing of 17,382 samples. We present analysis of statistical methods along with functional characterization of trans-eQTLs using this large multi-tissue cohort. Statistical considerations for trans-eQTL detection include improving power, filtering rampant false positive associations, detecting regulatory hotspots, and assessing the false discovery rate at multiple levels including genomic loci and target genes. Functionally, while trans-eQTL variants overlap greatly with cis-eQTL variants, they display some notable differences. As we and others have previously found, we show that trans-eQTLs are more tissue-specific than cis-eQTLs, often only displaying strong effects in just a single tissue and independently replicating only in data for matched tissues. We evaluate the enrichment of trans-eQTLs in regulatory elements, protein-coding regions, and near transcription factors, and investigate tissue-specific regulatory mechanisms including overlap with regulatory networks. These results provide the broadest characterization to date of trans-eQTLs across human tissues, essential to understanding tissue-specific gene regulation and effects of genetic variation that propagate through cellular networks to affect complex phenotypes with tissue-specific origins.

1992W

Fine-mapping regulatory variants across 49 tissues. A. Brown¹, F. Hormozdiari², F. Aguet³, E.T. Dermizakis¹, X. Wen³, the GTEx consortium. 1) Department of Genetic Medicine and Development, University of Geneva, Geneva, Switzerland; 2) The Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, Massachusetts 02142, USA; 3) Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109, USA.

Identification of causal regulatory variants can substantially enhance our understanding of the mechanisms of gene expression and the links between expression and disease. The latest GTEx data provided RNA-seq and WGS from 838 individuals across 49 tissues. We applied 3 published fine-mapping methods to these data: CaVEMaN, CAVIAR and dap-g. We find 68,948 high confidence causal variants ($P > 0.8$), with a median number of 6 variants in a 95% credible set. Fine-mapping approaches can also find independent eQTLs, accounting for multiple testing. We find 7.2% (kidney-cortex) to 45% (nerve-tibial) of eGenes have >1 eQTL, with differences explained by sample size ($p = 0.95$). We used these results to derive properties of regulatory variants; these could be used, e.g., for developing personalised genome interpretation methods to predict deleterious variants solely from genomic context. We annotated eQTLs to 715 ChIP-seq peaks and 54 transcription factors motifs, greatest enrichment is in peaks for a mark of active promoters (H3K4me3 peaks contain 16.5% of regulatory variants, a 372 fold enrichment (FE)). Motifs show more enrichment (up to 1,100FE), but explain a far smaller proportion of regulatory variation (0.3%). We also explored using combinations of annotations to predict causal regulatory variants, to produce more tailored inference on specific variants. Assigning priors to variants, 32% of variants have unique prior weight due to unique genomic context. Using these priors we see a significant increase in lead variant causal probability ($P < 1e-316$, median $P = 0.41$ vs 0.27). In 58% of cases we find a different variant to be causal than when considering only statistical information; these new variants had a slightly higher replication rate in a new dataset (88.1% vs 87.6%). We used the fine-mapping results in whole blood to look for colocalization with GWAS hits for 4 lipid traits. We find 36 out of 148 GWAS hits colocalize with eQTLs ($P > 50\%$), explaining 21 new GWAS hits compared to the previous GTEx release. This demonstrates the improvement due to the new release, with 332 new samples and genotypes from WGS. We show here that leveraging molecular phenotype associations, annotations and posterior probabilities can bring benefits to diverse sets of analyses. These include looking at properties of regulatory variants, assessing allelic heterogeneity, and exploring the relationship with GWAS signals, moving us towards the goal of personalized non-coding genome interpretation.

1993T

Tissue-variability of eQTL effects and the role of transcription factors. E.D. Flynn^{1,2}, S.E. Castel^{1,2}, P. Mohammadi^{1,2,3,4}, T. Lappalainen^{1,2,4}, GTEx Consortium. 1) Department of Systems Biology, Columbia University, New York, NY; 2) New York Genome Center, New York, NY; 3) The Scripps Translational Science Institute, The Scripps Research Institute, La Jolla, CA; 4) Shared last author.

The GTEx Consortium has identified thousands of local expression quantitative trait loci (*cis*-eQTLs) in the human population that affect gene expression in multiple tissues. Over a third of GTEx eQTLs are active across all or almost all surveyed tissues and may have different effect sizes in different tissues, but the molecular mechanisms of this cross-tissue variability of eQTL effect size are poorly understood. Since eQTLs are enriched in transcription factor binding sites (TFBS), we developed a model of eQTL effect size as a function of transcription factor (TF) activity levels across tissues – low (absence of TF binding) or high (saturation of TF binding) tissue TF level results in reduced eQTL tissue effect size, with maximum eQTL effect at mid-range TF levels. We investigated mechanisms of tissue-variability of eQTL effect size (allelic fold change) in the GTEx v8 data release, which includes 838 individuals' whole genome sequences and their RNA-seq data from a total of 15,201 samples across 49 tissues. We first examined the correlation of eGene expression with eQTL absolute effect size across tissues and identified 3,477 expression-correlated eQTLs (Spearman corr., 5% FDR, $N = 26,499$), with 60% positive and 40% negative correlations. This result shows that expression level and genetic regulatory effects are not independent phenomena, and we found that the expression-correlated eQTLs have distinct genetics, gene features, and functional annotations. Importantly, they are enriched for TFBS overlap compared to non-correlated eQTLs, and a portion of these TFBS-overlapping eQTLs are also correlated with the relevant TF gene expression level across tissues. This indicates that TF activity levels may indeed modify eQTL effects across tissues, and it allows identification of upstream molecular drivers of eQTL activity. We proceeded to find several examples of eQTLs that co-localize with GWAS signals, directly disrupt TFBS, and correlate with TF levels across tissues, thus pinpointing the likely causal variant and the direct molecular regulator of disease associations. Our results provide a basis for identifying regulatory mechanisms of eQTL activity and for understanding eQTL effect size variability across tissues in the context of TF activity. This has major implications for future work investigating context-specific mechanisms of GWAS loci and their effects on human phenotype and disease.

1994F

Human skeletal muscle *trans*-eQTL meta-analysis identifies potential biological connections across chromosomes. L. Guan¹, D.L. Taylor^{2,3}, R.P. Welch⁴, A.U. Jackson⁴, N. Narisu², A. Varshney⁶, R.D. Albanus¹, H.M. Stringham⁴, X. Wen⁴, L.L. Bonnycastle², M.R. Erdos², J. Tuomilehto^{5,7,8,9}, H.A. Koistinen^{6,10,11}, F.S. Collins², M. Boehnke⁴, S.C. Parke^{1,5}, L.J. Scott⁴. 1) Department of Computational Medicine & Bioinformatics, University of Michigan, Ann Arbor, Michigan, United States of America; 2) National Human Genome Research Institute, National Institutes of Health, Bethesda, United States of America; 3) European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridgeshire, United Kingdom; 4) Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan, United States of America; 5) Department of Human Genetics, University of Michigan, Ann Arbor, Michigan, United States of America; 6) Department of Public Health Solutions, National Institute for Health and Welfare, Helsinki, Finland; 7) Department of Neurosciences and Preventive Medicine, Danube University Krems, Krems, Austria; 8) Diabetes Research Group, King Abdulaziz University, Jeddah, Saudi Arabia; 9) Dasman Diabetes Institute, Dasman, Kuwait; 10) Department of Medicine and Abdominal Center: Endocrinology, University of Helsinki and Helsinki University Central Hospital, Haartmaninkatu 4, Helsinki, Finland; 11) Minerva Foundation Institute for Medical Research, Biomedicum 2U, Tukholmankatu 8, Helsinki, Finland.

Skeletal muscle represents ~40% of lean body weight and constitutes the largest organ. Many diseases affect skeletal muscle and differences in gene expression may cause or be caused by disease states. Genetic variants that affect RNA expression levels are called expression quantitative trait loci (eQTL). To date, most studies have focused on *cis*-eQTLs, which affect expression of nearby genes; fewer studies have examined *trans*-eQTLs, which affect expression of distal genes. Here we identify skeletal muscle *trans*-eQTLs from 301 FUSION (Finland-United States Investigation of NIDDM Genetics) study participants and 491 GTEx (Genotype-Tissue Expression project) V7 donors. We performed inter-chromosomal *trans*-eQTL analysis for protein-coding and lincRNA genes within each study for SNPs with minor allele count >10 (7.1M SNPs in FUSION, 9.2M in GTEx) imputed using the Haplotype Reference Consortium panel. We used an additive genetic model to test for the association of inverse normalized gene expression with imputed autosomal allele count, adjusting for potential known and latent confounders. Using an approximate Bonferroni-corrected threshold (5×10^{-8} divided by the number of tested genes), we identified 4 *trans*-eGenes (genes whose expression levels are significantly associated with a genetic variant) out of 15906 genes in FUSION and 3 *trans*-eGenes out of 16216 genes in GTEx. Of 9 genes identified in the GTEx V6p publication (10% genome-wide FDR), in FUSION we replicated SNP-gene pairs in 4 genes ($p < 8 \times 10^{-5}$); of the remaining genes, 4 genes had SNP-gene pairs $p > .05$, and 1 gene was not tested in FUSION (SNP not imputed). As an initial analysis, we performed a sample size-based meta-analysis of 245M variant-gene pairs with p -values < 0.05 in both FUSION and GTEx and identified 21 *trans*-eGenes ($p \leq 3 \times 10^{-12}$) out of 15348 genes. Within these, we identified *trans*-eQTL signals for *DYNLL1* ($p = 4.0 \times 10^{-14}$, Chr 12) and *DYNLL2* ($p = 1.4 \times 10^{-16}$, Chr 17) that both overlapped a *cis*-eQTL signal for the transcription factor *ATMIN*. *ATMIN* has been shown to influence the expression of *DYNLL1*. We calculated enrichment for the lead *trans*-eQTLs in genomic regions with different classes of skeletal muscle chromatin states and found significant enrichment in the active transcription start site state ($p = 0.0012$, fold-enrichment 3.8). These findings will help increase our understanding of cross-chromosomal regulation of genetic expression by genetic variation in skeletal muscle.

1995W

Conserved DNA methylation helps drive tissue-specific gene expression differences across primates. L.E. Blake¹, J. Roux^{1,2}, I. Hernando-Herraez³, N.E. Banovich^{1,4}, R. Garcia Perez³, C. Joyce Hsiao⁵, C. Chavarria¹, T. Marques-Bonet⁶, Y. Gilad^{1,5}. 1) Department of Human Genetics, University of Chicago, IL; 2) Department of Biomedicine, University of Basel, Basel, Switzerland; 3) Institute of Evolutionary Biology, Universitat Pompeu Fabra, Barcelona, Spain; 4) Integrated Cancer Genomics Division, Translational Genomics Research Institute, Phoenix, AZ; 5) Department of Medicine, University of Chicago, Chicago, IL.

Gene regulatory changes are hypothesized to play an important role in primate adaptive evolution and human disease. Previous work has identified tissue-specific gene expression patterns that are conserved across primates, suggesting that these patterns are important for tissue-specific functions. However, we still know remarkably little about the mechanisms driving the conserved tissue-specific gene expression profiles. Therefore, we chose to assess the contribution of one epigenetic mechanism, DNA methylation, to primate regulatory evolution across multiple tissues. We collected RNA and DNA from 4 tissues (heart, kidney, liver and lung) from each of our 4 human, chimpanzee and rhesus macaque individuals. We then interrogated the CpG methylation status genome-wide with whole-genome bisulfite sequencing and collected gene expression profiles using RNA-sequencing. Consistent with previous studies, we found that tissue-specific differentially expressed (DE) genes in our dataset were highly conserved. Furthermore, we observed that tissue-specific differentially methylated regions (T-DMRs) are also conserved across the three species. Ultimately, we are interested in identifying DNA methylation differences that drive interspecies and intertissue gene expression differences in primates. To this end, we estimated that at least 14% of interspecies gene expression differences across primates in a given tissue can be explained, at least in part, by DNA methylation differences. DNA methylation differences could contribute to a similar percentage (11-22%) of gene expression differences across tissues in these species. Finally, we discovered that conserved DNA methylation significantly contributes to 10-30% of genes with tissue-specific gene expression. Overall, this work identifies DNA methylation differences that may underlie species-specific traits and tissue-specific functions.

1996T

Gene circuitry mapping is a powerful tool for identification of drug-gable targets and appropriate chemistries for therapeutically relevant outcomes. *D. Bumcrot, E. Cohick, Y. Liu, O. Chung, C. Smith, J. Weinstein, B. Schwartz, G. Whissell, A. Sehgal, C. O'Donnell, A. Sigova, I. Grossman.* Camp4 Therapeutics, Cambridge, MA.

Background: A wide range of under-served diseases could be therapeutically addressed by increasing or decreasing the expression of specific genes. This includes rare disorders caused by loss- or gain-of-function mutations, as well as common diseases with well validated genetic modifiers. Genetic studies of human disease, including GWAS, are increasing the number of well supported associations between diseases and specific genes, thus presenting opportunities for therapeutic intervention. We have built a gene circuitry mapping platform that discovers novel connections between signaling pathways that control transcription and the regulation of specific disease-associated genes. **Hypothesis:** Our goal is to utilize this platform across primary human cells derived from a range of tissues to develop testable therapeutic hypotheses for a broad spectrum of diseases. To test the validity of the platform we selected a well-characterized disease gene, *PCSK9*, and elucidated the control circuitry, relevant druggable nodes and existing chemistries that affect it at clinically meaningful levels. **Methods:** We have generated a gene circuitry map for primary human hepatocytes through the application of functional genomics techniques. These include RNA-seq profiles in response to more than 300 different experimental conditions and ChIP-seq profiles measuring the binding of approximately 60 DNA binding proteins and chromatin marks. Together with the results of DNA-DNA interaction assays (including HiChIP) these data inform an algorithmic approach to decoding hepatocyte gene regulatory circuitry. This enables prediction of novel mechanisms for modulating genes of interest via signaling pathway nodes; many of which are targets of approved drugs. As an example, we mapped the transcriptional control of the metabolic disease-associated gene, *PCSK9*. **Results:** The map correctly identifies known mechanisms regulating *PCSK9*, such as elevation in response to statin treatment, but importantly also predicts pathways that heretofore have not been linked to its transcriptional regulation. Follow-up testing validates our predictions and proposes novel approaches for therapeutic intervention. **Conclusions:** The Gene Circuitry platform yields readily testable predictions for druggable targets, often with existing clinical-stage chemistries. In this way, it can provide a significant head start to drug development programs for a wide range of diseases.

1997F

Mapping gene regulatory dynamics during cardiomyocyte differentiation at single cell resolution. *R. Elorbany^{1,2}, R. Dohn², K. Rhodes³, H. Eckart⁴, A. Selewa⁵, S. Pott^{3,4}, A. Basu^{4,6}, Y. Gilad^{3,4}.* 1) Interdisciplinary Scientist Training Program, University of Chicago, Chicago, IL; 2) Committee on Genetics, Genomics, and Systems Biology, University of Chicago, Chicago, IL; 3) Department of Human Genetics, University of Chicago, Chicago, IL; 4) Department of Medicine, University of Chicago, Chicago, IL; 5) The Graduate Program in Biophysical Sciences, University of Chicago, Chicago, IL; 6) Center for Nanoscale Materials, Argonne National Laboratory.

Characterizing the impact of genetic variation on gene expression is essential for understanding human development and disease. As cells differentiate over time, the regulation of gene expression must also change dynamically. In this study, we aim to detect genetic variants associated with the change in gene expression over time, which we term “dynamic QTLs.” Dynamic QTLs can provide insight into the temporal patterns of gene regulation, and how these dynamic processes vary between individuals. However, during differentiation, the composition of cell types in a sample is heterogeneous and may also differ between individuals. To determine whether observed variation in gene expression is due to differences in gene regulation or in cell composition, we investigate gene expression over time at single cell resolution. Using the Drop-seq platform, we collected single cell RNA-seq data at 5 time points during differentiation from human induced pluripotent stem cells (iPSCs) to cardiomyocytes in 10 individuals. The 5 time points used in this study were chosen to be representative of broad developmental stages from iPSC to cardiomyocyte, as determined by bulk gene expression data from cells collected daily during the differentiation. Using single cell expression data, we can characterize the cell composition of each sample at each time point. Cells in the same differentiation stage can be directly compared between individuals for QTL mapping. We expect to identify QTLs associated with inter-individual differences in gene expression, both at each time point independently (static QTLs) and across the entire timecourse (dynamic QTLs). Our single-cell expression data will allow us to investigate variation between individuals in both cell composition and gene regulation over time. Thus far, we have collected single cell RNA-seq data using Drop-seq on two individuals at five time points during differentiation. By the final stage of our cardiomyocyte differentiation, gene expression profiles of single cells show clustering into distinct cell populations. In preliminary analysis, we can identify genes most significantly associated with distinct cell populations at each time point, including cardiac-specific gene *TNNI2* which increases in expression in a particular cell population over time. As such, we have confidence in our ability to investigate gene expression within a heterogeneous sample of cells as they transition between cell states during differentiation.

1998W

Hi-C-based characterization of the landscape of physically interacting regions and interaction mechanisms across six human cell lines using HIPPIE2. *P.P. Kuksa¹, A. Amlie-Wolf^{1,2}, Y.-C. Hwang³, B.D. Gregory⁴, L.-S. Wang^{1,2}.* 1) Penn Neurodegeneration Genomics Center, Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA; 2) Genomics and Computational Biology Graduate Group, University of Pennsylvania Perelman School of Medicine; 3) DNAnexus, Inc., Mountain View, California, USA; 4) Department of Biology, University of Pennsylvania, Philadelphia, PA.

Most regulatory chromatin interactions are mediated by various transcription factors (TFs) and involve physically-interacting regions (PIRs) that are regulatory elements such as enhancers, insulators, or promoters. To map these interactions across cell lines, we developed a novel approach, HIPPIE2. HIPPIE2 accepts raw Hi-C reads as input and provides extensive characterization of (1) PIRs and their potential regulatory roles using epigenomic annotations, (2) high-confidence PIR-PIR interactions, and (3) mediating TF complexes based on binding motifs and ChIP-seq experiments. Unlike standard genome binning approaches with fixed-sized bins at 10K-1Mbp resolution, HIPPIE2 calls physical locations of PIRs with better precision and higher resolution by using restriction enzyme cutting sites and strand orientations in the ligation step of the Hi-C protocol. We applied HIPPIE2 to in situ Hi-C datasets across 6 human cell lines (GM12878, IMR90, K562, HMEC, HUVEC, and NHEK; Rao et al, 2014) with matched functional genomics datasets from ENCODE and Roadmap. HIPPIE2 detected 1,042,738 unique PIRs genome-wide across cell lines, with an average length of 1,005bps and high reproducibility (92.3% shared PIRs in GM12878 replicates). We discovered PIRs are enriched for epigenetic marks (H3K27ac 2.18x and H3K4me1 1.96x), suggesting active regulatory roles. With respect to cell-type specificity, on average 32.8% of PIRs were shared between cell lines. HIPPIE2 identified 2.8M significant (Fit-Hi-C) intrachromosomal PIR-to-PIR interactions. PIR-PIR interactions were enriched for TF binding with experimental evidence of TF binding from ChIP-seq in 27.2% of interactions. 50,608 PIR-PIR interactions were between enhancers and promoters based on genomic and epigenomic annotations of PIRs. These enhancer-promoter interactions were mediated by 45 distinct TFs (41 in enhancers and 40 in promoters) obtained from overrepresented TF-TF pairs. 35 TFs were enriched in >1 cell lines and 10 were cell line-specific. Several of these factors are known to be involved in DNA looping and long-distance regulation (e.g., *YY1*, *SP1*, *CTCF*, *MYC*, *NR4A1*, *MEF2*, *PAX5*). Our analysis demonstrated HIPPIE2 identifies PIRs, TFs, and interactions with excellent sensitivity and specificity thanks to its high resolution and reproducibility of PIR detection. HIPPIE2 (<https://bitbucket.com/wanglab-upenn/HIPPIE2>) is useful for studying genome-wide regulatory mechanisms and how TF binding mediates these interactions.

1999T

High resolution genetic mapping of causal regulatory interactions in the human genome. *N. Kumasaka, A. Knights, D. Gaffney.* Wellcome Sanger Institute, Cambridge, United Kingdom.

Physical interaction of distal regulatory elements in three-dimensional space poses a challenge for studies of common disease because noncoding risk variants may be substantial distances from the genes they regulate. Experimental methods to capture these interactions, such as chromosome conformation capture (CCC), usually cannot assign causal direction of effect between regulatory elements, an important component of disease fine-mapping. Here, we developed a Bayesian hierarchical approach that models causal interactions between regions of open chromatin using two-stage least squares. We applied our model to a novel ATAC-seq data from 100 individuals mapping over 15,000 high confidence causal interactions. Strikingly, the majority (>60%) of interactions we detected were over distances of <20Kb, a range where CCC-based methods perform poorly. For a fraction of loci, we identify a single variant that alters accessibility across multiple peaks, and experimentally validate one, the BLK/FAM167A locus associated with risk for multiple autoimmune disorders, using CRISPR engineering. Our study highlights how association genetics of chromatin state provides a powerful complement to CCC-methods to map regulatory variants to genes.

2000F**Characterizing the genetic basis for variation in m⁶A methylation.** K.

Luo¹, Z. Zhang^{2,3,4,5}, M. Qiao⁶, X. He¹, M. Stephens^{1,7}, C. He^{2,3,4,5}. 1) Department of Human Genetics, University of Chicago, Chicago, IL; 2) Department of Chemistry, University of Chicago, Chicago, IL; 3) Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL; 4) Institute for Biophysical Dynamics, University of Chicago, Chicago, IL; 5) Howard Hughes Medical Institute, University of Chicago, Chicago, IL; 6) School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX; 7) Department of Statistics, University of Chicago, Chicago, IL.

There has been great progress over the past few years in linking genetic variation to molecular phenotypes by mapping QTLs across the gene regulatory cascade from DNA to RNA and protein. Recent breakthroughs in transcriptome-wide profiling of N⁶-methyladenosine (m⁶A), the most abundant internal modification in eukaryotic mRNA, has opened a new frontier in gene regulation. This reversible RNA methylation affects almost every stage of mRNA metabolism, from processing to translation and decay. However, the mechanistic role that genetic variants have in guiding m⁶A changes and the functional interactions between m⁶A and other layers of gene regulation are still largely unknown. To understand the genetic basis for inter-individual variation in m⁶A methylation, we measured m⁶A methylation using m⁶A-seq in 60 Yoruba lymphoblastoid cell lines (LCLs), for which genotypes are available and multiple other molecular phenotypes have been extensively characterized. We identified genetic variants that are associated with m⁶A methylation level (m⁶A-QTLs). We observed that m⁶A-QTLs are strongly enriched within binding sites of many important RNA-binding proteins and also active chromatin features. To dissect the causal relationships between m⁶A and other molecular phenotypes in the cascade of gene regulation, we integrated m⁶A-QTLs with other layers of molecular QTLs in LCLs identified in earlier studies (such as transcription rate, translation, and protein). Our integrative analysis uncovered interesting yet complex relationships between m⁶A methylation and other layers of gene regulation. Finally, we discovered significant enrichment of m⁶A-QTLs for GWAS traits, which may suggest novel mechanisms for complex diseases at the RNA methylation level.

2001W**Genetic loci associated with inter-individual variation in alternative polyadenylation and PolII pausing provide new insight into the gene regulatory code.** B.E. Mittleman¹, K. Patterson², S. Warland², S. Pott^{2,3}, Y.

Gilad^{1,2,3}, Y.I. Li². 1) Genetics, Genomics and Systems biology, University of Chicago, Chicago, IL; 2) Genetic Medicine, University of Chicago, Chicago, IL; 3) Human Genetics, University of Chicago, Chicago IL.

Characterizing the effects of genetic variation on gene regulation can shed light on the mechanisms that underlie inter-individual differences in complex traits and disease. While there has been immense progress in our ability to identify the effects of genetic variants on gene expression and splicing, we still have a poor understanding of many regulatory steps, including active transcription and polyadenylation. To begin addressing this gap, we assayed inter-individual variation in two post transcription initiation regulatory steps, polymerase II pausing and polyadenylation site usage. We obtained high quality Native elongation transcript sequencing (NET-seq) data to detect strand-specific polymerase II density along gene bodies at base-pair resolution. As expected, NET-seq reads are enriched at promoter proximal regions and at exon-intron boundaries. These data allow us to detect native elongating transcription-QTLs (NetQTLs) associated with polymerase II pausing. We also use these data to examine the alternative splicing kinetic coupling model, namely, the hypothesis that internal pol II pausing reduces splice site competition leading to increased exon inclusion. In turn, we use 3' sequencing data collected in the same LCLs to detect polyadenylation QTLs (apaQTLs) associated with alternative polyadenylation site usage. We collected these data for whole cell and nuclear specific mRNAs to detect short-lived transcripts prior to nuclear export. We can thus test the role of alternative polyadenylation usage on mRNA localization and export. Finally, by integrating both data types from this study, we ask whether Pol II pausing can be used to explain inter-individual variation in alternative polyadenylation usage.

2002T

Detection and replication of rare variant driven gene expression outliers via joint modeling of total and allele-specific expression. *B. van de Geijn¹, G. McVicker², J.D. Blischak³, Y. Gilad³, A.L. Price¹.* 1) Harvard T.H. Chan School of Public Health, Boston, MA; 2) Salk Institute for Biological Studies, La Jolla, CA; 3) University of Chicago, Chicago, IL.

Non-coding genetic variants can lead to large changes in gene expression and may underly instances of genetic disease. Indeed, a recent study found that individuals with outlier expression ($|z\text{-score}| > 2$) for a given gene often harbor a rare variant nearby (Li et al. Nature 2017); however, it is challenging to determine whether a particular outlier is genetically driven by a rare variant. We developed a rigorous statistical method to identify rare variant driven outliers by jointly modeling total and allele-specific expression, leveraging a previous method for identifying quantitative trait loci (WASP: van de Geijn, McVicker et al. Nature Methods 2015). Total expression for each individual is directly modeled as count data, taking into account gene- and individual-level overdispersion as well as technical effects on expression measurements; this boosts power to detect outliers over methods that require normalization. Similarly, we modeled allele specific expression directly as read counts while taking into account overdispersion and technical effects. We distinguish between four possible scenarios based on model likelihood. The first is no-effect, where the individual being tested has total expression similar to the others and no allelic imbalance. The second, which we are most interested in, is cis effect, where an individual has outlier total expression and allelic imbalance. The third and fourth are trans or monoallelic effects where there either outlier expression or allelic imbalance, but not both. To evaluate our method, we analyzed RNA-seq data from lymphoblastoid cell lines of 30 Yoruba trios. We expect any heritable outlier in one of the parents to also occur in the child half of the time. At 5% FDR, we identified 15 parent-gene pairs with evidence of a cis effect (likelihood ratio test for cis vs. no-effect model and higher likelihood for the cis model vs. trans or monoallelic models). Of these, 11 were overexpressed (median 20x change in expression) and 4 were underexpressed (median 1/10x change). 5 of the 15 genes were also outlier genes in the corresponding children. We also identified 7 parent-gene pairs that would have been labeled as outliers based on total expression, but that showed stronger evidence of trans or monoallelic effects and were thus not labeled as outliers by our method. None of these replicated in the corresponding children. Our findings highlight the need for an outlier model that can specifically identify cis effects.

2003F

Conserved enhancer domains reflect gene pathogenicity and evolutionary buffering of expression. *X. Wang^{1,2}, D.B. Goldstein^{1,2}.* 1) Columbia University, New York, NY; 2) Institute of Genomic Medicine, New York, NY.

Non-coding transcriptional regulatory elements are critical for controlling the precise spatiotemporal expression of genes. We used maps of enhancer activity across human tissues to define a genome-wide "enhancer domain score" that represents the number of nucleotides implicated in regulating the expression of a human gene. We demonstrate that the number of conserved bases in enhancers linked to a gene is reflective of disease pathogenicity, and provides information independent of and complementary to commonly-used metrics including pLI and RVIS. The enhancer domain score can also predict the specific affected tissues for developmental disease genes. Additionally, genes with large conserved enhancer domains are depleted for *cis*-acting genetic variants that disrupt gene expression. Finally, we show that the cumulative burden of rare genetic variation in enhancers is associated with gene expression differences, and that genes with large enhancer domains are buffered against these disruptive effects. Our results demonstrate that dosage-sensitive disease genes have evolved robustness to the disruptive effects of genetic variation by expanding their transcriptional regulatory domains. This resolves a puzzle in the genetic literature on why disease genes are depleted for *cis*-eQTLs and establishes a framework for the identification of non-coding regulatory variation with phenotypic consequences. .

2004W

Methylation quantitative trait loci in CD4⁺ T lymphocytes, monocytes, and neutrophils in Japanese individuals. S. Komaki¹, Y. Shiwa¹, R. Furukawa¹, T. Hachiya¹, H. Ohmomo¹, Y. Sutoh¹, R. Otomo¹, M. Satoh¹, J. Hitomi², K. Sobue², M. Sasaki², A. Shimizu¹. 1) Division of Biomedical Information Analysis, Iwate Tohoku Medical Megabank Organization, Iwate Medical University, Yahaba, Iwate, Japan; 2) Iwate Tohoku Medical Megabank Organization, Iwate Medical University, Yahaba, Iwate, Japan.

Purpose DNA methylation (DNAm) centrally regulates gene expression; inter-individual variations of DNAm are associated with various human traits. Numerous trait-associated single-nucleotide polymorphisms (SNPs) identified via genome-wide association studies (GWAS) are not located in coding regions and mechanisms underlying the effects of genomic variations on traits remain unclear; DNAm may cause this effect. This study aimed to reveal how genomic variations affect DNAm via methylation quantitative trait locus (mQTL) analysis. Methods We used CD4⁺ T lymphocytes (CD4T; n = 102), monocytes (n = 102), and neutrophils (n = 94) isolated from peripheral blood by fluorescence-activated cell sorting with >95 % purities. DNAm profile at ~24 million CpGs and genotypic data of ~9 million SNPs were obtained from each cell type. Based on these datasets, we conducted cis-mQTL analysis with cis-proximity defined as a distance < 1 Mbp. False discovery ratio for multiple testing correction was applied with a significance threshold q-value ≤ 0.05. We further investigated the enrichment of the mQTLs among cardiovascular/immunological disease-associated SNPs listed in the PheGenI database. Results & Discussion Among >3 billion SNP-CpG pairs, mQTL analysis revealed 115,390,389 SNP-CpG associations (mQTLs). mQTL number was largest in neutrophils, followed by CD4T and monocytes. Cell type-specific mQTLs accounted for >50% of all detected mQTLs. Among 5997 cardiovascular disease-associated SNPs and 3887 immunological disease-associated SNPs, 71% and 55%, respectively, were associated with DNAm level (mQTL), of which 35% were cell type-specific mQTLs. Hence, numerous phenotype-altering SNPs are accompanied by changes in DNAm. Expression quantitative trait methylation analysis is needed to understand how genomic variation affects gene expression through changes in DNAm. Moreover, the association between genomic variations and DNAm levels was highly cell type-specific, thereby potentially contributing to specific functions of each cell type. We have published the resulting mQTL data on iMETHYL, a genome browser (<http://imethyl.iwate-megabank.org>). This is the first cell type-specific and comprehensive multi-omic genome browser that provides QTL-related information, which will greatly contribute to understanding the biological consequences of genomic variation in terms of phenotypic variation.

2005T

Penalized functional regression method for across-platform imputation of methylation levels. G. Li¹, L. Raffield², M. Logue^{3,4}, M.W. Miller^{2,4}, Y. Li^{2,5,6}.

1) Dept. of Statistics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Dept. of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC; 3) National Center for PTSD at VA Boston Healthcare System; 4) Dept. of Psychiatry, Boston University School of Medicine; 5) Dept. of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC; 6) Dept. of Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, NC.

DNA methylation of cytosine residues at CpG dinucleotides is one of the most extensively studied epigenetic marks. CpG site methylation is involved in not only normal development but also risk and progression to many diseases. With the advancement of powerful technologies such as DNA methylation microarrays and bisulfite sequencing techniques, geneticists are able to profile DNA methylation at increasingly higher resolution. However, different DNA methylation profiling platforms employ different designs with varying resolutions, which hinders joint analysis of methylation data. Thus, we proposed a penalized functional regression (PFR) model to impute missing-by-design methylation levels. We have previously demonstrated the power of our PFR approach: it improved imputation accuracy when imputing the Illumina HumanMethylation450 (HM450) BeadChip from the HumanMethylation27 (HM27) BeadChip, compared with alternative methods. In this study, we further extend our method to impute from HM450 to 850,000 methylation sites on the recently released HumanMethylationEPIC BeadChip. We compared our PFR model with state-of-the-art methods including lasso, elastic net regression and machine learning based methods. We applied our PFR model to 144 samples measured both by HM450 and EPIC. Among the most challenging 3,000 CpG sites with the largest variations, 1036 probes exceeding a threshold of 0.25 for our under-dispersion statistic (a measure of estimated imputation quality) achieved a prediction accuracy > 75.1% when dichotomizing the DNA methylation level with a cutoff of 0.5; with a more stringent threshold of 0.5, 530 probes passing quality control had accuracy 80.0%. In addition, we found data harmonization across platforms before applying PFR model crucial: our non-parametric regression-based harmonization approach further improved quality by 1% with the more stringent 0.8 threshold. Our results validate that, by incorporating the nonlocal information in functional terms, the PFR model is a convenient and efficient imputation tool. The imputed methylation values can be used to boost power for discovery in subsequent epigenome-wide association studies (EWAS).

2006F

Exploring the longitudinal association between loneliness and methylation of genes associated with the Conserved Transcriptional Response to Adversity (CTRA). D. Phillips¹, Y. Wang², N.L. Pedersen², S. Hagg², C.A. Reynolds¹. 1) University of California, Riverside, Riverside, CA; 2) Karolinska Institutet, Stockholm, Sweden.

Perceived loneliness is a potentially modifiable factor linked with a wide range of adverse physical, cognitive, and mental health outcomes. While the mechanisms of these associations remain largely unknown, prior work has shown altered gene expression in blood leukocytes of chronically lonely (vs. non-lonely) individuals, with increased expression of genes associated with inflammation and reduced expression of genes associated with fighting viral infections in lonely compared to non-lonely persons, an expressional pattern which has been referred to as a *conserved transcriptional response to adversity* (CTRA) and has been linked with other circumstances involving objective and perceived stress. Despite these observations, the relation between loneliness and DNA methylation remains unexplored in published work. To address this gap, we assessed the longitudinal association between loneliness and methylation at 1,586 CpG sites associated with 105 CTRA genes in up to 375 twins from the Swedish Adoption Twin Study of Aging (SATSA) across up to 18 years. Partial correlations adjusting for age, sex, and number of longitudinal assessments were calculated between baseline loneliness levels ($Base_{Lonely}$), longitudinal means (M_{Lonely}), and variability (SD_{Lonely}), with longitudinal patterns of methylation (M_{CpG} , SD_{CpG}) at each site. Partial correlations exceeded an absolute effect size of .1 or higher for 504 of 1,586 sites (32%) across 4 sets of comparisons of $Base_{Lonely}$ with M_{CpG} and SD_{CpG} , M_{Lonely} with M_{CpG} , and SD_{Lonely} with SD_{CpG} . $Base_{Lonely}$ was associated with M_{CpG} at 149 sites at .1 or higher (r_p range = -.20 to .18), and with SD_{CpG} at 200 sites (r_p range = -.24 to .21). M_{Lonely} was associated with M_{CpG} at 177 sites (r_p range = -.20 to .20). SD_{Lonely} was associated with SD_{CpG} at 162 sites (r_p range = -.19 to .22). Some of the CpGs flagged suggest an enrichment of sites within particular genes. For example, 48 of 147 CpG sites near or within the *SLC12A7* gene suggested loneliness may correlate differentially with methylation levels and variability across time; among these, baseline loneliness correlated with higher M_{CpG} at 18 of 20 sites, and negatively with SD_{CpG} at 17 of 20 sites. These findings indicate that longitudinal growth analyses are warranted to explore the link between feelings of loneliness and level and change in methylation at these sites.

2007W

Identification of rare genetic variants contributing to tissue-specific patterns of extreme expression. N.M. Ferraro¹, J.R. Davis², E.K. Tsang¹, X. Li³, M. Gloudemans¹, A.J. Scott⁴, I.M. Hall⁴, A. Battle⁵, S.B. Montgomery^{2,3}, GTEx Consortium. 1) Biomedical Informatics Program, Stanford University, Stanford, CA; 2) Genetics Department, Stanford University, Stanford, CA; 3) Pathology Department, Stanford University, Stanford, CA; 4) McDonnell Genome Institute, Washington University School of Medicine, St. Louis, MO; 5) Department of Computer Science, Johns Hopkins University, Baltimore, MD.

While numerous genetic association studies have associated common variants with human traits, the specific molecular consequences of rare variation are often unknown or difficult to predict, despite their implications in both rare and common disease. Recent work has shown that some rare variants induce large changes in gene expression across many human tissues. However, for other variants, effects are likely to manifest differently in distinct tissues. To identify such effects, we have developed a method to discover gene expression outliers that directly accounts for patterns of correlation of expression levels across tissues, and can therefore detect tissue-specific changes. We applied this method in a dataset of gene expression data of 550 individuals and 23 tissues, with a median of 12 tissues available per individual, from the Genotype-Tissue Expression project. Outliers discovered by our correlation-aware method, like the primarily distinct set of outliers discovered via standard methods, are significantly enriched for nearby rare variation, including SNPs, indels, and structural variants ($p < 2e-16$). Further, nearby rare variants identified in both methods are enriched in highly conserved regions of the genome and are predicted to be consequential via fitCons and CADD scores. However, unlike existing approaches, our correlation-aware approach is able to discover outliers that manifest in a single tissue. These tissue-specific outliers are significantly enriched ($p=0.008$) for nearby rare variants located in matched-tissue specific enhancer regions, and are depleted for rare variants in enhancer regions specific to other tissues ($p=4.35e-10$), indicating that rare variants in tissue-specific enhancers may be driving the extreme expression observed in that tissue. Our results provide a new avenue to study rare variant biology in tissue-specific contexts. In doing so, we improve the discovery of functional rare variants with tissue-specific effects, which can provide novel insights into genome interpretation and genetic disease risk.

2008T

High-throughput untranslated region engineering and screening. J.

Cao^{1,2}, E. Novoa^{2,3}, Z. Zhang^{2,3}, M. Kellis^{2,3}, T. Lu^{1,3}. 1) Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA; 2) Department of Electrical Engineering & Computer Science, Massachusetts Institute of Technology, Cambridge, MA; 3) Broad Institute of MIT and Harvard, Cambridge, MA.

Gene therapy holds the promise of treating the diseases that cannot be treated using traditional small molecule or protein drugs. The *in vivo* drug expression levels depend highly on the genetic elements near the coding region. In this work, we developed a comprehensive strategy to identify novel 5' untranslated region (5' UTR) sequences that can enhance protein expression from the human cytomegalovirus (CMV) promoter in human muscle cells. We first designed a library of 12,000 5' UTR sequences based on genomic searches and *in silico* computation, then designed a novel recombinase-based screening strategy to screen the library in mammalian cells. In addition, we developed next-generation sequencing (NGS) analysis pipelines that allow investigating the correlations between the 5'UTR sequences and protein expression levels. We identified two 5' UTR candidates that can improve protein expression by greater than 40% in human rhabdomyosarcoma (RD) cells, which outperformed the commonly used introns. We envision these synthetic UTRs can be used to enhance *in vivo* drug productivity for DNA vaccines and mRNA therapeutics.

2009F

Extensive differential gene expression by sex in human skeletal muscle

tissue. S.C. Hanks¹, A.U. Jackson¹, R.P. Welch¹, M.R. Erdos², N. Narisu², H. Stringham¹, J. Tuomilehto^{3,4,5,6}, H.A. Koistinen^{7,8,9}, S.C.J. Parker^{10,11}, F.S. Collins², M. Boehnke¹, L.J. Scott¹. 1) Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI; 2) National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 3) Chronic Disease Prevention Unit, National Institute for Health and Welfare, Helsinki, Finland; 4) Center for Vascular Prevention, Danube University Krems, Krems, Austria; 5) Diabetes Research Group, King Abdulaziz University, Jeddah, Saudi Arabia; 6) Dasman Diabetes Institute, Dasman, Kuwait; 7) Department of Health, National Institute for Health and Welfare, Helsinki, Finland; 8) Department of Medicine and Abdominal Center: Endocrinology, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland; 9) Minerva Foundation Institute for Medical Research, Helsinki, Finland; 10) Department of Computational Medicine & Bioinformatics, University of Michigan, Ann Arbor, MI; 11) Department of Human Genetics, University of Michigan, Ann Arbor, MI.

Skeletal muscle biology is important in multiple human diseases including obesity and diabetes. These and many other diseases exhibit differences between sexes. Identifying genes that are differentially expressed by sex in human skeletal muscle will enhance our understanding of biological differences in this disease relevant tissue. To quantify gene expression levels, we performed strand-specific mRNA-seq on 301 vastus lateralis biopsies from Finnish individuals with a range of glucose tolerances as part of the FUSION (Finland-United States Investigation of NIDDM Genetics) study. We tested for differential gene expression by sex using a linear mixed model with random effects adjusting for technical covariates, batch, age, BMI, and estimates of muscle fiber type proportions. Out of 21,743 tested genes, we found 9,512 genes, including 9,117 on autosomal chromosomes, to be significantly differentially expressed by sex (false discovery rate=0.05). 2,923 genes show higher expression in males while 6,589 genes show higher expression in females. The majority of signals were unaffected by adjusting for muscle fiber type estimates. GRTP1 Antisense RNA 1 transcript had higher expression in males and showed the most significant differential expression by sex (p-value=2.1x10⁻⁶) of autosomal genes. Gene set enrichment testing, performed using RNA-Enrich, revealed that mitochondria-related genes from the nuclear genome had higher expression in males. Our results show a high degree of sex-biased gene expression in skeletal muscle tissue, indicating significant differences in gene regulation or cell type composition by sex. Further analyses are needed to study the possible relationships between the differences in expression levels and their biological impact on health and disease.

2010W**Variance QTL mapping in human induced pluripotent stem cells.** A.K.

Sarkar¹, P.Y. Tung¹, J.D. Blischak^{1,2}, J. Hsiao¹, Y.I. Li^{1,3}, M. Stephens^{1,4}, Y. Gilad^{1,3}. 1) Department of Human Genetics, University of Chicago, Chicago, IL; 2) Committee on Genetics, Genomics, and Systems Biology, University of Chicago, Chicago, Illinois, USA; 3) Department of Medicine, University of Chicago, Chicago, Illinois, USA; 4) Department of Statistics, University of Chicago, Chicago, Illinois, USA.

Advances in single cell sequencing have revealed that the variance of gene expression across cells is a complex trait which could provide a previously unconsidered regulatory mechanism to explain human development and disease. However, previous studies of gene expression variance have been limited to yeast, or to carefully chosen gene pathways in humans. Here, we take an unbiased, genome-wide approach to identify variance quantitative trait loci (vQTLs). We reprogrammed 53 Yoruba lymphoblastoid cell lines into induced pluripotent stem cells and generated RNA-Seq data for median 95 cells per individual and total 5,447 single cells. We find 279 eQTLs (FDR 10%), of which 83% replicate in matched bulk RNA-Seq data, and 148 vQTLs (FDR 10%). We find that the vast majority of vQTLs replicate as eQTLs, suggesting they can be explained away by mean effects. We demonstrate the current study is underpowered to directly find dispersion effects, or variance effects which are independent of mean effects, and estimate at least 500 cells per individual (condition) would be required to have 80% power to find a two-fold change in dispersion. Overall, our results do not rule out the existence of genetic effects on dispersion, and will guide the design of future studies.

2011T**Multi-tissue analysis of mitochondrial post-transcriptional methylation.**

A.T. Ali¹, K. Small², A. Hodgkinson¹. 1) Department of Medical and Molecular Genetics, School of Basic and Medical Biosciences, King's College London, London, United Kingdom; 2) Department of Twin Research and Epidemiology, School of Life Course Sciences, King's College London, London, United Kingdom.

The mitochondrial genome is transcribed as poly-cistronic strands that are subsequently cleaved and processed by nuclear encoded proteins to release mature mitochondrial RNA products. Post-transcriptional processes therefore play a crucial role in regulating the expression of the mitochondrial genome, and in particular, variable methylation of mitochondrial RNA can lead to changes in transcript abundance and protein levels, ultimately impacting mitochondrial function. Here, we characterise variation in mitochondrial post-transcriptional methylation levels at functionally important positions in the mitochondrial genome across a total of 11,552 samples from 39 tissue/cell types. Using genome-wide association studies, we replicate previously found associations between missense mutations in nuclear encoded mitochondrial genes (*MRPP3*, *SLC25A26* and *MTPAP*) and mitochondrial RNA methylation levels across multiple tissue types, and we additionally identify 4 novel independent associations at genome-wide significance ($P < 5 \times 10^{-8}$). Newly identified genetic variants are predicted to be functional or show evidence of significantly ($P < 0.05$) mediating RNA methylation levels through the expression of nearby nuclear-encoded mitochondrial genes. Finally, these nuclear loci co-locate with genetic variants associated with complex disease, suggesting that mitochondrial post-transcriptional methylation may be involved in disease progression.

2012F

A limited set of transcriptional programs define major histological types and provide the molecular basis for a cellular taxonomy of the human body. A. Breschi^{1,2,3}, M. Muñoz-Aguirre^{1,4}, D. Garrido-Martín^{1,2}, C. Davis⁵, V. Wucher¹, S. Djebali⁶, J. Gillis⁷, D. Pervouchine⁷, A. Vlasova¹, A. Dobin⁵, C. Zaleski⁶, J. Drenkow⁵, C. Dnyko⁵, A. Scavelli⁶, F. Reverter-Comas¹, T. Gingeras⁵, R. Guigó^{1,2}. 1) Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Dr. Aiguader 88, 08003 Barcelona, Catalonia, E-08003; 2) Universitat Pompeu Fabra (UPF), Barcelona, Catalonia, E-08003; 3) Stanford University, Department of Genetics, Stanford, 94305, USA; 4) Universitat Politècnica de Catalunya. Departament d'Estadística i Investigació Operativa. 08034 Barcelona, Catalonia; 5) Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 11742, USA; 6) GenPhySE, Université de Toulouse, INRA, INPT, INP-ENVIT, Castanet Tolosan, France; 7) Skolkovo Institute for Science and Technology, 3 Nobel st., Moscow, Russia 143025.

Organs and tissues are complex structures composed of millions of cells of diverse morphology and function. Tissue transcriptomes, thus, represent the average behaviour of genes across a highly heterogeneous collection of primary cells. How the transcriptional profiles of each of these cells types relate to the profiles of tissues and organs is still poorly understood. Here we have monitored by RNA-Seq the transcriptome of multiple human primary cells from multiple tissues. We found that their transcriptional profiles cluster into a few broad cell types: endothelial, epithelial, mesenchymal, neuronal and blood cells. The clustering is recapitulated using independent transcriptomic and epigenomic data. The influence of tissue of origin contributes very little (<4%) to the transcriptomes of primary cells. We identified about 3000 genes specific to the three cell types. These include a core set of transcription factors (TFs), showing strong co-expression patterns, and thus likely candidates to drive cell type specificity. Cell type specific genes are enriched for motifs of TFs specific to the same cell type. Cell type specific genes are mostly a vertebrate innovation, appearing early in the evolution of this lineage, with epithelial specific genes being the most evolutionary dynamic. We employed a method to estimate the proportion of cells from each type in a given tissue from gene expression values in that tissue in the GTEx collection of tissue transcriptomes. We found that, although we characterized a very limited number of primary cells from a limited number of tissues, the basic cell types capture a large proportion (>70%) of cellular composition of many human tissues. Through the analysis of the GTEx catalog of histological images and associated annotations, we show that our inferred cellular composition precisely defines tissue type and captures morphological heterogeneity in the samples. We identified changes in the cell type composition occurring with age and sex in a few tissues. We found that departures from the normal cellular composition of tissues correlate with histological phenotypes associated to diseases. Alterations of the cellular composition are particularly relevant in cancer, where they can be even associated to different stages of disease progression. The collection of primary cells transcriptomes produced here is a unique resource to understand tissue biology, serving as interface between tissue and single cell transcriptomics. .

2013W

Investigation of human elements regulating escape from X-chromosome inactivation in mouse models. S. Peeters¹, A. Korecki^{1,2}, E. Simpson^{1,2}, C. Brown¹. 1) Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada; 2) Centre for Molecular Medicine and Therapeutics, University of British Columbia, Vancouver, BC, Canada.

X-chromosome inactivation (XCI) epigenetically silences one X chromosome in every cell in female mammals; however, approximately 12–20% of genes escape from XCI in humans and continue to be expressed from both the active and inactive X chromosomes. Only 3–7% of X-linked genes escape in mouse, and tend to be distributed as singletons while human escape genes often cluster in larger domains, suggesting both promoter-centric and regional mechanisms of escape. We have recently shown that despite known species differences between escape gene number and distribution on the X chromosome, human escape genes *RPS4X* and *CITED1* retain their ability to be expressed from the inactive X chromosome in multiple tissues of a transgenic mouse model. The recognition of human elements necessary for escape from XCI in mouse provides a tractable system to assess additional human genes for intrinsic escape ability, and characterize the critical regions responsible for continued expression from an inactive X. As the *RPS4X* construct used above included an entire escape domain bounded by genes subject to inactivation, we next targeted the human escape gene *KDM5C*, which is part of a larger domain of escape not included in its construct. In this case, the well-characterized escape gene became silenced, supporting a need for more distal elements to retain expression in a heterochromatic environment. To validate these results, as well as further explore promoter-centric versus regional models for escape, we are targeting additional constructs to the same X-linked docking site in mouse embryonic stem cells and analyzing expression and promoter DNA methylation for evidence of escape following differentiation and XCI. Expression of genes escaping from XCI may contribute to the phenotypes of X aneuploidies as well as influence male and female susceptibility to disease. As X-chromosome reactivation is being examined therapeutically for some X-linked disease, understanding of how X-linked gene regulatory mechanisms work may provide insights into utilization of such therapeutic approaches.

2014T

Male-female subject-specific XCI-adjusted differential gene expression reveals changes in active-X expression of escape genes. *R. Sauteraud, D. Liu.* Biostatistics, Penn State University, Hershey, PA.

X chromosome inactivation (XCI) mechanism randomly silences one female X (Xi) and only expresses the alleles on the active X (Xa). ~10% of the X-linked genes escape XCI and express both copies of alleles and it is estimated that ~30% of the X-linked genes may escape in a subset of individuals, but remain silenced in others. Often used to assess X-chromosome inactivation (XCI) status of X-linked genes and gain insight into the mechanisms of inactivation and escape, Male-female differential gene expression analysis (DGEA) suffers from the major drawback that it only allows gene level calls and does not account for variable escape between individuals. DGEA is further complicated by the changes in the level of expression of the Xi in escape genes at the subject level. Using our recently developed RNA-Seq based subject-specific XCI inference method to assess Xa/Xi expression in 217 samples of the GEUVADIS project, we adjusted the expression levels to compare expression of the female Xa with males. While the XCI field is usually concerned with the study of the Xi and assumes that the Xa remains largely unaffected, we show that there are significant male-female differences in the active copy of X that may be affected by subject-specific levels of escape. Our results indicate that X-linked escape genes with no Y homologs are up-regulated in females, even after adjusting for Xi expression. We further investigated the role of inter-individual differences on tissue-specific escape and showed that along with the previously reported limited tissue-specificity of XCI, there are differences at the subject level that need to be accounted for when comparing expression levels. XCI-adjusted DGEA could shed some light into the mechanism of escape from inactivation and our XCI-inference allows to circumvent the usual limitations of differential expression by implementing additional subject-level information.

2015F

Comparative analysis of the effects of sex chromosome dosage changes in Turner syndrome and Klinefelter syndrome on local and global gene expression. *X. Zhang^{1,2}, D. Hong¹, S. Ma³, A. Stankov¹, S. Shrestha¹, J. Hallmayer¹, W. Wong³, A. Reiss¹, A. Urban^{1,2}.* 1) Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Palo Alto, CA; 2) Department of Genetics, Stanford University School of Medicine, Palo Alto, CA; 3) Department of Statistics, Stanford University, Palo Alto, CA.

Background: In both Turner syndrome (TS; XO) and Klinefelter syndrome (KS; XXY) copy number aberrations of the X chromosome lead to various developmental symptoms. To date there has not been a comprehensive and directly comparative analysis of TS vs. KS regarding the molecular effects of these sex chromosome dosage changes, on the expression levels of the genes on chromosome X as well as for genes on the autosomes, and of transcriptome-wide network effects. **Methods:** We analyzed gene expression patterns with RNA-Seq and DNA methylation patterns with the CpGiant assay, in blood from 55 clinically well-characterized individuals (13 TS, 14 KS, 13 male controls, 14 female controls). The data from TS patients were analyzed vs. the female controls and the data from KS patients were analyzed vs. the male controls, followed by comparative analyses between KS and TS of differentially expressed genes (DEGs), of transcriptome network effects, and of patterns of DNA methylation. **Results:** On chromosome X we detected 67 DEGs in TS and 42 DEGs in KS. DEGs were enriched on the short arm of chromosome X in both TS and KS. Genome-wide, there are 974 DEGs in TS and 172 DEGs in KS. Interestingly, there is a 'core-group' of differentially expressed genes, where 81 genes are differentially expressed in both TS and KS. These 81 shared DEGs are located on autosomes (50 DEGs) as well as on chromosome X (31 DEGs). Strikingly, all of these 81 core-group DEGs change their levels of expression into opposite directions in TS vs. KS. Weighted correlation network analysis (WGCNA) shows significant co-expression of DEGs on the X chromosome and on the autosomes in both TS and KS, indicating that there are ripple effects of the changes in X chromosome dosage that extend to the autosomes. Broad hypomethylation of the X chromosome is observed in TS whereas hypermethylation of chromosome X is present in KS. Both hypomethylated and hypermethylated regions are present on the autosomes in both TS and KS. **Conclusions:** We carried out direct comparative analysis on the molecular level in TS vs. KS. Multiple DEGs on chromosome X were identified, as well as a large number of DEGs transcriptome-wide. The discovery of shared core-genes indicates the existence of common molecular mechanisms in both TS and KS for transmitting the gene dosage changes to the transcriptome. The findings regarding positional effects and reciprocal DNA methylation patterns on chromosome X further underscore this possibility.

2016W

Gene therapy prevents hepatic tumor development in glycogen storage disease type Ia mice at the tumor-developing stage. J.H. Cho¹, Y.M. Lee^{1,3}, M.F. Starost¹, B.C. Mansfield^{1,2}, J.Y. Chou¹. 1) NIH, Bethesda, MD; 2) Foundation Fighting Blindness, Columbia, MD; 3) University of Connecticut School of Medicine, Farmington, Connecticut.

Hepatocellular adenoma/carcinoma (HCA/HCC) is a severe long-term complication of glycogen storage disease type Ia (GSD-Ia) caused by a deficiency in glucose-6-phosphatase- α (G6Pase- α or G6PC), a key enzyme in gluconeogenesis. Using liver-specific *G6pc*-knockout (L-*G6pc*^{-/-}) mice, we have shown that hepatic G6Pase- α deficiency leads to metabolic reprogramming and defective autophagy, both contribute to tumorigenesis. Here, we show that from pre-tumor to tumor-developing stages, the livers of L-*G6pc*^{-/-} mice displayed metabolic abnormalities and autophagy impairment along with elevated accumulation of p62, a specific autophagy substrate critical for tumorigenesis. Treating tumor-free and tumor-bearing L-*G6pc*^{-/-} mice at the tumor-developing stage with rAAV-G6PC, a G6PC-expressing recombinant adeno-associated virus (rAAV) vector led to restoration of hepatic G6Pase- α expression and prevention of de novo HCA/HCC development. However, rAAV-G6PC-mediated gene transfer could not abrogate the pre-existing HCA/HCC nor restore G6Pase- α expression in these tumors. Significantly, the expression of 11 β -hydroxysteroid dehydrogenase type-1 that mediates local glucocorticoid activation is down-regulated in HCA/HCC lesions, leading to impaired glucocorticoid signaling critical for activation of gluconeogenesis. This suggests local glucocorticoid action downregulation may prevent restoration of G6Pase- α expression in HCA/HCC lesions of rAAV-G6PC-treated L-*G6pc*^{-/-} mice. Examining the tumor lesions in untreated and rAAV-G6PC-treated L-*G6pc*^{-/-} mice showed that in addition to continued p62 accumulation, two tumor promoting pathways, nuclear factor erythroid 2-related factor 2 (Nrf2) and mammalian target of rapamycin (mTOR) signaling were activated. In summary, rAAV-G6PC-mediated gene transfer prevented HCA/HCC development but failed to abrogate the pre-existing tumors that were characterized by the lack of G6Pase- α expression, accumulation of p62, and activation of Nrf2 and mTOR signaling.

2017T

Clec16a knock out mice suffer excessive weight loss, a process mediated through dysregulated lipophagy and rescued using a JAK/STAT inhibitor. R. Pandey¹, H. Heather¹, B. Strenkowski¹, M. Bakay¹, H. Hakonarson^{1,2}. 1) Children's Hospital Of Philadelphia, Philadelphia, PA; 2) Department of Pediatrics, the Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

CLEC16A is implicated in multiple autoimmune diseases. *CLEC16A*'s genomic location next to the suppressor of cytokine signaling 1 (SOCS1) gene and role in modulating mitophagy led us to hypothesize that CLEC16A exerts its effect on a wide variety of immune cells through modulation of SOCS expression and regulation of cytokine signaling, suggesting that perturbations in the molecular link between CLEC16A, mitophagy, lipophagy, and SOCS1 may underlie inflammatory and autoimmune disorders. Consequently, we generated *Clec16a* inducible KO mice to examine the mechanism underlying altered CLEC16A expression and autoimmune susceptibility. Our results demonstrate that *Clec16a* KO mice exhibit metabolic disturbances resulting in severe weight loss which is accompanied by a reduction in white adipose tissue and enhanced inflammatory response and energy wasting despite increased food intake. In addition, our ubiquitous *Clec16a* KO mice exhibited dysregulated mitophagy and lipophagy as well as altered cytokine/chemokine profile, thereby contributing to the immune dysregulation and risk of autoimmunity associated with specific *CLEC16A* variants. Indeed, treatment of *Clec16a* KO mice with the pan-JAK/STAT inhibitor, Tofacitinib, partially rescues the pathological lipodystrophic phenotype and improves survival. Taken together, loss of function variants in *CLEC16A* that are associated with decreased CLEC16A levels may contribute to autoimmunity through dysregulated mitophagy and lipophagy, while drugs modulating lipolysis or the JAK/STAT pathway reverse the process and may be effective in treating and preventing autoimmunity in individuals with risk associated *CLEC16A* variants. Our study underscores for the first time, critical association of CLEC16A with lipid metabolism and unravels the therapeutic potential of the combination of mitophagy and JAK/STAT inhibitors in a myriad of autoimmune diseases and may unveil new directions in treatment interventions of obesity.

2018F

A novel role for a long noncoding RNA in airway differentiation during allergic asthma. G.J. Kwon^{1,2}, M. Yurieva¹, A. Williams^{1,2}. 1) The Jackson Laboratory for Genomic Medicine, Farmington, CT; 2) Department of Genetics and Genome Sciences, University of Connecticut Health, Farmington, CT.

Allergic asthma is characterized by airway hyperresponsiveness to a type 2 adaptive immune response. Immune cells release type 2 cytokines, such as IL-4 and IL-13, which drive airway inflammation. Airway epithelial cells are essential in orchestrating this immune response and undergo distinct morphological changes following type 2 cytokine exposure. However, the molecular pathways regulating these alterations are not fully understood. Long noncoding RNAs (lncRNAs) are defined as non-protein-coding transcripts longer than 200 nucleotides with an increasingly diverse set of functions and are generally more cell-specific than protein-coding transcripts. Despite evidence of their function in regulating the immune response, few lncRNAs have been functionally identified in airway epithelial cells, leading us to speculate their potential role in airway immunity. We cultured primary human bronchial epithelial cells under air-liquid interface (ALI) conditions and performed RNA sequencing from eight individual donors following IL-13 stimulation. The most significantly induced lncRNA was *WFDC21P*, a lncRNA reported to modulate STAT3 dephosphorylation in dendritic cells, but whose function is unknown in airway epithelium. *WFDC21P* is cytoplasmic, highly expressed relative to previously identified lncRNAs in the airway and displays tissue specificity. IL-13 is a critical mediator of differentiation in airway epithelial cells, and we hypothesized *WFDC21P* may mediate differentiation via regulating STAT phosphorylation, similar to dendritic cells. Knockdown via short-hairpin-mediated RNA of *WFDC21P* in an immortalized bronchial epithelial cell line resulted in increased STAT3 signaling, contrary to its effect in dendritic cells. Furthermore, knockdown of *WFDC21P* in primary human bronchial epithelial cells under ALI conditions resulted in morphological disturbances and defective cilia development, as shown by an absence of beating cilia in culture via time lapse and lack of *FOXJ1* expression via quantitative PCR. Our studies reveal a novel role for *WFDC21P* in airway epithelium and type 2 immunity. Current functional studies are underway to further define the effects of *WFDC21P* on the morphology of airway epithelium, in addition to downstream pathways affected by loss of *WFDC21P* following IL-13 exposure via RNA-sequencing. Future studies to understand *WFDC21P*'s mechanism of action will include identifying interacting partners via RNA pulldown followed by mass spectrometry.

2019W

Loss of Cdk5rap3 impairs liver regeneration after partial hepatectomy. Y. Jia, S. Yang, H. Wang, R. Yang, Y. Huang. Institute of Basic Medical Sciences, Peking Union Medical College, Beijing, China.

Liver regeneration can be induced by hepatic damage, such as partial hepatectomy (PH). Patients with terminal liver failure would be prone to high mortality due to impaired regenerative process. Liver regeneration after PH induced the endoplasmic reticulum stress (ER stress) in remnant hepatocytes. Cdk5rap3, also known as C53 and Lzap, exerts the role in regulating cell cycle, proliferation, apoptosis and adherence/invasion. It was also reported as a potential tumor suppressor. Our previous work demonstrated that Cdk5rap3 plays an important role in mammalian liver development and maintaining endoplasmic reticulum homeostasis. Here, we investigated its function in liver regeneration. Cdk5rap3 liver-specific knockout mice (Cdk5rap3^{tm1atm1a} : Alb-Cre, designated as CKO hereafter) were subjected to standard two-thirds partial hepatectomy (PH). We found that Cdk5rap3 deficiency reduced hepatocyte proliferation, induced lipid accumulation, impaired glycogen synthesis, decreased PCNA, PHH3, BrdU-positive cells in the livers of mice after PH. Moreover, the blood glucose was significantly lower in CKO mice compared to WT mice after PH. In regenerating livers, Cdk5rap3 deficiency aggravated ER stress by upregulation of phosphorylation of PERK and Chop expression. Taken together, Cdk5rap3 deficiency impairs liver regeneration, and it might be a possible therapeutic target for patients with terminal liver failure.

2020T

Non-coding *TBX4* and *FGF10* variants suggest complex genetics and potentially compound inheritance in acinar dysplasia of the lungs: Bridging genetics of rare and common variants. J.A. Karolak^{1,2}, M. Vincent^{3,4}, T. Gambin^{5,6}, B. Cogné^{3,4}, O. Pichon³, D. Mowat^{7,8}, A.J. Gifford^{8,9}, H.C. Mefford¹⁰, C. Ward-Melver¹¹, T. Besnard^{3,4}, F. Petit¹², I. Bache^{13,14}, Z. Tümer¹⁵, M. Joubert¹⁶, J. Martinovic¹⁷, D. Carles¹⁸, E. Bieth¹⁹, L. Pasquier²⁰, V. Secc²¹, S. Küry^{3,4}, S. Bézieau^{3,4}, G.M. Schauer²², T.M. Bartel²³, Z.H. Coban Akdemir²⁴, J.R. Lupski^{1,23,24}, P. Szafranski¹, G.H. Deutsch²⁵, B. Isidor³, C. Le Caignec³, P. Stankiewicz¹. 1) Department of Molecular and Human Genetics, Baylor College Medicine, Houston, TX; 2) Department of Genetics and Pharmaceutical Microbiology, Poznan University of Medical Sciences, Poland; 3) CHU de Nantes, Service de Génétique Médicale, Nantes, France; 4) Inserm, CNRS, Univ Nantes, l'institut du thorax, Nantes, France; 5) Institute of Computer Science, Warsaw University of Technology, Warsaw, Poland; 6) Department of Medical Genetics, Institute of Mother and Child, Warsaw, Poland; 7) Department of Medical Genetics, Sydney Children's Hospital, The University of NSW, Sydney, NSW Australia; 8) School of Women's and Children's Health, The University of NSW, Sydney, NSW Australia; 9) Department of Anatomical Pathology, Prince of Wales Hospital, Randwick NSW 2031, Australia; 10) Division of Genetic Medicine, University of Washington, WA, USA; 11) Division of Medical Genetics, Akron Children's Hospital, Akron, OH, USA; 12) Service de Génétique Clinique, CHU Lille, Lille, France; 13) Wilhelm Johannsen Centre for Functional Genome Research, Department of Cellular and Molecular Medicine, University of Copenhagen, Copenhagen, Denmark; 14) Department of Clinical Genetics, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark; 15) Kennedy Center, Department of Clinical Genetics, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark; 16) Service d'anatomo-pathologie, CHU Nantes, Nantes, France; 17) Unit of Fetal Pathology, AP-HP, Antoine Beclere Hospital, Paris; 18) Service d'anatomo-pathologie, CHU Bordeaux, France; 19) Service de génétique médicale, CHU Toulouse, France; 20) Service de génétique médicale, CHU Rennes, France; 21) Service d'anatomo-pathologie, CHU Marseille, France; 22) Department of Pathology, Kaiser Permanente Oakland Medical Center, Oakland, CA, USA; 23) Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA; 24) Texas Children's Hospital, Houston, TX, USA; 25) Department of Pathology, Seattle Children's Hospital, Seattle, WA, USA.

Acinar dysplasia (AcDys) is an extremely rare neonatal lethal lung hypoplasia characterized by diffuse maldevelopment of respiratory bronchioles, alveolar ducts, and alveoli with unclear molecular etiology. In a newborn with AcDys, we reported a *de novo* missense variant in *TBX4* (PMID:27374786), known to be involved in lung branching morphogenesis in mice and associated with childhood onset pulmonary hypertension (PMID:23592887) and ischiocoxopodopatellar syndrome (OMIM147891). In addition, a homozygous missense variant in *FGFR2* was described in a patient with AcDys and ectrodactyly (PMID:27323706). Using SNP and CGH arrays, Sanger sequencing, WES, WGS, and reporter assays, we analyzed 25 AcDys and 28 non-AcDys samples. In unrelated patients with AcDys, we identified one novel heterozygous *de novo* missense variant in *TBX4*, five *de novo* recurrent 2.2 Mb CNV deletions at 17q23.1q23.2 and one overlapping *de novo* deletion of 2.12 Mb at 17q23.2q23.3, all including *TBX4*. Additionally, in two AcDys sibs, WGS analyses revealed an 8.6 kb deletion spanning two *TBX4* exons, inherited from their healthy mother. No coding variants in *TBX4* were identified by WES on the second allele in these patients. Intriguingly, the recurrent 17q23.1q23.2 deletion has been reported in patients with heart and limb abnormalities without the AcDys features (OMIM613355). Importantly, WGS of the second allele revealed the presence of at least one non-coding variant in each AcDys patient with structural or sequence variant involving *TBX4* that are absent in non-AcDys patients with 17q23.1q23.2 deletion and showed a statistically significant enrichment of non-coding variants within and upstream to *TBX4* that overlap putative regulatory elements. In four other patients with lung hypoplasia and AcDys features, we identified heterozygous variants involving *FGF10*, known to be functionally interacting with *TBX4* during lung development: two sequence variants (nonsense and frameshift) in *FGF10* and in two family members, 2.5 Mb non-recurrent CNV deletion at 5p12 encompassing *FGF10*. The occurrence of the *TBX4* or *FGF10* rare coding pathogenic variants on one allele in concomitance with the putative hypomorphic non-coding common variants within the regulatory elements on the other allele, suggests a complex autosomal recessive inheritance in patients with AcDys and support the previously proposed role of *TBX4*-*FGF10*-*FGFR2* signaling pathway in human lung organogenesis. (NIH-R01HL137203).

2021F

Long read assay of DMD mRNA reveals isoform heterogeneity in normal and Duchenne muscle. R.T. Wang^{1,2}, F. Barthelemy^{2,3}, E. Tseng⁴, K. Eng⁴, P. Babayan⁴, M.C. Miceli⁵, S.F. Nelson^{1,2,5}. 1) Human Genetics, University of California Los Angeles, Los Angeles, CA; 2) Center for Duchenne Muscular Dystrophy, University of California Los Angeles, Los Angeles, CA; 3) Department of Microbiology, Immunology, and Molecular Genetics, David Geffen School of Medicine and College of Letters and Sciences, University of California, Los Angeles, Los Angeles, CA; 4) Pacific Biosciences, Inc. Menlo Park, CA; 5) Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA.

The *DMD* gene is one of the largest and most complex genes in the human genome. It encompasses 2.3Mb (0.1%) of the genome and its 11.3kb long mRNA encodes the essential structural protein dystrophin. Defects in *DMD* give rise to Duchenne muscular dystrophy, a X-linked progressive muscle wasting disease with an incidence of approximately 1 in 3500 male births. Dystrophin is encoded by 79 exons of the *DMD* transcript and consists of an N-terminal actin binding domain, 24 spectrin-like repeats, a cysteine-rich domain and a C-terminal domain. Despite much study into the structure and composition of the *DMD* transcript, its isoforms have been difficult to elucidate due to its enormous length and large number of exons. Using PacBio Iso-Seq technology, we performed long read, single molecule sequencing of cDNA generated from healthy human myoblasts. Transcript lengths ranged from 0.14-11.3kb. We found evidence of extensive isoform heterogeneity with over 300 unique transcripts identified. The primary differentiation between isoforms was their differential exon inclusion and exclusion. The canonical muscle-specific 79-exon isoform was most prevalent, but several novel abundant isoforms excluded nearly all spectrin repeats regions. Exons 71 and 78, which both code for the C-terminal region of dystrophin, were commonly spliced out in multiple isoforms. Isoform diversity was approximately 75% lower in the dystrophic sample from a patient with a deletion in exons 3-23. Overall, long read sequencing identified a large number of novel *DMD* isoforms in wild type and dystrophic human muscle. These results provide a reference of known *DMD* transcripts for use in the development of splice modulating therapies currently in development for Duchenne as well as helping understand the role of dystrophin in muscle biology.

2022W

Somatic activating mutations in *MAP2K1* cause melorheostosis. J.C. Marini¹, H. Kang¹, S. Jha¹, Z. Deng², N. Fratzl-Zelman³, W.A. Cabral¹, A. Ivovic², F. Meylan², E.P. Hanson², E. Lange², J. Katz², P. Roschger², K. Klaushofer², E.W. Cowen², R.M. Siegel², T. Bhattacharyya². 1) National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD; 2) National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, MD; 3) Ludwig Boltzmann Institute of Osteology at the Hanusch Hospital of WGKK, and AUVA Trauma Center Meidling, 1st Medical Department Hanusch Hospital, Vienna, Austria.

Melorheostosis is a benign bone overgrowth condition, with “dripping candle wax” appearance on X-rays. Because melorheostosis occurs sporadically, somatic mutations were hypothesized. We investigated bone tissue with paired biopsies of affected and contralateral unaffected bone of 15 melorheostosis patients. The bone samples were utilized for DNA extraction, osteoblast cultures and histology. Using WES and ddPCR, we identified somatic mutations in *MAP2K1*, encoding the kinase MEK1, in affected, but not unaffected, bone of 8 melorheostosis patients. The variant allele frequency ranged from 3-34%. We found no genomic mutations in *LEMD3* or MAPK-related genes that would support a two-hit mechanism. Mosaicism was shown by immunohistochemistry of p-ERK1/2 in bone tissue and two populations of cultured osteoblasts with distinct p-ERK1/2 levels on flow cytometry. All MEK1 mutations occur in two adjacent residues (p.Q56 and p.K57) in the Negative Regulatory Domain, and produce normal levels of mutant transcripts and protein, supporting an activating mechanism. Osteoblasts from affected bone displayed increased MEK1-ERK1/2 signaling and faster proliferation than cells from unaffected bone. The histology of melorheostotic bone featured parallel layers of lamellar bone on the surface, followed by intense remodeling in the interior of the lesions, with increased osteoblasts and osteoclasts. The elevated osteoclast number reflected the increased RANKL/OPG ratio in affected osteoblasts, stimulating osteoclast development. Also, BMP2-stimulated mineralization was significantly suppressed in cultured affected osteoblasts, explaining the 50-fold increase in unmineralized bone matrix (osteoid per bone surface) seen in the interior of the bone lesions, and reduced osteoblast *RUNX2*, *COL1A1* and *ALPL* expression. Melorheostosis provides the first demonstration that the *MAP2K1* oncogene is important to human bone formation. Interestingly, germline mutations in the catalytic core of *MAP2K1* and the negative regulatory domain of *MAP2K1* and *MAP2K2* were identified in Cardio-Facio-Cutaneous syndrome (CFC). CFC and other RASopathies are associated with variable cardiac, facial and neurodevelopmental defects, but not with bone overgrowth. Somatic *MAP2K1* mutations found in melorheostotic osteoblasts and extracranial arteriovenous malformations apparently have consequences distinct from germline mutations. We conclude that outcomes of *MAP2K1* mutations depend on tissue and timing.

2023T

Regulation of *RUNX2* gene during evolution: Functional implications in nonsyndromic craniosynostosis patients. W. Lattanzi¹, M. Barba¹, L. Di Pietro¹, P. Frassanito², G. Tamburrini², C. Prampolini¹, M. Baranzini¹, V. Orticelli¹, S.A. Boyadjiev³, O. Parolini¹, A. Benitez-Burraco⁴, A. Arcovito⁵. 1) Università Cattolica del Sacro Cuore, Roma, Italy; 2) Pediatric Neurosurgery Unit, Fondazione Policlinico Universitario Agostino Gemelli, Rome, Italy; 3) Section of Genetics, Department of Pediatrics, University of California Davis, Sacramento, CA; 4) Dept. of Spanish, Linguistics, and Theory of Literature, Faculty of Philology, Universidad de Sevilla, Sevilla, Spain; 5) Institute of Biochemistry and Clinical Biochemistry, Università Cattolica del Sacro Cuore, Rome, Italy.

Runt related transcription factor 2 (*RUNX2*) encodes the master transcription factor in bone development. Most *RUNX2* mutations described in the literature result in loss-of-function, which causes cleidocranial dysplasia, whereas increased gene dosage is associated with craniosynostosis. *RUNX2* affects indeed skull ossification and is thought to be involved in the different skull morphology of anatomically-modern humans (AMH) compared with extinct species. The aim of our study was to comparatively characterize the genomic structure of *RUNX2* in AMH and extinct hominins, and to infer putative functional correlations for human skull morphogenesis, using nonsyndromic craniosynostosis (NCS) as a disease model. *RUNX2* sequences from extinct hominins (Neanderthal and Denisova) and AMH genomes were compared, and their functional consequences were predicted *in silico*. The expression of *RUNX2* isoforms and of miRNAs binding to the 3'UTRs of the gene, were then analyzed in NCS suture tissues and cells. The affinity of selected miRNAs to specifically interact with the binding site on *RUNX2* mRNA has been evaluated by SPR-based analysis with the Biacore™ X100 biosensor. *RUNX2* counts 11 splice variants, alternatively transcribed from two promoters, and featuring two alternative 3'UTRs. *In silico* analysis allowed detecting all nt changes between AMH and Neanderthal/Denisova, within noncoding regulatory sequences, and predicting the effect of the 3'UTRs sequence changes on miRNA binding. Gene expression analysis of miRNAs and *RUNX2* splice variants during *in vitro* osteogenic differentiation of suture-derived cells highlighted that miRNAs expression was modulated during the ossification process and correlated with *RUNX2* expression. Particularly the isoform containing the second 3'UTR was apparently stabilized by miRNAs binding. SPR-based biomolecular interaction analysis validated the interaction between selected miRNAs and their *RUNX2* mRNA binding site. Our data suggest that *RUNX2* genomic evolution may have affected regulatory sequence of the gene, affecting a specific splice isoform. Changes at the 3'UTR of the AMH gene modified the epigenetic regulation of *RUNX2*, by altering the binding of selected miRNAs and influencing the expression of bone-specific genes plausibly involved in the evolution of skull shape and volume.

2024F

The role for ribosome biogenesis in disorders of craniofacial development. K.E.N. Watt¹, K. Terrazas¹, S. Shannon¹, A. Achilleos¹, P.A. Trainor^{1,2}. 1) Stowers Institute for Medical Research, Kansas City, MO; 2) Department of Anatomy and Cell Biology, University of Kansas Medical Center, Kansas City, KS.

The synthesis of ribosomes is an essential and ubiquitous process for cell survival, growth, and proliferation; however, disruptions in this process result in a group of disorders termed ribosomopathies which display tissue-specific phenotypes. Two examples of ribosomopathies are Acrofacial Dysostosis, Cincinnati Type (AFDCin) and Treacher Collins syndrome (TCS), which exhibit similar facial dysmorphism. AFDCin is also distinguished by various limb anomalies. Pathogenic variants in the RNA Polymerase I (PolI) subunit *POLR1A* are associated with AFDCin, while pathogenic variants in PolI subunits *POLR1C* and *POLR1D*, as well as PolI-associated factor *TCOF1*, underlie the pathogenesis of TCS. Previous studies in zebrafish and mouse models have shown that alterations in these genes result in reduced rDNA transcription, a rate-limiting step of ribosome biogenesis. This leads to both increased cell death and reduced proliferation particularly of neural crest cells, which give rise to the majority of craniofacial cartilage and bone, resulting in cranioskeletal anomalies. In order to fully understand the function of ribosome biogenesis in neural crest cell and skeletal biology, we have conditionally deleted *Polr1a*, *Polr1c*, and *Tcof1* in neural crest cells or at specific times during embryogenesis. These mouse models have revealed that PolI subunits are essential for embryonic survival beyond the blastocyst stage and that there is a specific requirement for PolI genes and associated factors within the neural crest for craniofacial development and embryo survival past mid-gestation. Furthermore, temporal excision of these genes indicates that different tissues may have different threshold requirements for ribosome biogenesis, as some tissues, including cartilage and bone, are more specifically perturbed upon removal of PolI genes. In conclusion, these mouse models provide new information on the role of PolI specifically in neural crest cell and skeletal development. This is important for understanding the role for ribosome biogenesis during normal craniofacial development, and will help in identifying mechanisms underlying the pathogenesis and possible prevention of craniofacial anomalies arising from perturbations in ribosome biogenesis.

2025W

ER-stress involvement in the pathogenic mechanism of Spinocerebellar Ataxia 38 (SCA38). E. Di Gregorio¹, M. Marta², M. Manes³, E. Hoxha⁴, C. Costanz⁵, D. Tripathy⁶, S. Cavalieri⁷, E. Giorgio⁸, C. Mancini², E. Pozzi², E. Riberi⁸, E. Chierto², N. Mitro², D. Caruso², M. Basso³, M. Sallase⁸, F. Tempia⁴, B. Borroni³, A. Brusco^{1,2}. 1) Medical Genetics Unit, Città della Salute e della Scienza University Hospital, Turin, Italy; 2) Department of Medical Sciences, University of Turin, Turin, Italy; 3) Neurology Unit, Department of Clinical and Experimental Sciences, University of Brescia, Brescia, Italy; 4) Neuroscience Institute Cavalieri Ottolenghi, Orbassano, Italy; 5) Centre for Integrative Biology, University of Trento, Trento, Italy; 6) University of Turin, Department of Public Health and Pediatrics, Turin, Italy; 7) Department of Pharmacological and Biomolecular Sciences, University of Milan, Milan, Italy; 8) Department of Medical, Oral and Biotechnological Sciences, University G. D'Annunzio, Chieti, Italy.

ELOVL5 gene is associated with autosomal dominant Spinocerebellar Ataxia 38 (SCA38, MIM#611805), a rare adult-onset cerebellar neurodegeneration. This gene encodes for an elongase, an enzyme localized in the endoplasmic reticulum (ER) where it is involved in the synthesis of a subset of polyunsaturated fatty acids. We explored pathogenic mechanism of SCA38, studying aberrant ELOVL5-p.Gly230Val protein. We demonstrated a subcellular mislocalization in perinuclear area of aberrant protein in different cellular models. Based on these, we hypothesized p.Gly230Val-ELOVL5 is a misfolded protein able to activate the cellular unfolded protein response (UPR). Supporting that idea, we showed a significant increase of ELOVL5 protein in SCA38 fibroblasts after a treatment with the proteasome inhibitor MG-132. In COS7 cells stably expressing p.Gly230Val ELOVL5 we demonstrated the activation of ER-stress response by a significant increase of UPR markers CHOP, ATF-4 and XBP1 and an alteration of ER homeostasis by a slow-down protein transport from ER to the Golgi. Moreover, the use of the chemical chaperone PBA, acting on unfolded p.Gly230Val-ELOVL5 in COS7 cells, led to a physiological ER relocalization of the protein. To determine whether the activation of UPR was associated with neuronal degeneration, we measured cell viability of primary cortical neurons overexpressing wild type or aberrant ELOVL5 and we demonstrated a slightly increased of cells death in p.Gly230Val ELOVL5 neurons. In conclusion, our results support a role for altered ER-stress response in SCA38 pathogenesis, suggesting chemical chaperones might be useful in the treatment.

2026T

Sacs R272C missense homozygous mice develop a milder ataxia phenotype than knock-out mice. R. Larivière¹, N. Sgarioto¹, B. Toscano Marquez², R. Gaudet³, K. Choquet³, R.A. McKinney⁴, A.J. Watt², B. Brais¹. 1) Department of Neurology and Neurosurgery, McGill University, Montreal Neurological Institute, Montreal, Qc, Canada; 2) Department of Biology, McGill University, Montreal, Qc, Canada; 3) Department of Human Genetics, McGill University, Montreal Neurological Institute, Montreal, Qc, Canada; 4) Department of Pharmacology and Therapeutics, McGill University, Montreal, Qc, Canada.

Autosomal recessive spastic ataxia of Charlevoix-Saguenay (ARSACS [MIM 270550]) is an early-onset neurodegenerative disorder caused by mutations in the SACS gene. Over 200 SACS mutations worldwide have now been identified. Most mutations lead to a complete loss of saccin, a large 520 kD protein, although some missense mutations are associated with low levels of saccin expression. We previously showed that Sacs knock-out mice demonstrate early-onset ataxic phenotype with neurofilament bundling in many neuronal populations. To determine if preservation of some mutated saccin protein resulted in the same cellular and behavioral alterations, we generated mice expressing the R272C missense mutation, a homozygous mutation found in some affected patients. Sacs^{R272C} mice express 21% of wild-type brain saccin and display similar abnormalities to Sacs knock-out mice: they have reduced Purkinje cell firing rate, exhibit somatodendritic neurofilament bundles in neurons, and develop a similar, but milder ataxic phenotype. These results support the hypothesis that low levels of saccin, even if mutated, mitigates the severity of the clinical phenotype.

2027F

Characterization of STUB1 in zebrafish: Development of a new knock-out model to study neurodegeneration. Y. Pakdaman^{1,2}, S. Johansson^{1,2}, I. Aukrust^{1,2}, P.M. Knappskog^{1,2}, S. Ellingsen³. 1) Clinical Science Dept, University of Bergen, Bergen, Norway; 2) Medical Genetics Dept, University of Bergen, Bergen, Norway; 3) Biology Dept, University of Bergen, Bergen, Norway.

Spinocerebellar ataxia-16 (SCAR16) is an autosomal recessive neurodegeneration disorder that develops due to deficiencies in *STUB1* gene encoding the ubiquitin E3 ligase and dimeric co-chaperone CHIP. There is limited knowledge regarding the pathogenic role of mutant CHIP *in vivo* and mechanisms underlying SCAR16. In this project, we aim to study the zebrafish *STUB1* orthologue (*stub1*) by CRISPR/Cas9 mediated knock-out of the *stub1* gene. Single guide RNAs (sgRNAs) targeting a 20 nucleotide-sequence in the last exon of *stub1* gene were constructed and injected together with Cas9 proteins into the 1-cell stage of fertilized wild type zebrafish embryos. High resolution melting (HRM) and MluI-restriction digest of the target sites were used to calculate the frequency of somatic mutations after injection. All injected eggs showed a relatively high frequency of modifications at the *stub1* locus (33-100%) at 48 hours post injection (hpi). Sequencing data showed a wide range of insertions and deletions (indels) within the targeted locus. Analysis of F1 siblings was performed to investigate the frequency of germ-line transmission of *stub1* mutations. One founder showed 100% mutation frequency of offspring and was used to produce F1 heterozygous carriers for the *stub1* mutation with 7nt deletions at position 10800 in the *stub1* gene. F1 siblings from the selected founder were raised and identified carriers of the *stub1* mutation were mated in order to obtain homozygous offspring. So far, we have not detected major differences in viability and morphology of carriers compared to the wild-type, however, further studies are required in order to investigate neurobehavioral and physiological consequences of CHIP deficiency in these fishes. Our findings provide the first evidence of customizing CRISPR/Cas9 system to induce modifications and very likely knock out of the *stub1* gene in zebrafish and contribute to establishing a new animal model for studying *STUB1*-related ataxia disease. Future work will focus on phenotypic analysis of homozygous F2 fish. The homozygous fish will be analyzed for the main phenotypes found in humans with mutation in *STUB1* such as ataxia, movement abnormalities and atrophy of different regions of the brain.

2028W

High throughput screen (HTS) using ARSACS fibroblast cytoskeletal bundling assay. *N. Sgarioto¹, R. Lariviere¹, J.P. Chapple², B. Brais¹.* 1) Neurology and Neurosurgery, McGill University, Montreal Neurological Institute Montreal, Quebec, Canada; 2) William Harvey Research Institute, Barts and the London School of Medicine, Queen Mary University of London, London, United Kingdom.

Autosomal recessive spastic ataxia of Charlevoix-Saguenay (ARSACS) is a Quebec founder disease caused by mutations in the SACS gene. Our data show that the cytoskeletal defects first observed in *Sacs*^{-/-} mice and patient brains, are also found in cultured fibroblasts derived from ARSACS patient skin biopsies (HDF). In both primary and immortalized HDF, vimentin intermediate filaments form in the great majority of cells perinuclear bundles. To search for therapeutic molecules we developed a cellular model for high throughput screen (HTS) by expressing a fluorescent-tagged vimentin protein in patient HDFs. Using high-resolution confocal plate reader, we are in the process of screening 20 000 molecules. Our first HTS of 2 000 FDA-approved drugs has uncovered 8 active compounds that had the potential to reverse the ARSACS cytoskeletal phenotype. Validation on 6 different patients HDF carrying 4 different mutations as well as a CRISPR/Cas9 SACS knockout line is under way. The identification of compounds that may influence cytoskeletal bundling formation serve as proof of concept that further HTS screens using this model make uncover therapeutic molecules for ARSACS.

2029T

Disruption of *Tmem30a* results in cerebellar ataxia and degeneration of Purkinje cells. *X. Zhu^{1,2}, Y. Yang².* 1) Sichuan Academy of Medical Sciences and Sichuan People's Hospital, Chengdu, Sichuan, China; 2) School of Medicine, University of Electronic Science and Technology of China, Chengdu, Sichuan, China.

Mutations in phosphatidylserine (PS) flippase *ATP8A2* cause cerebellar atrophy, mental retardation and disequilibrium syndrome. The proper transport and function of *ATP8A2* require a β -subunit transmembrane protein 30A (TME-M30A). However, the roles of *TMEM30A* in the central nervous system remain elusive. To investigate the role of *Tmem30a* in the cerebellum, we developed a *Tmem30a* Purkinje cell (PC)-specific knockout (KO) mouse model. The *Tmem30a* KO mice displayed early-onset ataxia and progressive PC death. Deficiency in *Tmem30a* led to an increased expression of Glial fibrillary acidic protein (GFAP) and astrogliosis in regions with PC loss. Elevated C/EBP homologous protein (CHOP) and BiP expression levels indicated the presence of endoplasmic reticulum (ER) stress in the PCs prior to visible cell loss. The TUNEL analysis suggested that apoptotic cell death occurred in the cerebellum. Our data demonstrate that loss of *Tmem30a* in PCs results in protein folding and transport defects, a substantial decrease in dendritic spine density, increased astrogliosis and PC death. Altogether, our data demonstrate the essential role of *Tmem30a* in the cerebellum.

2030F

Neuroprotective efficacy of *Bacopa monnieri* (Bramhi) extract: Molecular approach in Parkinson's disease models: *In vivo* and *in vitro* study. B. Singh^{1,2}, S. Pandey¹, R. Verma², A.A. Mahdi³. 1) Biochemistry, King George's Medical University, Lucknow, India; 2) Neurology, King George's Medical University, Lucknow; 3) Era University, Lucknow.

Parkinson's disease is a second most common neurodegenerative disease, with the selective death of dopaminergic neurons and motor impairment. The disease has no absolute cure yet; all the therapeutic strategies involve prescription of dopamine agonist drugs which turn ineffective after prolonged use. We previously reported the propensity of *Bacopa monnieri* (BM) extract against oxidative stress markers and neurogenic parameters at a molecular level in the brain of MPTP treated mice. In this study, we tested the hypothesis whether pretreatment or post-treatment of BM could provide better neuroprotection against rotenone-induced neuroinflammation under in vitro and in-vivo conditions in the different region of brain tissues. BM exhibited multiple free radical scavenging and anti-inflammatory capacity in-vitro study. Pretreatment of neuroblastoma SH-SY5Y cells with BM not only abrogated the generation of inflammation but also offered significant marked protection against rotenone mediated cytotoxicity as compared to post-treatment. These findings were further validated employing rotenone rat model through in-vivo study. BM pretreatment completely abolished rotenone-induced inflammatory response in the different region of brain substantia nigra, striatum, hippocampus as compared to standard control treatment L-dopa. We determined the degree of oxidative stress induction, redox status, enzymatic antioxidants and dopamine function in various brain regions of PD rat model provided with BM for one month followed by rotenone challenge (2mg/kg body weight) subcutaneous last four weeks via RT-PCR analysis. BM prophylaxis completely prevented rotenone-induced oxidative dysfunctions and inflammation. Interestingly, BM prophylaxis also prevented the aggregation of alpha-synuclein. Collectively these findings provide evidence on the significant prophylactic neuroprotective efficacy of BM in in-vitro as well as in-vivo study. Based on these data, it is hypothesized that BM can serve as a useful adjuvant in protecting the brain against oxidative-mediated neurodegenerative disorders.

2031W

Neurobehavioral issues in the phenotypic spectrum of variations in OTX2. A. Kumar, G. Mainali, M. Yelton, G. Raymond. Penn State Milton S Hershey Medical Center, Hershey, PA.

Introduction: The majority of pathogenic variants in the OTX2 gene have been reported in the patients with ocular malformation, pituitary dysfunction with growth hormone deficiency and rarely patients with neurological issues such as developmental delay, learning disabilities, behavioral problems or seizure. We describe a 7-year-old boy with varied neurological presentation found to have two variants in OTX2 which may be related to his phenotype. **Case Presentation:** The proband initially presented at 5 years of age for evaluation of periodic dystonia and ataxia. His birth history and family history was largely unknown as he was adopted but he had complex medical history including developmental delay, failure to thrive, tremors, hypotonia, recurrent vomiting, OSA, periodic dystonia and ataxia. On examination he was short, developmentally delayed with autistic features, dysarthric, appeared anxious with perseveration, hypotonia, titubation, unsteadiness and tremors. MRI of his brain showed thinning of splenium of the corpus callosum, hypoplasia of pituitary gland with ectopic posterior pituitary and a Dandy-Walker variant with prominent fourth ventricle and cisterna magna. 16 gene dystonia panel was negative. Whole exome sequencing showed two variants in OTX2 in cis, p.L61R and p.R69W. He is presently doing well on oxcarbazepine and lamotrigine for dystonia and behavioral issues. **Discussion:** The OTX2 gene, a bicoid-type homeodomain gene, is implicated in the development of brain, eye, and the pituitary. Most of the patients reported have ocular abnormalities most commonly anophthalmia-microphthalmia, short stature due to growth failure from pituitary dysfunction and other neurological abnormalities. Our patient had no eye abnormality but he had pituitary hypoplasia and multiple neurological problems including hypotonia, dystonia, ataxia which has not been reported in the literature. His WES showed two cis-variants in OTX2 p.L61R and p.R69W, that is possibly related to his phenotype. These variants have not previously been reported but are predicted to have a detrimental effect. We are reporting a single individual with intellectual disability, autism, and paroxysmal movement disorder potentially expanding the phenotype seen in variations in OTX2.

2032T

Functional modeling of NMIHBA-causing PRUNE1 variants reveals a requirement for its exopolyphosphatase activity. *H. Nistala¹, J. Dronzek¹, S. Rajamani², S. Nuwayhid², D. Delgado², E. Burke³, C. Gonzaga-Jauregi¹, M. Franklin², C. Schoenherr^{1,2}, P. Sarangapani², M.C. Malicdan³, J.R. Lupski¹, W.K. Chung², A.N. Economides^{1,2}.* 1) Regeneron Genetics Center, Tarrytown, NY; 2) Regeneron Pharmaceuticals Inc., Tarrytown, NY; 3) Undiagnosed Diseases Program Translational Laboratory, NHGRI, National Institutes of Health, Bethesda, MD; 4) Department of Human Genetics, Baylor College of Medicine, Houston, TX; 5) Columbia University Irving Medical Center, New York NY.

Neurodevelopmental disorder with microcephaly, hypotonia, and variable brain anomalies (NMIHBA, OMIM# 617481) is an autosomal recessive disorder caused by variants in the prune exopolyphosphatase 1 gene (*PRUNE1*). NMIHBA is a debilitating neurodevelopmental and neurodegenerative disorder characterized by global developmental delay and intellectual disability. Microcephaly, progressive cortical atrophy, cerebellar hypoplasia and delayed myelination are neurological hallmarks of affected individuals. In this report, we provide clinical description of two patients harboring compound heterozygous variants c.G383A (p.Arg128Gln), c.G520T (p.Gly174*) in *PRUNE1*. Furthermore, we demonstrate biochemically that missense variants within the conserved N-terminal aspartic acid-histidine-histidine (DHH) motif result in the destabilization of protein structure and consequent loss of exopolyphosphatase activity, leading to *PRUNE1* loss-of-function. Lastly, we demonstrate that loss of *Prune1* results in embryonic lethality in mouse secondary to structural brain malformations. Taken together, these studies provide a molecular definition of the pathogenesis underlying NMIHBA while underscoring the critical role of *PRUNE1* exopolyphosphatase activity in neurodevelopment.

2033F

CEDNIK syndrome, a rare neuro-cutaneous disorder. *H.K. Tiwana, G.V. Raymond, A. Johri, A. Kumar.* Penn State Milton S. Hershey Medical Center, Hershey, PA.

Introduction: CEDNIK (cerebral dysgenesis, neuropathy, ichthyosis and keratoderma) syndrome is a rare autosomal recessive neuro-cutaneous disorder characterized by mutation in Synaptosomal-associated protein 29 (*SNAP29*) gene. We present a patient with alteration in *SNAP29* and present the clinical, radiographic, and discuss the role of *SNAP29* in development. Case Presentation: A 26-month old male born to non-consanguineous parents without any prenatal complications, had severe retrognathia at birth requiring G-tube placement and mandibular enhancement procedure. He had severe developmental delay with inability to sit, no speech development, severe cognitive issues and seizures developed at 9 months. Examination showed bilateral epicanthic folds with mid-facial hypoplasia, generalized ichthyosis and diffuse hypotonia. MRI showed multiple brain malformations including partial agenesis of corpus callosum and diffuse polymicrogyria. A sequencing panel showed two mutations, a homozygous pathogenic frame shifts sequence change in the *SNAP29* gene, associated with the CEDNIK syndrome. He has continued to have profound developmental delay and at 2 years is functioning in the 4-6 month range. CONCLUSION: Our patient's DNA analysis of *SNAP29* gene demonstrated a homozygous one base pair duplication in exon 2, c.354dup. This pathogenic sequence change results in an amino acid frameshift and creates a premature stop codon 14 amino acids downstream of the mutation, p.Leu119Alafs*15. This pathogenic change is predicted to result in an abnormal transcript, causing truncated *SNAP29* protein with abnormal function. *SNAP* proteins are t-SNARE proteins that function in intracellular vesicle fusion. *SNAP29* has been determined to bind to a broad range of syntaxin fusion proteins and is involved in a number of intracellular transport steps and would impact a variety of neuroectodermal tissues. Recent study suggest that *SNAP* has additional role as a negative modulator of neurotransmitter release by slowing the recycling of the SNARE-associated fusion machinery and synaptic vesicle turnover. This may be related to the severe neurologic manifestations shown by patients with CEDNIK syndrome.

2034W

CSF1R allelic and biallelic loss-of-function lead to microglia loss in adult and respectively their complete absence in congenital brain disease. T.J. van Ham¹, N. Oosterhof¹, I.J. Chang², L.E. Kuil³, D. Jensen³, R. Daza⁴, L. Astle⁵, H. van der Linde¹, C. Latimer⁶, D. Keene⁶, E. Loter⁷, W.F.J. van IJcken⁸, J.A.A. Demmers⁹, J.C. van Swieten^{10,11}, M.H.G. Verheijen¹², R.F. Hevner^{1,6}, J.T. Bennett^{2,3,7}. 1) Clinical Genetics, Erasmus University Medical Center, Rotterdam, Netherlands; 2) Department of Pediatrics, Division of Genetic Medicine, University of Washington School of Medicine, Seattle, WA, USA; 3) Center for Developmental Biology and Regenerative Medicine, Seattle Children's Research Institute, Seattle, WA, USA; 4) Center for Integrative Brain Research, Seattle Children's Research Institute, Seattle, WA, USA; 5) Department of Laboratory and Pathology, Alaska Native Tribal Health Consortium, Anchorage, AK, USA; 6) Department of Pathology, University of Washington School of Medicine, Seattle WA USA; 7) Department of Laboratories, Seattle Children's Hospital, Seattle, WA, USA; 8) Center for Biomimics, Erasmus Medical Center, Rotterdam, the Netherlands; 9) Proteomics Center, Erasmus University Medical Center, Rotterdam, the Netherlands; 10) Department of Neurology, Erasmus Medical Center, Rotterdam, the Netherlands; 11) Department of Clinical Genetics, VU Medical Center, Amsterdam, the Netherlands; 12) Department of Molecular and Cellular Neurobiology, CNCR, Amsterdam Neuroscience, VU University, Amsterdam, the Netherlands.

Haploinsufficiency of *CSF1R*, a key regulator of microglia, brain resident macrophages, causes adult onset leukoencephalopathy with axonal spheroids (ALSP). ALSP is a rapidly progressive disorder and is considered a microgliopathy, as it is thought that microglial dysfunction is the primary cause. Apart from phagocytosis, microglia functions in the human brain are poorly understood and it is unknown how *CSF1R* variants affect microglia and lead to disease. Here, we report the first neuropathological description of an infant whose brain contained almost no microglia, caused by homozygous pathogenic variants in *CSF1R*. This patient had structural brain anomalies, including agenesis of the corpus callosum (ACC). Neuropathology revealed calcifications, loss of white matter and axonal spheroids, which are also characteristic for ALSP. To investigate underlying mechanisms, we generated zebrafish deficient in both *CSF1R* homologs (*csf1^{pm}*). We found that, in *csf1^{pm}* zebrafish, microglia were almost absent, but also that haploinsufficiency caused aberrant microglia density and local loss of microglia. We also observed lower microglia numbers and even widespread microglia depletion in post mortem brain tissue of ALSP patients, suggesting loss of microglia may contribute to ALSP pathogenesis. To investigate the consequences of a lack of microglia, we performed brain proteome analysis in control and *csf1^{pm}* brain, which identified Cux1a as a highly downregulated protein in *csf1^{pm}* brain. CUX1 is expressed in, and regulates, neurons forming the corpus callosum in mice and we found that CUX1-expressing neurons were reduced in the *CSF1R* homozygous infant. This suggests that there is a conserved, possibly trophic role of microglia affecting CUX1-neurons. As white matter diseases caused by variants in other genes, e.g. *TREM2* and *TYROBP*, are also considered to be microgliopathies, this implies that microglia have essential roles in human white matter. Phenotypic similarities between *CSF1R* haploinsufficiency and biallelic loss-of-function variants, imply that related mechanisms, likely functional loss of microglia, cause severe abnormalities in the embryonic and adult brain. This suggests that having too few microglia is severely detrimental to the adult and developing human brain.

2035T

Loss of function mutations in *EPHB4* are responsible for vein of Galen aneurysmal malformation. A. Vivanti¹, A. Ozanne², C. Grondin¹, G. Saliou³, L. Quevarec⁴, H. Maurey⁴, P. Aubourg⁴, A. Benaché⁵, M. Gut⁶, I. Gut⁶, J. Martinovic⁷, M.V. Senat⁸, M. Tawak⁹, J. Melki¹. 1) UMR-1169 and University Paris Sud, Inserm, Le Kremlin-Bicêtre, Ile de France, France; 2) Department of Neuroradiology, Assistance publique des Hôpitaux de Paris, Hôpital Bicêtre, Le Kremlin-Bicêtre, 94276, France; 3) Department of Diagnostic and Interventional Radiology, Centre Hospitalier Universitaire Vaudois, Lausanne University Hospital, Lausanne, CH-1011, Suisse; 4) Department of Paediatric Neurology, Hôpital Bicêtre, Assistance publique des Hôpitaux de Paris, Le Kremlin-Bicêtre, 94276, France; 5) Department of Obstetrics and Gynecology, Hôpital Antoine-Béclère, Assistance publique des Hôpitaux de Paris, 92140, Clamart, France; 6) CNAG-CRG, Centre for Genomic Regulation, Barcelona Institute of Science and Technology, 08028, Barcelona, Spain; 7) Unit of Fetal Pathology, Hôpital Antoine-Béclère, Assistance publique des Hôpitaux de Paris, 92140, Clamart, France; 8) Department of Obstetrics and Gynecology, Hôpital Bicêtre, Assistance Publique des Hôpitaux de Paris, Le Kremlin-Bicêtre, 94276, France; 9) INSERM UMR-1195, University Paris Sud, Le Kremlin-Bicêtre, 94276, France.

Vein of Galen aneurysmal malformation (VGAM) is one of the most common fetal brain vascular malformations. This results from the persistence of the embryonic prosencephalic vein of Markowski and may lead to congestive heart failure and pulmonary hypertension in neonates, macrocrania, ventriculomegaly, seizures, mental retardation, haemorrhage or venous ischaemia. Clinically, VGAM cases appear to be sporadic. We conducted whole exome sequencing in 19 unrelated VGAM patients and screened candidate gene in a cohort of 32 additional patients using either targeted exome or Sanger sequencing. In a cohort of 51 VGAM patients, we found 5 affected individuals with heterozygous mutations in *EPHB4* including *de novo* frameshift (p.His191Alafs*32) or inherited deleterious splice or missense mutations predicted to be pathogenic by *in silico* tools. Knockdown of *EPHB4* in zebrafish embryos leads to specific anomalies of dorsal cranial vessels including dorsal longitudinal vein, the ortholog of the median prosencephalic vein, the embryonic precursor of the vein of Galen. This model allowed to demonstrate *EPHB4* loss of function mutations in VGAM by the ability to rescue the brain vascular defect in knockdown zebrafish co-injected with wild type but not truncated *EPHB4* mimicking the p.His191Alafs mutation. Our data showed that in both species, loss of function mutations of *EPHB4* result in specific and similar brain vascular development anomalies (Vivanti et al. 2018). Recently, *EPHB4* germline mutations have been reported in non-immune hydrops fetalis and in cutaneous capillary malformation–arteriovenous malformation. Here, we show that *EPHB4* mutations are also responsible for VGAM indicating that heterozygous germline mutations of *EPHB4* result in a large clinical spectrum. The identification of *EPHB4* pathogenic mutation in patients presenting capillary malformation or VGAM should lead to careful follow up of pregnancy of carriers for early detection of VGAM in order to propose optimal neonatal care. Endovascular embolization indeed greatly improved the prognosis of VGAM patients.

2036F

Neuron-specific activation of a patient mutation reveals the prominent role of forebrain excitatory neurons in *SCN8A* epileptic encephalopathy. J.L. Waggon¹, R.K.A. Bunton-Stashyshyn¹, A. Faulkner¹, E.R. Wengert², P. Wagley³, Y. Yuan⁴, K. Bhatia⁵, J.M. Jones¹, L.L. Isom¹, J.M. Parent⁵, H.P. Goodkin¹, M.K. Patel¹, M.H. Meisler^{1,5}. 1) Human Genetics, University of Michigan, Ann Arbor, MI; 2) Anesthesiology, University of Virginia, Charlottesville, VA; 3) Neurology, University of Virginia Health System, Charlottesville, VA; 4) Pharmacology, University of Michigan, Ann Arbor, MI; 5) Neurology, University of Michigan, Ann Arbor, MI.

De novo mutations of the sodium channel gene *SCN8A* cause severe epileptic encephalopathy with early seizure onset, developmental delay, movement disorders, and elevated risk of sudden death. A previously described knock-in mouse model carrying the patient mutation p.Asn1768Asp (N1768D) recapitulated the seizures and sudden death characteristic of the disorder. To define the contributions of specific neuronal populations to the seizure phenotype, we generated a conditional allele of *Scn8a* that can be activated by Cre recombinase. We chose the recurrent *de novo* mutation p.Arg1872Trp (R1872W), which has been seen in 14 unrelated individuals with severe *SCN8A* encephalopathy. We used TALENs and a targeting vector to insert two copies of the final *Scn8a* exon 26 into the endogenous locus. The upstream copy of exon 26 is flanked by loxP sites and encodes wildtype Arg1872. Upon exposure to CRE, the upstream copy is removed and expression of the downstream exon containing the R1872W mutation is activated. We crossed mice with the conditional *Scn8a* allele with mice expressing Ella-Cre to generate heterozygous mice with global, constitutive expression of R1872W. We also used Nestin-Cre (pan-neural), Emx1-Cre (excitatory neurons), and Gad2-Cre (inhibitory neurons) to evaluate contributions of neuronal subpopulations to *SCN8A* encephalopathy. We evaluated age of seizure onset, survival, video-EEG, and neuronal firing in brain slice recordings. Global activation of R1872W with Ella-cre resulted in seizure onset at 2 weeks of age followed by death within 24 hours, often immediately after the first seizure. Brain slice recordings revealed sustained spontaneous firing in cortex and hippocampus indicative of extreme neuronal hyperexcitability. Pan-neural activation by Nestin-Cre also resulted in early seizure onset and sudden death. Activation of R1872W in excitatory neurons by Emx1-Cre was sufficient to cause seizures and sudden death, with onset between 3 weeks and 2 months of age. In contrast, activation of R1872W in inhibitory neurons with Gad2-Cre did not induce seizures, neurological dysfunction or abnormal EEG. These findings provide insight into the pathogenic mechanism of gain-of-function mutations of *SCN8A*. The work underscores the differences between *SCN8A* encephalopathy and Dravet Syndrome caused by haploinsufficiency of *SCN1A* in inhibitory neurons. Our observations identify excitatory neurons as critical targets for therapeutic intervention.

2037W

Motor and metabolic impairments in a *Mecp2-e1* mutant mouse model of Rett syndrome. D.H. Yasui¹, A. Vogel Ciernia¹, T.E. Grant¹, B.P. Durbin-Johnson², A.B. Noronha¹, A.R. Chang¹, J.N. Crawley³, J.M. LaSalle⁴. 1) Department of Medical Microbiology and Immunology, UC Davis, CA School of Medicine, 95616, USA; 2) Division of Biostatistics, Department of Public Health Sciences, UC Davis School of Medicine, Sacramento, CA, 95817, USA; 3) Robert E. Chason Endowed Chair in Translational Research, MIND Institute Distinguished Professor, Department of Psychiatry and Behavioral Sciences University of California Davis School of Medicine Sacramento, CA 95817, USA; 4) Department of Medical Microbiology and Immunology, Genome Center, MIND Institute, UC Davis School of Medicine, CA, 95616, USA.

Rett syndrome (RTT) is an X linked neurodevelopmental disorder that predominantly occurs in females. RTT, caused by mutations in the *MECP2* gene encoding the MeCP2-e1 and MeCP2-e2 isoforms, is diagnosed within the first 24-36 months of life and presents early as loss of speech, mobility, and purposeful hand movements followed by breathing abnormalities, seizures and gastrointestinal symptoms in later disease stages. To model the RTT disease process and identify potential treatments we have recapitulated an *MECP2* mutation from a RTT patient in a mouse model. This model establishes that loss of the MeCP2-e1 isoform and not the MeCP2-e2 isoform underlies the RTT phenotype as *Mecp2-e1*^{-/-} deficient mice have severe motor and behavioral deficits and die by 30 weeks of age. For the current studies we have expanded behavioral and motor characterization of *Mecp2-e1*^{-/-} females and *Mecp2-e1*^{-/-} males by performing repeated analysis of motor tasks using rotarod, beam walking and gait assays throughout the disease process up to 18 weeks of age. *Mecp2-e1*^{-/-} females and *Mecp2-e1*^{-/-} males have significant increases (p<0.05) in hind limb clasping while *Mecp2-e1*^{-/-} females show increased beam walking times (p<0.05), along with reduced latency to fall in rotarod and altered gait (p<0.05). In an independent cohort of early disease stage *Mecp2-e1* deficient animals, metabolism, sleep, and total activity was determined using Comprehensive Lab Animal Monitoring System analysis, while body composition was assayed by Dual-energy X-ray Absorptiometry and untargeted liver lipid profiling was performed by mass spectrometric analysis. *Mecp2-e1*^{-/-} mice show a significant elevation of respiratory exchange rate (p=0.0368), significant reductions in energy intake (p=0.0174) and activity (p=0.023) along with reduced sleep (p=0.0085) and elevated percent body fat (p=0.0096). Analyses of *Mecp2-e1*^{-/-} females revealed significantly elevated body fat content (p=0.0183) along with elevated liver cholesterol esters 204 and 181 and triglycerides 502, 568 and 589. The identification of early metabolic alterations, body composition changes and sleep defects in *Mecp2-e1*^{-/-} male mice along with body composition and motor defects in *Mecp2-e1*^{-/-} females suggests novel mechanisms underlying disease progression in RTT model mice. Identification of these early phenotypic effects may lead to development of potential RTT translatable therapies.

2038T

Role of the DUF1220 protein domain in neurogenesis. E.A. Werren^{1,2}, A.W. Bigham², S.L. Bielas¹. 1) Human Genetics, University of Michigan, Ann Arbor, MI, USA; 2) Anthropology, University of Michigan, Ann Arbor, MI, USA.

Recurrent, reciprocal genomic disorders resulting from segmental duplications in humans are a major cause of disease and a source for highly plastic genomic regions subject to human-lineage specific evolution. The DUF1220 (Olduvai) protein domain has undergone a significant expansion in copy number in the human lineage. This protein domain is encoded by the neuroblastoma breakpoint family (NBPF) genes, located primarily on chromosome 1 and enriched in the 1q21 region. DUF1220 can be sub-classified into six clades – three of which are conserved across primates, and three are human-lineage specific (HLS). Interestingly, in NBPF genes located within the 1q21 region, the three DUF1220 HLS domains have undergone extreme amplification. Mutations in 1q21 have been linked to several neurodevelopmental disorders with clinical features including microcephaly, macrocephaly, and autism spectrum disorder. Recent studies have shown that increased copy number of the DUF1220 domain is associated with increased proliferation in human neural stem cells. Likewise, peak DUF1220 expression in early- to mid-fetal brain tissues during development, overlapping with periods of cortical neurogenesis. However, the precise molecular mechanism behind the domain's effect on neural proliferation is unknown. We detect NBPF in neural progenitor cells (NPCs) differentiated from H9 embryonic stem cells. NBPF co-localizes with vesicles and the cleavage furrow at the late stages of cytokinesis, consistent with cell biology described as the molecular basis of primary microcephaly. We will test the hypothesis that the type of DUF1220 expansion and the domain clade contribute to early human brain development using 3D cerebral organoids differentiated from human embryonic stem cells.

2039F

Gabapentin therapy: Improvement in neurologic and behavioral phenotype of patient with GABRB2 and CHAMP1 dual diagnosis. A. Basunia¹, M. Byler¹, N. Havermick², J.A. Bassetti³, H. Cui⁴, C. Eng⁵, R.R. Lebel¹. 1) Section of Medical Genetics, SUNY Upstate Medical University, Syracuse, NY; 2) Department of Neurology, SUNY Upstate Medical University, Syracuse, NY; 3) Department of Pediatrics, Weill Cornell Medicine, New York, NY; 4) School of Information, University of Arizona, Tucson, Az; 5) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

Our female patient was born to a non-consanguineous 28 year old Caucasian father and a 26 year primigravida Caucasian mother at 40 weeks gestation following pregnancy complicated by bleeding in the late second trimester. There were no suspected teratogen exposures and family history revealed no intellectual disability or genetic disorders. She developed severely adverse phenotype including developmental delays, dysmorphic features and hypotonia at 2 months and partial complex seizures at 12 months. Whole-exome analysis revealed a pathogenic *de novo* CHAMP1 frameshift mutation (c.1876_1877del; p.S626fs) as the underlying cause. Levetiracetam was used with beneficial effects. The GABRB2 mutation (c.236T>C; p.M79T) was discovered upon revisiting the database, and Gabapentin was then added to the regimen. Soon thereafter, it was apparent that in addition to excellent seizure control the patient was also demonstrating improved global neurologic function. GABA_A receptors are multi-subunit inhibitory chloride channels of the central nervous system. Mutations in GABA_A genes, such as GABRB2, have all been associated with epilepsy and/or developmental delays. The M79T GABRB2 variant is found in the N-terminal extracellular domain associated with GABA binding, particularly in allosteric changes. Thus, the variant may reduce functionality of GABA_A receptor by hampering fast synaptic inhibition and decreasing the threshold for seizure. Gabapentin modulates the activity of GABA biosynthesis enzymes, increasing GABA concentration in the postsynaptic cleft and thus the probability of GABA - GABA_A receptor interactions. A comparative analysis of the response to various anti-epileptic drugs (AEDs) in children with pathogenic GABRB2 mutations was subsequently performed. While response to AED combination therapies varied markedly even within recurrent *de novo* mutations, Gabapentin was efficacious in alleviating motor symptoms in a 4 year old male patient multifocal seizures and dystonic movements. Until the genomic era, the treatment of seizures in neonates and children was strictly empiric and based on pharmacokinetics, adverse effects and potential drug-drug interactions. Targeting underlying molecular mutations identified through genomic analysis should be therapeutically effective for management of such patients.

2040W

Using zebrafish as a tool to model patient mutations of *KCNB1* and its role in epilepsy. M.Y. Dennis¹, A. Colón-Rodríguez¹, K. Burbach¹, E. Ferino¹, M. Bader¹, P. Lein², L. Jao³. 1) Genome Center, MIND Institute, and Department of Biochemistry & Molecular Medicine, University of California, Davis, CA; 2) Department of Molecular Biosciences, University of California, Davis, CA; 3) Department of Cell Biology and Human Anatomy, University of California, Davis, CA.

Genetic factors responsible for complex neurological disorders, including epilepsy, are not fully understood. *De novo* mutations in *KCNB1* encoding K.2.1 voltage gated potassium channel have been identified in patients with sporadic infantile epilepsy. Nearly all variants identified are missense making their pathogenicity uncertain. *Kcnb1* homozygous null mice exhibit neuronal hyperexcitability, sensitivity to convulsant-induced seizures, defects in spatial learning and motor activity, and anxiety-like behavior. A recent study examining *kcnb1* null mutants in zebrafish resulted in a reduced ventricular system and gain of function caused hydrocephalus; however, additional phenotypic and neuronal alterations were not examined. Zebrafish are a compelling organism to model neurodevelopmental disorders due to their considerable orthology of genes (~70%) and synaptic function with humans. The relative ease to gene edit and produce great numbers of offspring allow for higher-throughput experiments compared with mouse. Using gene editing, we produced knockout alleles of *kcnb1* and performed subsequent morphological and behavioral analyses in zebrafish larvae. Our results show no significant differences between mutant zebrafish carrying one (heterozygote) or two (homozygous) null alleles compared to wildtype fish in body or head size. However, there was a significant increase in activity (distance moved) between heterozygous and homozygous fish in the presence of PTZ, a GABA_A antagonist, recapitulating phenotypes observed previously in *Kcnb1* null mice. Considering the majority of patient mutations are amino-acid substitutions, efforts to introduce missense variants at evolutionarily conserved sites using gene editing are ongoing. The results of this study will lay the groundwork to: (1) identify neurodevelopmental alterations of *kcnb1* mutations in zebrafish; and (2) develop tools that we will continue to use as we move forward to characterize unknown disease-related mutations in a higher-throughput manner.

2041T

PCDH19 regulation of neural progenitor cell differentiation suggests asynchrony of neurogenesis as a mechanism contributing to PCDH19 Girls Clustering Epilepsy. L.A. Jolly^{1,2}, C.C. Homan^{1,2}, S. Pederson¹, T. To¹, C. Tan¹, S. Piltz^{1,3}, M. Corbett^{1,2}, E. Wolvetang⁴, P. Thomas^{1,3}, J. Gecz^{1,2,3}. 1) University of Adelaide, Adelaide, South Australia, Australia; 2) Robinson Research Institute, Adelaide, South Australia, Australia; 3) South Australian Health and Medical Research Institute, Adelaide South Australia, Australia; 4) Australian Institute for Bioengineering and Nanotechnology, The University of Queensland, Queensland, Australia.

PCDH19-Girls Clustering Epilepsy (PCDH19-GCE) is a female restricted childhood epileptic encephalopathy characterised by a spectrum of neurodevelopmental problems. PCDH19-GCE is caused by heterozygous loss-of-function mutations in the X-chromosome gene, *Protocadherin 19* (*PCDH19*) encoding a cell-cell adhesion molecule. Intriguingly, hemizygous males are generally unaffected. As *PCDH19* is subjected to random X-inactivation, heterozygous females are comprised of a mosaic of cells expressing either the normal or mutant allele, which is thought to drive pathology. In alignment, males displaying somatic mosaicism of *PCDH19* mutations also present with the disorder. Despite being the second most prevalent monogenic cause of epilepsy, little is known about the role of PCDH19 in brain development. Here we show that PCDH19 is highly expressed in human neural stem and progenitor cells (NSPCs) and investigate its function *in vitro* in these cells of both mouse and human origin. Transcriptomic analysis of mouse NSPCs lacking *Pcdh19* revealed changes to genes involved in regulation of neuronal differentiation, and we subsequently show that loss of *Pcdh19* causes increased NSPC neurogenesis. We reprogrammed human fibroblast cells harbouring a pathogenic *PCDH19* mutation into human induced pluripotent stem cells (hiPSC) and employed neural differentiation of these to extend our studies into human NSPCs. As in mouse, loss of PCDH19 function caused increased neurogenesis, and furthermore, we show this is associated with a loss of human NSPC apical-basal polarity. Overall our data suggests a conserved role for *PCDH19* in regulating mammalian cortical neurogenesis and has implications for the pathogenesis of PCDH19-GCE. We propose that the difference in timing or 'heterochrony' of neuronal cell production originating from PCDH19 wildtype and mutant NSPCs within the same individual contributes to downstream disruption to neuronal network formation, predisposing the individual to network dysfunction and epileptic activity.

2042F

***In vitro* functional studies of rare glutamate receptor variants define effects on receptor activity in patients.** S.J. Myers, S. Kim, S. Ross, V. Peterson, J. Allen, R. Mizu, H. Yuan, S.F. Traynelis. Center for Functional Evaluation of Rare Variants (CFERV), Department of Pharmacology, Emory University, Atlanta, GA 30322.

Whole exome and genome sequencing has identified numerous rare variants in the glutamate receptor family of *GRIN*, *GRIA*, *GRIK*, and *GRID* genes associated with neurological syndromes and phenotypes such as epilepsy (including infantile spasms, Lennox-Gastaut, Landau-Kleffner, and West syndromes), intellectual disability, developmental delay, language disorders, hypotonia and movement disorders, autism and schizophrenia. We compared specific glutamate receptor gene variant function identified in patients to those identified in a control population in a series of *in vitro* electrophysiological and biochemical studies. Briefly, variant glutamate receptors were expressed in *Xenopus laevis* oocytes and HEK cells and used to measure agonist potency (L-glutamate and glycine EC₅₀), agonist deactivation kinetics, receptor sensitivity to inhibition by the endogenous modulators Mg²⁺, Zn²⁺, and H⁺ (IC₅₀, % inhibition), and cell-surface receptor trafficking. Over 90 specific missense, frameshift, or short indel variants in the *GRIN* gene family (primarily *GRIN1*, *GRIN2A*, and *GRIN2B*) were functionally characterized by CFERV. Approximately 51% of the disease associated variants we evaluated possess altered function in at least one functional parameter (up to 8 parameters/variant measured), with 25% of disease variants with multiple functional parameters altered. Agonist potency for receptor activation (30%) was the most common functional parameter altered in the patient population of variants whereas the potencies for the endogenous negative modulators (pH, Mg²⁺, and Zn²⁺) were altered in 10-20% of these variants. Over twenty disease-associated variants resulted in a significant gain or loss of function of >5 fold shifts compared to the wild-type receptor; however several rare variants altered L-glutamate agonist affinity by over 100 fold. We present data on novel glutamate receptor variants and describe the effect of these mutations on receptor function. The goal of CFERV is to increase our understanding of the relationship between specific genetic variations in the glutamate receptor gene families and health consequences and to use these data to define limits of *in vitro* functional change that drive genetic phenotype.

2043W

De novo hotspot variants in CYFIP2 cause early-onset epileptic encephalopathy. M. Nakashima^{1,2}, M. Kato^{3,4}, K. Aoto¹, S. Masaaki⁵, H. Belal¹, S. Mukaida⁶, S. Kumada⁶, A. Sato^{6,7}, A. Zerem⁸, T. Lerman-Sagie⁹, D. Lev⁹, H.Y. Leong⁹, Y. Tsurusaki², T. Mizuguchi², S. Miyatake², N. Miyake², K. Ogata², H. Saitsu¹, N. Matsumoto². 1) Hamamatsu University School of Medicine, Hamamatsu, Shizuoka, Japan; 2) Yokohama City University Graduate School of Medicine, Yokohama, Japan; 3) Showa University School of Medicine, Tokyo, Japan; 4) Yamagata University Faculty of Medicine, Yamagata, Japan; 5) National Hospital Organization Utano Hospital, Kyoto, Japan; 6) Tokyo Metropolitan Neurological Hospital, Tokyo, Japan; 7) The University of Tokyo Hospital, Tokyo, Japan; 8) Wolfson Medical Center, Tel-Aviv, Israel; 9) Hospital Kuala Lumpur, Kuala Lumpur, Malaysia.

The cytoplasmic fragile X mental retardation 1 interacting proteins 2 (CYFIP2) is a component of the WAVE regulatory complex, which is involved in actin dynamics. An obvious association of CYFIP2 variants with human neurological disorders has never been reported. Here, we identified *de novo* hotspot CYFIP2 variants in neurodevelopmental disorders and explore the possible involvement of the CYFIP2 mutants in the WAVE signaling pathway. Our findings suggest that *de novo* Arg87 variants in CYFIP2 have gain-of-function effects on the WAVE signaling pathway, and are associated with severe neurological disorders. We performed trio-based whole exome sequencing (WES) in 210 families and case-only WES in 489 individuals with epileptic encephalopathies. The functional effect of CYFIP2 variants on WAVE signaling was evaluated by computational structural analysis and *in vitro* transfection experiments. We identified three *de novo* CYFIP2 variants at the Arg87 residue in four unrelated individuals with early-onset epileptic encephalopathy. Structural analysis indicated that the Arg87 residue is buried at an interface between CYFIP2 and WAVE1, and the Arg87 variant may disrupt hydrogen bonding, leading to structural instability and aberrant activation of the WAVE regulatory complex. All mutant CYFIP2 showed comparatively weaker interactions to the VCA domain than wild type CYFIP2. Immunofluorescence revealed that ectopic speckled accumulation of actin and CYFIP2 was significantly increased in cells transfected with mutant CYFIP2. Our findings suggest that *de novo* Arg87 variants in CYFIP2 have gain-of-function effects on the WAVE signaling pathway, and are associated with severe neurological disorders.

2044T

Fragile X syndrome drosophila mutants display defects in stress odorant response: A new model for sensory processing dysfunction. *F. Bolduc, A. Androschuk, C. Rosenfelt, R. He.* University of Alberta, Edmonton, Alberta, Canada.

Neurodevelopmental disorders such as intellectual disability (ID) and autism spectrum disorder (ASD) affect significantly sensory perception and therefore behavior in most affected individuals. Silencing expression of the Fragile X mental retardation 1 (*FMR1*) gene leads to Fragile X syndrome (FXS), the most common single gene cause of ID and ASD. Individuals with FXS frequently present with significant changes in their sensitivity to sound and pain for instance. The molecular basis of these changes in sensory processing remain largely unknown. *Drosophila* have a highly conserved *FMR1* orthologue, *dfmr1*. We and others have shown before that the *dfmr1* mutants display cognitive and social defects reminiscent of symptoms seen in individuals with FXS. To explore sensory processing, we took advantage of a well-established behavioral assay for sensory processing of the *Drosophila* stress odorant (dSO). Here, we show that *dfmr1* mutant flies present significant defects in dSO response. Importantly, using pharmacological manipulations, we show that the response defects to dSO in *dfmr1* mutant flies can be ameliorated. Together, our results show that treatment in FXS should be considered with the perspective of behavior specificity. Moreover, our results provide a novel model to further understand the impact of FXS on social interaction and sensory processing.

2045F

Additional fragile X related 2 (Fxr2) KO mice exhibit severe phenotypes. *Y. Gu, D.L. Nelson.* Human and Molecular Genetics, Baylor College of Medicine, Houston, TX.

Fragile X syndrome is the most common cause of inherited monogenic intellectual disability, often associated with autism spectrum disorder. To understand the function of FMR1 gene product, FMRP, it is important to determine how the paralogs FXR1 and FXR2 compensate for its absence in Fragile X syndrome. We have developed Fmr1, Fxr1, Fxr2 knockout, and Fmr1 and Fxr1 conditional knockout mouse models, and analyzed these models at molecular, pathological, electrophysiological and behavioral levels. To develop comprehensive FMR1 gene family mouse model tools, we recently characterized a conditional knockout mouse model for Fxr2, developed through the Knockout Mouse Project (KOMP). We obtained the Fxr2 knockout first model (Fxr2 Tm2a), which was created by insertion of lacZ and neomycin genes into Fxr2 intron 3. FRT sites surround lacZ and neomycin genes and three Lox P sites are included; two surround exons 4 and 5, and the other lox P site is located between lacZ and neomycin genes. We created an Fxr2 conditional knockout model (Fxr2 Tm2c) after lacZ and neomycin genes were deleted using ubiquitously expressed flip recombinase (FLP). A new Fxr2 knockout model (Fxr2 Tm2d) was developed by deletion of exons 4 and 5 through universally expressed Cre recombinase. The new Fxr2 KO (Fxr2 Tm2d) showed low body weight and reduced levels of anxiety in behavioral tests in contrast to the original Fxr2 KO mice, which did not show these abnormalities. Using RT-PCR, we found Fxr2 mRNA expressed by both KO lines. The original KO expresses Fxr2 mRNA lacking exon 7, leading to a predicted Fxr2 protein deleting 39 amino acids in KH0 region. The new Fxr2 KO (Fxr2 Tm2d) expresses Fxr2 mRNA without exons 4 and 5, leading to a predicted Fxr2 protein containing only the first 77 amino acids of the N-terminal region. We also characterized several lines containing insertion of lacZ and/or neomycin. All these lines showed partial lethality and body weight reduction, in addition to severe differences in behavioral tests. RT-PCR analysis showed Fxr2 mRNA contains inserted exogenous 115 bps immediately after exon 3, leading to new truncated proteins. In summary, we have developed several new Fxr2 mutant lines including the first Fxr2 conditional knockout line. These mice showed more severe phenotypes compared with the originally reported Fxr2 knockout due to different splicing mRNA. These new mouse models will provide new additional tools to further understand FMR1 gene family function.

2046W

CRISPR/Cas9 mediated repair of Gaucher-type 2 patient mutations in iPSC-derived neural progenitor cells. *K.M. Baumgarner^{1,2}, J. Dizon^{1,2}, C.E. McKinney^{1,2}.* 1) Edward Via College of Osteopathic Medicine, Spartanburg, SC; 2) Gibbs Cancer Center and Research Institute, Spartanburg, SC.

Gaucher disease type 2 (GD2) is a lethal autosomal recessive disorder with a life expectancy less than 2 years of age. This disease results from a defect in the glucocerebrosidase enzyme encoded by *GBA1*. Since glucocerebrosidase breaks down glucocerebroside, an accumulation of this lipid occurs in GD2 patient's cells throughout the body affecting the spleen, liver, lungs, bones, and the brain. With no therapies available for this form of the disease, studying the molecular mechanisms involved in neuronopathic Gaucher disease could lead to future treatments. Using patient-derived GD2 fibroblasts lines, we have produced induced pluripotent stem cell lines (iPSC) differentiated into neurons. To better understand the mechanisms involved in GD2 disease, we are also creating an isogenic control line for GM02627 through CRISPR/Cas9 directed homologous repair. GM02627 is a compound heterozygote in exon 8 of *GBA1*: one allele has a G>A transition that results in substitution of arginine for glycine (G325R), and the second allele has a T>G transition that results in a glycine to cysteine substitution (C342G). CRISPR targeting of this cell line occurred at the neural progenitor cell (NPC) stage of differentiation. GM02627 NPCs were co-transfected with CAS9 ribonucleoprotein and a 122bp oligo containing a wildtype sequence for exon 8 of *GBA1*. Three gRNA combinations were tested to find one that simultaneously repaired both alleles. After CRISPR/Cas9 targeting, the pools of NPCs were cloned using limiting dilution to produce clonally derived cell lines. Additionally, the NPC positive controls were analyzed for CRISPR/Cas9 efficiency using a genomic cleavage assay and the number of recombination events were quantified using amplicon next generation sequencing PCR products from NPC pooled DNA samples. By creating isogenic control lines for multiple GD2 cell lines, we will be able to compare the GD2 and 2627 repaired isogenic control. Exosome composition, electrophysiological status, global gene and protein expression data will be obtained from neurons that would normally be inaccessible for study.

2047T

Modeling microcephaly in a monkey model by disrupting the ASPM gene via CRISPR/Cas9. *X. Guo¹, Z. Tu¹, W. Yang¹, S. Yan¹, S. Li^{1,2}, X. Li^{1,2}.* 1) Guangdong-Hongkong-Macau Institute of CNS Regeneration, Jinan University, Guangzhou, Guangdong, China; 2) Department of Human Genetics, Emory University, Atlanta, USA.

The ASPM (abnormal spindle-like primary microcephaly) gene is evolutionally critical for brain volume expansion and intelligence development. Loss of function or down-regulation of the ASPM gene is the most common genetic cause for primary microcephaly in humans. To investigate the function of the ASPM gene in the primate brains, we disrupted the ASPM gene in one-cell stage embryos of cynomolgus monkeys (*Macaca fascicularis*) by injecting guide RNA and CRISPR/Cas9 into fertilized monkey eggs. We obtained 6 live monkeys in which the ASPM gene has been targeted. PCR and sequencing analysis revealed mosaic mutations in these targeted monkeys. One monkey exhibited strikingly head circumference decrease (<2SD to age-matched and same gender animals), which recapitulates the important clinical feature of microcephaly. Our findings demonstrated that the ASPM gene possess evolutionally similar function between old world monkeys and humans and also suggest that the ASPM gene knockout monkeys can provide an important animal model to investigate the ASPM function in primate brains.

2048F

Mutations in transmembrane domain 4 of *GRIA3* are responsible for a specific movement disorder. J. Piard¹, M. Bereau², T. Wirth³, D. Amsalem⁴, M. Raynaud⁵, J. Chelly⁵, L. Van Maldergem¹. 1) Centre de Génétique Humaine, CHU Besançon, Besançon, France; 2) Service de Neurologie, CHU Besançon, Besançon, France; 3) Département Médecine translationnelle et neurogénétique, IGBMC, Strasbourg, France; 4) Service de Pédiatrie, CHU Besançon, Besançon, France; 5) Laboratoire de génétique moléculaire, CHU Tours, Tours, France.

Ionotropic amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) glutamate receptors mediate the majority of excitatory synaptic transmission in the central nervous system. *GRIA3* is located on X chromosome and encodes the GLUA3 subunit of AMPA receptor. Mutations in *GRIA3* have been associated with intellectual disability, autism spectrum disorder and epilepsy in males. By whole exome sequencing, we identified the missense variant G826D in *GRIA3* in three male patients from a large Caucasian family and their unaffected mothers. The mutation involves a highly conserved residue in a key functional domain of iGluR3, the fourth transmembrane domain (TM4). Patients 1 and 2 are twin brothers evaluated for moderate intellectual disability and a specific movement disorder which started at three years of age and predominantly affects the upper limbs. Brain MRI is normal in both brothers. Examination showed generalized myoclonic jerks, mild ataxia and brisk reflexes. Patient 3 is a male proband evaluated for learning difficulties and continuous generalized myoclonic jerks. A diagnosis of hyperekplexia was initially suggested but not confirmed by molecular analyses. Examination showed brisk reflexes. Brain MRI was normal. In 2007, Wu et al. studied a patient with ID and myoclonic jerks and the G833R *GRIA3* missense variant localized in TM4 domain. Our observation suggests that missense *GRIA3* mutations in TM4 domain could be responsible for a specific movement disorder in patient with X-linked ID.

2049W

A *Drosophila* model of *PIGA* deficiency, an ultra-rare X-linked intellectual developmental disorder. E. Coelho, C.Y. Chow. Department of Human Genetics, University of Utah, Salt Lake City, UT.

Mutations in the *Phosphatidylinositol glycan class A (PIGA)* gene cause *PIGA* deficiency, a type of X-linked intellectual developmental disorder. *PIGA* deficiency is an ultra-rare disease with fewer than 12 patients reported. *PIGA* deficiency is characterized by neonatal hypotonia, myoclonic seizures, dysmorphic features, and a number of congenital anomalies. *PIGA* is involved in the first step of glycosylphosphatidylinositol (GPI) anchor biosynthesis by transferring N-acetylglucosamine (GlcNAc) from uridine 5'-diphospho N-acetylglucosamine (UDP-GlcNAc) to PI to form GlcNAc-PI. GPI is a glycoprotein that attaches the C-terminus of a protein to the cell surface. GPI-anchored proteins play a number of roles in cell-cell signaling, cell migration, and immunity. Here we report the development of a *Drosophila* model of *PIGA* deficiency using CRISPR technology. We generated a null allele and demonstrate that loss of *PIGA* function in *Drosophila* results in multisystem dysfunction and larval lethality. Because a number of the patient phenotypes include nervous system abnormalities, we use RNAi technology to knockdown *PIGA* expression in neurons. Neuron-specific loss of *PIGA* function results in behavioral and neurological abnormalities that are reminiscent of those observed in patients. We will also present a number of other tissue-specific knockdown studies, including muscle, to identify the primary defects caused by loss of *PIGA* function. The majority of human mutations in *PIGA* are missense variants in evolutionarily conserved amino acid residues. This provides the opportunity to study the structure and function of the *PIGA* protein and how missense variants contribute to disease. We are generating transgenic *Drosophila* strains expressing a number of the disease-causing missense variants that will be crossed into the null background we generated. Comparing the disease allele phenotypes to the null model will provide insight into the disease pathogenesis, how much function is retained, and genotype/phenotype relationships. Importantly, this study paves the way for identification of genetic modifiers and small molecule screens that may lead to therapeutic approaches for this ultra-rare disease.

2050T

The different regions in the central of exon 7 (+11~+40) has different effect on the splicing of SMN exon7. Y. Qu, J. Bai, Y. Cao, H. Wang, Y. Jin, F. Song. Capital Institute of Pediatrics, Beijing, China.

Humans have two copies of Survival Motor Neuron gene: SMN1 and SMN2. Homozygous deletions or subtle variants of survival motor neuron gene 1(SMN1) can cause spinal muscular atrophy (SMA). Using site-specific mutations and ex vivo splicing assay we revealed that the site +11~+15, site +15~+21,site +17~+21 ,site 31~+35,and site +36~+40 in exon 7 of SMN gene might be the exonic splicing silencer regions. For the deletion of these regions can cause the increase of exon7 inclusion. While site from +22 to +30 might be exonic splicing enhancer region, and the key site were +22,+27,+29. For the rate of exon7 skipping were increased in the deletion vectors of region +22~+27 (93%),region +27~+31(70%), and region +22~+30 (100%) in SMN1 gene. Here, we also revealed that the antisense-oligonucleotide targeting the intronic splicing silencer N1 (ISS-N1) fully corrects SMN2 exon 7 splicing and restores high levels of SMN both in homozygous deletions SMA patient cells and in the subtle variants SMA patient cells. Based our previous results and this research results, it might be one silencer regions binding with hnRNP A1 in the central of SMN exon 7 and the Tra2 β 1 binding region might be in site+22 to site +30 of SMN exon 7.Our results demonstrated that different regions in the central of exon 7 (site +11 to +40)has different effect on the splicing of SMN1 exon7. This research was supported by the Beijing Natural Science Foundation (Project No.7172039), the Science Foundation of the Capital of Institute of Pediatrics (Project No.Fangxiang-2017-02).the National Key Research and Development Program of China (No. 2016YFC0901505), CAMS Initiative for Innovative Medicine (CAMS-I2M), and the National Natural Science Foundation of China (Project Nos. 81100933, 85009719,81501850)

*Corresponding author :Fang Song , Capital Institute of Pediatrics, No 2, Ya Bao Road, Beijing, 100020, China. Email : songf_558@263.net.

2051F

Over-expression of human *SMOOTHENED* (*SMO*) induces an abnormal head phenotype in zebrafish. M. Tanima-Nagai, S. Hong, E. Roessler, M. Muenke. The Medical Genetics Branch, National Human Genome Research Institute (NHGRI), National Institutes of Health (NIH), Bethesda, MD.

Background: The Hedgehog (Hh) signaling pathway plays several crucial roles in the development of most of animals, including the brain and neurogenesis. Accordingly, aberrant Hh pathway function has been implicated in many human disorders ranging from birth defects to cancer. Holoprosencephaly (HPE) is the most common congenital anomaly of the forebrain and face in humans. *SONIC HEDGEHOG* (*SHH*) is one of the best known HPE genes. Recent clinical and molecular data have identified mutations in *SMOOTHENED* (*SMO*), which is an obligate part of the Shh pathway, in HPE patients. Here we describe experimental approaches to reveal the function of *SMO* variants for forebrain or craniofacial development using zebrafish. **Materials and methods:** Wild-type (TAB5) zebrafish were used in this study under an NHGRI animal protocol. For micro-injection assays, *SHH* and *SMO* synthetic mRNA was injected separately or co-injected into fertilized embryos at the one or two-cell stage. A 100 pg bolus of *gfp* was used as a negative control. After 24 hours post fertilization (hpf) and 48 hpf, the phenotype of the injected embryos were analyzed. Expression of Shh pathway gene markers (*shh*, *smo*, *ptch2*, *nk2.2a*, *pax6*, *Gli1*, *Gli2a*) were confirmed via whole mount *in situ* hybridization. All experimental mRNA injections were performed in triplicate and on different days. **Results:** Single injection of *SMO* (100, 200, 300, 600 pg) alone typically did not induce any abnormal phenotypes. Low dose *SHH* (1, 2 pg) injections most frequently caused a normal phenotype. Interestingly, co-injection of a low amount of *SHH* (1 or 2 pg) and *SMO* 200 pg or higher dose(s) induced an abnormal phenotype. This phenotype showed less pigmentation of the eye region at 24 hpf and an incomplete eye shape at 48 hpf. *ptch2* expression at 48 hpf (a known target of the Hh pathway) showed more intensity in co-injection of *SHH* and *SMO* than *SMO* single injection alone. **Conclusion:** These experiments show that co-injection of high amounts of *SMO* and a minimal dose of *SHH* induces eye abnormalities. We are currently investigating the mechanism of this synergy. We plan to investigate the consequences to gene expression targets, which related to the Shh pathway, via whole mount *in situ* hybridization and western blotting. Ultimately, we hope this work will lead to a better understanding of pathway dynamics and opportunities for variant assay testing.

2052W

Establishment of cell lines for drug screening for Alexander disease using glial fibrillary acidic protein cellular localization. J. Tulyeu, E. Jimbo, Sh. Aoki, T. Yamagata, H. Osaka. Jichi Medical University, Tochigi, Japan.

Background: Alexander disease (AxD) is a devastating leukodystrophy caused by mutations in the *glial fibrillary acidic protein* (GFAP) gene. The pathological feature of the disease is the cytosolic aggregation of Rosenthal fibers in astrocytes. The clinical symptoms of severe AxD are motor and mental retardation, bulbar dysfunction, seizures, microcephalus and early death. Mild AxD patients manifest muscle weakness, spasticity and ataxia in later in life. No effective treatments have been established for any type of AxD. In this study, we screened drugs using *in vitro* cell lines with an Incell Analyzer to quantify aggregate formation. **Methods:** We transiently transfected GFAP wild-type and eight mutants (R79C, F80S, R88C, R239C, R239P, A244V, R258C and R416W) fused to green fluorescent protein [GFP] into HeLa and U2-OS cell lines, and quantified the aggregate formation using an Incell Analyzer. We then screened chemical libraries to quantify the aggregate formation in HeLa cell line expressing GFAP-F80S. **Results:** In both cell lines, all GFAP-GFP mutants made more aggregate compared to wild type. Using HeLa cells expressing GFAP-F80S, we performed *in vitro* screen of 138 chemicals and identified 7 that reduced the formation of aggregates. **Discussion:** The Incell Analyzer quantitatively proved the enhanced aggregate formation in GFP mutants. We are now undertaking to screen drugs that reduce the aggregate formation using this system.

2053T

De novo AMPAR *GRIA* variants associated with neurological diseases.

W. Xiangwei^{1,2}, H. Yuan^{1,3}, S.J. Myers^{1,3}, Y. Jiang², S.F. Traynelis^{1,3}. 1) Department of Pharmacology, Emory University School of Medicine, Atlanta, GA, 30322; 2) Department of Pediatrics, Peking University First Hospital, Beijing, 100034, China; 3) Center for Functional Evaluation of Rare Variants (CFERV), Emory University School of Medicine, Atlanta, GA, 30322.

α -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptors (AMPA receptors) play important roles in brain development and a wide range of pathological conditions. Here, we evaluated functional changes on 11 missense *GRIA* variants (1 *GRIA1* variants, 2 *GRIA2* variants and 8 *GRIA3* variants) associated with epilepsy and/or intellectual disabilities. Of these 11 patients, 45% (5/11) show epilepsy and intellectual disability, and 27% (3/11) present only with intellectual disability with no seizures. 18% (2/11) are in the agonist binding domain (ABD), 46% (5/11) are in transmembrane domains (TMs), and 36% (4/11) of variants are in linker regions. Functional evaluation using two-electrode voltage clamp current recordings revealed 2 gain-of-function variants that enhanced glutamate and kainate potency by 13~52-fold, compared to the corresponding WT receptors. We also found 7 loss-of-function variants that reduced glutamate and kainate potency by 3~5-fold, compared to the WT receptors. All of the gain-of-function variants and 6 of the 7 loss-of-function variants are located in TMs or the linker regions. Notably, similar diseases were associated with both gain- and loss-of-function variants in the same gene. Overall, our results suggest that AMPAR variants are highly relevant with epilepsy and intellectual disabilities in children, and the transmembrane domain or the linker regions between transmembrane domains were particularly intolerant to functional variation.

2054F

Relapsing remitting multiple sclerosis and its relationship with the immunology system and with the oxidative stress. A.P. Soto Brambila¹, P. Rivero Moragrega², G. Gabriel Ortiz³, F.P. Pacheco Moises⁴. 1) División de Genética Humana, Centro De Investigación Biomédica de Occidente, Cibo, Imss, Universidad De Guadalajara, Cucs, Mexico. Sierra Mojada; 2) División de Genética Humana, Centro De Investigación Biomédica de Occidente, Cibo, Imss, Universidad De Guadalajara, Cucs, Mexico. Sierra Mojada; 3) División de Genética Humana, Centro De Investigación Biomédica de Occidente, Cibo, Imss, Universidad De Guadalajara, Cucs Neurociencias, Mexico; 4) División de Genética Humana, Centro De Investigación Biomédica de Occidente, Cibo, Imss, Universidad De Guadalajara, Cucei, Departamento de Bioquímica, Mexico.

Background. Multiple Sclerosis (MS) is a chronic autoimmune neurological disease characterized by a demyelinating inflammatory response of the Central Nervous System (CNS). During the natural evolution of the disease there are episodes of inflammation and high levels of oxidative stress products that have been related to the progression and severity of the disease. **Objective.** To evaluate the expression of the mRNA of the genes that encode inducible and constitutive nitric oxide synthase and its association with the indicators of oxidative stress in plasma (Nitric Oxide Synthase activity, Glutathione Peroxidase enzymatic activity, Nitrate-Nitrite, Lipoperoxide, Fluidity Membrane) in patients with Relapsing Remitting Multiple Sclerosis (RRMS). **Conclusions.** The MRI of patients with RRMS had a very similar pattern in terms of the presence or recent appearance of new lesions. The expression of *iNOS* is increased in leukocytes from patients with RRMS compared to healthy controls. The expression of *iNOS* is decreased in leukocytes of patients with RRMS at the end of the month of evolution with respect to the first intake and at the end of the week of the outbreak compared with healthy controls. Also nitric oxide metabolites, lipid peroxidation products, nitric oxide synthase activity and glutathione peroxidase activity were significantly increased in plasma from subjects with RRMS compared to healthy controls. The plasma membrane fluidity of subjects with RRMS has values similar to that of plasma of healthy controls. Both the biochemical indicators of oxidative stress and the expression of the *iNOS* gene are increased in the plasma of patients with RRMS. Therefore, there is a positive association of the biochemical indicators of oxidative stress in plasma with the expression of *iNOS*.

2055W

Homozygous variants in newly described genes in early childhood-onset autism spectrum disorder: How to proceed? Y. Anikster^{1,2}, D. Marek-Yagel^{1,2}, E. Ganelin-Cohen², B. Poded-Shakked^{1,2}, L. Gabi^{1,2}, H. Cassuto^{1,3}. 1) Metabolic Disease Unit, Edmond and Lily Safra Children's Hospital, Tel Hashomer, Israel; 2) Sackler Faculty of Medicine, Tel-Aviv University, Tel-Aviv, Israel; 3) Child Development, Leumit HMO, Jerusalem.

Next Generation Sequencing (NGS) techniques have revolutionized molecular diagnosis and brought to the identification of numerous genes not previously associated with human disease. Occasionally, NGS can unveil rare biallelic deleterious variants in such genes in a single family, posing a challenge for the clinician in inferring and establishing their pathogenicity. Two illustrative examples in families with early childhood-onset Autism Spectrum Disorder (ASD) are described. In family A, the proband, 5.5 year old boy diagnosed with high functioning ASD presented with acute psychosis and significant developmental regression. His younger 3 years-old brother had also been diagnosed with ASD. The siblings were born to non-consanguineous parents of Sephardic-Jewish descent. As an extensive neurological evaluation failed to yield a diagnosis, whole exome sequencing (WES) was pursued for the proband and his parents, followed by targeted Sanger sequencing in additional family members. This brought to the identification of a novel nonsense mutation in *TMEM101*, for which both affected siblings are homozygous, and which further fully segregated in the extended family. In family B, the proband is a 4 years-old girl who was referred due to ASD diagnosed at 30 months of age. She is the first child born to non-consanguineous parents of Ashkenazi-Jewish descent. The patient has no clear dysmorphic features and her cognitive status is age-appropriate. Single WES revealed a novel homozygous mutation in *CHRNA7*, which encodes the alpha7 subunit of the neuronal nicotinic acetylcholine receptor, and has long been suggested as a candidate gene associated with ASD. Data sharing, via match-making software or networks of clinicians and researchers may facilitate the recognition of additional affected families worldwide and aid in the elucidation of further monogenic causes of autism spectrum disorder and ultra-rare neurodevelopmental disorders.

2056T

Evaluating genetic causation and personalized pharmacological treatment of an ultra-rare disease associated with deletion of *CACNG2*. M.L. Kleiber¹, T.R. Chapman¹, M.L. Maile^{1,3}, O. Hong¹, R. Purcell², A. Lipka², A. Roberts⁴, J.L. Sebat¹. 1) University of California, San Diego, San Diego, CA; 2) RespireRx Pharmaceuticals, Glen Rock, NJ; 3) Neuropore Therapies, San Diego, CA; 4) The Scripps Research Institute, San Diego, CA.

The characterization of genomic sequence variants in autism spectrum disorder (ASD) and neurodevelopmental disorders (NDD) studies has led to the identification of causal genes that may provide novel opportunities for individualized treatment strategies. Effective treatments for most ASDs and NDDs have been elusive due to a lack of understanding of how a genetic variant within a patient plays a causal role, as well as the establishment of an intervention that can demonstrably target the underlying molecular pathology and improve cognitive function *in vivo*. To demonstrate the feasibility of personalized medicine for a rare ASD/NDD, we have generated and evaluated a mouse model using CRISPR/Cas9 genome editing. This model recapitulates a previously reported case of autosomal dominant mental retardation 10 (MRD10, OMIM: 614256) and is characterized by an exon 2 in-frame deletion of the AMPA-receptor binding domain of *Cacng2* (*Cacng2*^{Δex2}), a transmembrane AMPA receptor regulatory protein that mediates rapid glutamatergic excitatory signaling facilitating receptor trafficking, gating, and developmental synaptic plasticity. In *Cacng2*^{Δex2/Δex2} mice, mRNA is expressed at normal levels in the cerebellum; however, no protein is detectable in cerebellar whole-cell lysate or post-synaptic fractions, suggesting that the mutant protein may be targeted to the lysosome for degradation. Behaviourally, homozygous *Cacng2*^{Δex2/Δex2} C57BL/6J mice present with severe motor ataxia. Heterozygous *Cacng2*^{Δex2/Δex2} males and females show significant cognitive and behavioral phenotypes such as hyperactivity, decreased anxiety-related behaviour, decreased startle response, deficits in reversal learning, and heightened male aggression. Maternal care deficits in female *Cacng2*^{Δex2/Δex2} carriers resulted in significant loss of fecundity (<40% pup survival by postnatal day 5). With the goal of reversing cognitive impairments associated with MRD10, we have initiated treatment studies using the ampakine compound CX717 at doses of 0.3 and 3 mg/kg. Given the function of CX717 as a positive allosteric modulator of AMPA receptor activity, this represents a valuable opportunity to evaluate a potential treatment for MRD10. Ultimately, ASD is a complex and challenging disorder that chronically affects both individuals and families. The biological causes for ASD and associated disorders are diverse – our data suggests that a personalized approach may provide valuable insights and effective treatment options.

2057F

***NSD1* overexpression recapitulates the phenotype of the 5q35.2-5q35.3 duplication syndrome and unveils a potential treatment for its undergrowth phenotype.** C.C. Eno¹, K. Jones², M. Nowaczyk³, D. Wong⁴, G. Mirzaa⁵, A. Beck⁶, J.A. Martinez-Agosto^{4,7,8}, F. Quintero-Rivera^{1,9}. 1) Pathology and Laboratory Medicine, University of California, Los Angeles, Los Angeles, CA; 2) Children's Hospital of The King's Daughters, Norfolk, Virginia; 3) McMaster University, Hamilton, Ontario, Canada; 4) Department of Pediatrics, UCLA Health, Los Angeles, California; 5) University of Washington, Seattle, Washington; 6) Seattle Children's Hospital, Seattle, Washington; 7) Department of Human Genetics, David Geffen School of Medicine at UCLA; 8) Department of Psychiatry, Semel Institute for Neuroscience and Human Behavior David Geffen School of Medicine at UCLA; 9) UCLA Medical Center, UCLA Clinical Genomics Center, David Geffen School of Medicine at UCLA.

Due to gene dosage sensitivity, Nuclear receptor binding SET domain protein 1 (*NSD1*), a H3K36 histone methyl transferase, located at 5q35.3, is a candidate gene responsible for the phenotype of the reciprocal microdeletion-microduplication syndromes. Heterozygous deletions of or variants in *NSD1* are associated with Sotos syndrome, which is characterized by overgrowth with advanced bone age, and height and head circumference (OFC) ≥ 2 standard deviations (SD). In contrast, the duplication of 5q35.2-q35.3, encompassing *NSD1*, is characterized by undergrowth with delayed bone age, as well as height and OFC ≤ 2 SD. We report nine new patients (ages: 1 month to 33 years) with a duplication of 5q35.2-q35.3, thus contributing to the emerging phenotype, the long-term prognosis, penetrance, and expressivity of this recently described microduplication syndrome. It also highlights psychiatric phenotypes of the syndrome, and suggests its inclusion as part of a differential diagnosis for children with delayed bone age and short stature. Focusing specifically on the main phenotype, undergrowth, we have modeled the effects of *NSD1* overexpression in *Drosophila melanogaster*. *Mes-4* (*NSD1* *Drosophila* homolog) overexpression results in growth restriction through increased cell death associated with decreased PI3K signaling downstream Insulin Receptor (InR). Nutrients (i.e. leucine) can activate mTOR independent of PI3K upstream signaling, thus bypassing the InR/PI3K pathway. We also show that a leucine-rich diet throughout all larval stages rescues *Mes-4/NSD1* induced cell death and suggests a possible treatment to ameliorate the human phenotype. Our findings emphasize the tight regulation of epigenetic factors involved in tissue growth and organism size. We provide the first evidence in a model organism of the effect of *NSD1* dosage on growth, and its association with opposing phenotypes in human syndromes.

2058W

Oligogenic effects of *16p11.2* copy number variation on craniofacial development. Y. Qiu¹, T. Arbogast², H. Li¹, S. Tang³, E. Richardson², O. Hong³, S. Cho³, O. Shanta^{3,4}, T. Pang³, C. Corsello^{3,5}, C. Deutsch⁶, C. Chevalier^{7,8,9}, S. Lorenzo^{7,8,9}, E. Davis², L. Iakoucheva³, Y. Hérault^{7,8,9}, N. Katsanis², K. Messer¹, J. Sebat^{10,11,12,13}. 1) Division of Biostatistics and Bioinformatics, Department of Family Medicine and Public Health, University of California San Diego, La Jolla, CA; 2) Center for Human Disease Modeling, Duke University, Durham, NC; 3) Department of Psychiatry, University of California San Diego, La Jolla, CA; 4) Department of Electrical Engineering, University of California San Diego, La Jolla, CA; 5) Rady Children's Hospital, San Diego, CA; 6) Eunice Kennedy Shriver Center UMMS, Charlestown and Worcester MA; 7) Institut de Génétique et de Biologie Moléculaire et Cellulaire, Université de Strasbourg, Illkirch, France; 8) Centre National de la Recherche Scientifique, UMR7104, Illkirch, France; 9) Institut National de la Santé et de la Recherche Médicale, U964, Illkirch, France; 10) Beyster Center for Genomics of Psychiatric Diseases, University of California, San Diego, La Jolla, CA; 11) Institute for Genomic Medicine, University of California, San Diego, La Jolla, CA; 12) Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, CA; 13) Department of Pediatrics, University of California, San Diego, La Jolla, CA.

Copy number variation (CNV) of *16p11.2* (*BP4-BP5*) is associated with developmental and psychiatric disorders and differences in head size. However, the effect of *16p11.2* genes on craniofacial development is not understood. We investigated the effect of *16p11.2* copy number on craniofacial development in humans, mice and zebrafish. 3D morphometric facial imaging was performed on 228 subjects including 45 deletion carriers, 44 duplication carriers and 139 familial controls. We used a series of linear mixed-effects models to test for the effect of the deletion and the duplication on human facial features. We show that copy number of *16p11.2* is positively correlated with growth of the nose and mandible and negatively correlated with growth of the brow and maxilla. Results were then validated by computed tomography (CT) scans of skulls from mouse models of the deletion and duplication. Structure-specific effects of gene dosage on mandible and frontonasal growth were recapitulated in zebrafish by modulating the expression of the genes *MAPK3*, *MVP* and *KCTD13* in combination, and we find evidence for synergistic effects between genes. Our results suggest that the deletion and duplication of *16p11.2* have distinctive facial features. These effects are attributable the combined influence of multiple genes and their differential effects on growth of the major orofacial processes.

2059T

Alternative splicing of *FMR1* gene is complex: More spliced products and more splicing patterns are discerned in the single-gene transcriptome from the disease gene of fragile X syndrome and FXTAS. F. Lan, P. Qiu, W. Yang, J. Liao, X. Guo, X. Fu, A. Yan, D. Zhang. DongFang Hospital, Medical College of Xiamen University, Fuzhou, Fujian Province, China.

Fragile X syndrome (FXS) is a common inherited mental retardation disorder. Its disease-determining gene is fragile X mental retardation 1 gene (*FMR1*), encoding fragile X mental retardation protein (FMRP). The *FMR1* gene contains 17 exons and 16 introns, and complex alternative splicing phenomenon of *FMR1* gene was observed in its initial transcript. Different alternatively spliced variants of *FMR1* gene may encode FMRPs with different structures and functions. However, how many alternatively spliced isoforms are present and what are their biological functions remain largely elusive. To investigate the alternative splicing of *FMR1* gene in human peripheral blood, RT-PCR and cloning-sequencing technology was employed to detect the types of alternatively spliced products of peripheral blood leukocytes in 10 normal individuals and to statistically analyze their expression abundance. The results showed that a total of 50 alternatively spliced products were identified from 505 PCR product-T vector recombinant plasmids, including 27 novel alternatively spliced products that have not been reported. The highest abundance in the alternatively spliced products belongs to ISO 17 and ISO 7, both of which involve exon 12 skipping, followed by ISO 1 (containing the full-length sequence of the coding region) and ISO 13 (using the second acceptor site of exon 17), while the frequency of other types of spliced products is not higher than 5.0%. The novel alternatively spliced variants identified in this study involved skipping of exons 3, 4, and 11, insertion of 30bp and 87bp sequence spliced from the middle of intron 7, and 46 bp and 140bp sequence spliced from the middle of intron 9. We further focused on the 140bp exon and revealed that this novel exon from the middle of intron 9 of human *FMR1* gene was commonly expressed in various human tissues and cells. Hybrid minigene splicing assay and bioinformatics analysis also provided evidences that the 140bp sequence held the potentiality of being an alternative exon. To explore the biological functions of the new transcript, recombinant eukaryotic expression vectors and lentiviral overexpression vectors were constructed. Our experiments showed the spliced transcript encoded a truncated protein with altered subcellular localization, and RNA microarray revealed that the expression of a few genes involved in FXS-related signaling pathways was significantly influenced when the truncated FMRP was overexpressed.

2060F

The hnRNP gene family as a model for understanding neurodevelopmental disorder subtypes. M.A. Gillentine¹, T. Wang¹, B. Coe¹, J. Rosenfeld², E. Torti³, A.S.I.D Network⁴, R. Bernier^{5,6,7}, E.E. Eichler¹. 1) Genome Sciences, University of Washington, Seattle, WA; 2) Baylor College of Medicine, Houston, TX; 3) GeneDx, Gaithersburg, MD; 4) ASID Network; 5) Seattle Children's Autism Center, Seattle, WA; 6) Psychiatry and Behavioral Sciences, University of Washington; 7) Center on Human Development and Disability, University of Washington.

With the increasing number of whole-exome and -genome sequencing studies, the number of genes implicated in neurodevelopmental disorders (NDDs) has risen dramatically. It is likely that gene families and cellular pathways will have shared clinical presentations and molecular pathomechanisms, potentially allowing for targeted therapies. We performed an analysis of 10,927 published exomes of NDD probands from denovo-db (<http://denovo-db.gs.washington.edu/denovo-db/>) to identify potential risk genes. Among the genes with an excess of *de novo* mutations ($p < 2.5 \times 10^{-6}$), genes associated with RNA processing, specifically heterogeneous ribonucleoproteins (hnRNPs), stood out with significantly more *de novo* likely gene-disruptive (LGD) variants than expected ($p < 2.2 \times 10^{-16}$). In total, variations in ten hnRNPs were identified. The top two hnRNPs, *HNRNPU* and *SYNCRIP*, were significantly enriched for both LGD and missense variants for all NDDs, as well as intellectual disability/developmental delay alone. We targeted six hnRNP genes for sequencing using molecular inversion probes in a cohort (ASID Network) of 18,836 patients and identified ten LGD, three severe missense mutations, and two splice variants in 17 new patients. Combining these data with clinical exome sequencing data generated from Baylor Genetics Laboratories and GeneDx, we have identified a total of 117 new patients (a third of which are variations in *HNRNPU* and *SYNCRIP*), with a subset having further phenotypic information available or are being clinically assessed. Using this information and published cases, we are expanding the phenotypes associated with hnRNP variation and determining if these are shared across the gene family. hnRNPs are RNA-binding proteins implicated in many processes, including neuronal development, and often work cooperatively and/or compensatorily. To assess the functional impact of these mutations, molecular phenotypes are being assessed with CRISPR-edited induced pluripotent stem cells and differentiated cortical neurons with both proband-specific variants and gene knockouts. We hypothesize that variation in hnRNP-encoding genes may result in shared, targetable neuronal phenotypes, including changes in including cell viability and proliferation, neuronal morphology, and neuronal activity, as well as hnRNP-specific functionality, such as splicing and RNA/protein localization. Additionally, we will attempt to rescue these phenotypes at the cellular level.

2061W

Using *Xenopus laevis* as a model for studying genes associated with intellectual disabilities. M. Lasser¹, L. Pizzo², J. Tiber¹, C. Monahan¹, S. Kim¹, S. Lee¹, S. Girirajan², L.A. Lowery¹. 1) Boston College, Chestnut Hill, MA; 2) Penn State University, State College, PA.

Intellectual disabilities are complex neurodevelopmental disorders that affect a significant portion of the general population and can be an immense health issue. Genetic mutations caused by rare copy number variants (CNVs) have been shown to play a role in intellectual disabilities or severe developmental delays. It has been previously shown that one such CNV results in a deletion within chromosome 16 in a subset of children clinically diagnosed with an intellectual disability. In addition to having intellectual disabilities, individuals with this deletion, located at 16p12.1, also have severe craniofacial abnormalities, microcephaly, cardiac defects, seizures, and growth retardation. This deletion encompasses six genes, many of which have not yet been studied in relation to neurodevelopment. In order to elucidate the role of these six genes during development, we have created antisense morpholino oligonucleotides to knock down gene function and observe phenotypic differences in *Xenopus laevis*. Importantly, each gene is highly conserved between humans and *Xenopus laevis*, making it an ideal model organism to understand the developmental mechanisms of this genetic deletion. Here, we show phenotypic differences of the 16p12.1 deletion by quantifying changes in craniofacial morphology, cartilage morphology, brain morphology, and axon outgrowth patterns of each individual gene knock down. We have found that reduction of three of these genes, *POLR3E*, *UQCRC2*, and *C16orf52* leads to severe craniofacial and cartilage defects. Additionally, we have found that reduction of *C16orf52* also decreases axon length, and that knockdown of *POLR3E* and *C16orf52* causes smaller brain size. Together, our results suggest that several of the 16p12.1 genes do play an important role during neurodevelopment and future studies will focus on the effect of combinatorial knockdown of these genes, as well as the mechanisms by which they modulate aspects of craniofacial morphology, brain morphology, and axon outgrowth patterns.

2062T**Linking chromatin biology to autism neuropathology with ASXL3.** S.E.

Marlow¹, Y.C. Tsan¹, A. Srivastava¹, R. KC¹, B.T. McGrath², S.L. Bielas¹. 1) Department of Human Genetics, University of Michigan, Ann Arbor, MI; 2) Cellular and Molecular Biology Graduate Program, University of Michigan, Ann Arbor, MI.

Human genetics studies aimed at understanding the genetic etiology of autism spectrum disorders (ASD) implicate chromatin as a nexus of convergent neuropathology, with genetic variants enriched in gene networks linked to chromatin regulation. In particular, histone modifications contribute to transcriptional plasticity pivotal to differentiation and acquisition of new cellular identities. Mono-ubiquitination of histone H2A (H2Aub1) is a reversible, transcriptionally repressive histone modification. We identified *de novo* dominant truncating variants in ASXL3 as the genetic basis of Bainbridge Ropers Syndrome (BRS) and ASD, characterized by failure to thrive, global developmental delay, feeding problems, hypotonia, profound speech deficits, and intellectual disability. ASXL3 is a component of the polycomb repressive deubiquitinase (PR-DUB) complex which deubiquitinates H2A. We identified dysregulation of H2Aub1 as a key molecular pathology in primary cells derived from individuals with BRS, a finding we have confirmed in mouse and human neural progenitor cells (NPCs). 3D human cerebral organoids histologically and molecularly recapitulate the context-dependent features of *in vivo* cortical development, allowing prominent characteristics of early corticogenesis to be reproducibly evaluated. We generated cerebral organoid models from CRISPR-edited ESC and iPSC lines to investigate ASXL3-dependent neuropathology. We observe ASXL3-dependent defects in neural differentiation of pluripotent cells to the full spectrum of mature cortical neuron subtypes, confirming post-mortem evidence that ASD can result from cytoarchitectural disorganization of the cerebral cortex. We utilize transcriptomic and epigenomic techniques to probe the role of ASXL3-dependent H2Aub1 deubiquitination in regulating transcriptional profiles critical to NPC fate decisions during corticogenesis, providing molecular insights gained by functional investigation of genetic variants and uncovering the elusive role of H2Aub1 in neural development.

2063F

A novel homozygous splice site donor alteration in NT5C2 gene leading to spastic diplegia cerebral palsy, developmental delay and microcephaly. M.I. Naseer¹, A.A. Abdulkareem², A.M. Abuzenadah¹, A.G. Chaudhary^{1,4,5}, M.K. Alwasayah^{1,2}, C. Trujillo³, M.H. Al-Qahtani¹. 1) King Abdulaziz University, CEGMR, Jeddah, Saudi Arabia; 2) Aziziah Maternity & Children Hospital, Jeddah, Saudi Arabia; 3) Erfan & Bagedo Hospital, Genetics Unit, Jeddah, Saudi Arabia; 4) Center for Innovation in Personalized Medicine, King Abdulaziz University, 21589, Jeddah, Kingdom of Saudi Arabia; 5) Faculty of Applied Medical Sciences, King Abdulaziz University, 21589, Jeddah, Kingdom of Saudi Arabia.

NT5C2 gene (OMIM: 600417) encode a hydrolase enzyme 5'-Nucleotidase, Cytosolic II play an important role in maintaining the balance of purine nucleotides and free nucleobases in the spinal cord and brain. In this study we report a novel homozygous splice site donor alteration in NT5C2 gene leading to spastic diplegia cerebral palsy, developmental delay and microcephaly. We have identified a large consanguineous Saudi family segregating developmental delay, spastic diplegia cerebral palsy and intellectual disability along with microcephaly. Whole exome sequencing was performed for the affected members of the family to study the novel mutation. Whole exome sequencing data analysis, confirmed by subsequent Sanger sequencing validation, identifies a homozygous splice site donor alteration of possible interest in NT5C2 (ENST00000343289: c.539+1G>T). The mutation was ruled out in 100 unrelated healthy controls. The homozygous mutation detected in this study has not yet been reported as pathogenic in literature or variant databases. In conclusion, the here first time detected homozygous mutation in the intronic region of NT5C2 gene in Saudi family further explain the possibility that NT5C2 gene may play important role and essential for multiple aspects of normal human neurodevelopment.

2064W

Baraitser Winter cerebrofrontofacial syndrome patient iPSC derived cerebral organoids show abnormal neuronal development. I. Niehaus¹, K. Sekeres², K. Guan², A. Haase³, G. Goehring³, E. Schrock³, W. Huttner⁴, M. Heide⁴, N. Di Donato¹. 1) Institute for Clinical Genetics, Faculty of Medicine Carl Gustav Carus, TU Dresden, Germany; 2) Institute of Pharmacology and Toxicology, Technische Universität Dresden, Dresden, Germany; 3) Leibniz Research Laboratories for Biotechnology and Artificial Organs (LEBAO), Department of Cardiac, Thoracic, Transplantation, and Vascular Surgery, Hannover Medical School, Hannover, Germany; 4) Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany.

Baraitser Winter cerebrofrontofacial syndrome (BWCF) is caused by different mutations in one of the two cytoplasmatic actin genes *ACTB* and *ACTG1*. BWCF patients show multiple physical disabilities as well as cortical malformations, e.g. pachygyria. Pachygyria is considered to be a neuronal migration disorder, however the precise pathomechanism remains unknown. Therefore, we aim to assess neuronal development in cerebral organoids differentiated from BWCF patient derived induced pluripotent stem cells (iPSC). For generating cerebral organoids, we used two iPSC lines from two different patients (Patient 1: NM_001101.3 (*ACTB*): c.359C>T, p.(Thr120Ile), and Patient 2: NM_001614.3 (*ACTG1*): c.608C>T, p.(Thr203Met)). As control we used the commercially available iPSC lines SC102A-1 (System Biosciences), 409B2 (RIKEN BRC cell bank). The differentiation of iPSC to organoids was performed according to the self-patterning Lancaster *et al.*, 2013 protocol to generate cerebral organoids as well as the pre-patterning Qian *et al.*, 2016 protocol to generate forebrain-specific organoids. Additionally we applied a directed differentiation to obtain forebrain-type organoids. After 30 days in culture, organoids were fixed, cryosectioned and stained for an immunocytochemical analysis. All BWCF embryoid bodies (EBs) showed abnormal neuroectoderm formation in comparison to EBs derived from the control iPSC lines. Accordingly, the organoids of patient derived iPSC showed macroscopically abnormal ventricle formation and morphology. Sox2 and Pax6 positive progenitor cells were significantly reduced indicating premature neuronal differentiation with an enrichment of cells in advanced stages of differentiation. Intriguingly, iPSC-derived organoids from patients with another classical neuronal migration disorder - Miller-Dieker syndrome - also demonstrated considerably more complex developmental mechanisms beyond abnormal migration. We will present results of an extended immunocytochemical analysis for neuronal differentiation, proliferation and apoptosis, as well as transcriptome analyses of the EBs at neural induction phase. Isogenic control organoids will be generated using CRISPR repair of the actin missense mutations in both patient derived iPSC lines.

2065T

NUP50 biallelic pathogenic variants as a novel genetic cause of neurodevelopmental disorder? S. Moutton^{1,2}, AL. Bruehl³, C. Coubes³, A. Vitobello³, F. Tran-Mau-Them², A. Sorlin^{1,2}, S. Nambot^{1,2}, Y. Duffourd³, C. Philippe², L. Faivre^{1,2}, C. Thauvin-Robinet^{1,2}. 1) Reference Center for Developmental Anomalies, Department of Medical Genetics, Dijon University Hospital, Dijon, France; 2) INSERM U1231, LNC UMR1231 GAD, Burgundy University, F-21000, Dijon; 3) Department of medical genetics, Hôpital Arnaud de Villeneuve, Montpellier University Hospital.

We investigated by whole exome sequencing (WES) a proband who presented with intellectual disability (ID), attention deficiency, seizures and vermiform hypoplasia. Solo clinical WES analysis focusing on known disease-causing gene was negative. Reinterpretation in a research point of view allowed to highlight 2 variants within *NUP50* gene. This gene had previously been hypothesized to be associated with ID in a consanguineous North-African family where two sisters were affected by non-syndromic ID. A homozygous c.502C>T (p.R168W) missense variant had been identified within the largest region of identity by descent. This variant remains absent from ExAC and gnomAD databases and is predicted to be probably damaging. Our patient was found to carry a truncating variant absent from databases at the eighth codon inherited from the mother with a subsequent potential translation initiation codon at the 29th codon, and a missense variant p.Asn371Ser inherited from the father (2 occurrences in gnomAD). Prediction scores are conflicting. International datasharing, including genematcher, was non-contributory. Nucleoporins have been involved in various human diseases including some immune, viral, neurological, cardio-vascular and cancer diseases but *NUP50* is not associated with any of them. *NUP50* (or *NPAP60*) encodes for a component of the nuclear pore complex (NPC). It has also been shown to be located in the nucleoplasm and to regulate developmental gene expression in *Drosophila*. *NUP50* is expressed in various human tissues (brain, liver, lung, kidney, and testis). Homozygous knock-out mice present with neural tube defects and intrauterine growth retardation and die *in utero*. Two functional isoforms of *NUP50* exist in humans, *NUP50-L* (1-469 aa) also present in other organisms whereas *NUP50-S* (29-469 aa) is specifically expressed in humans. Both isoforms are affected by the missense variants. Instead, the frameshift variant only affects the longest isoform. *NUP50-S* was shown to stabilize the binding of importin- α to classical Nuclear Localization Signal-cargo, while *NUP50-L* promotes the release of importin- α from the NLS-cargo. Therefore, these variants could create an imbalance between these antagonizing roles of *NUP50* isoforms at the NPC and/or disrupt gene expression regulation. Additional clinical observations are required to better characterize the potential role of *NUP50* in neurodevelopmental disorders. An international collaborative study would be useful.

2066F

Biallelic mutations in the death domain of PIDD disrupt interactions between PIDD and CRADD and cause intellectual disability. *T.I. Sheikh¹, N. Vasli², K. Kharizi⁴, R. Harripaul^{1,3}, Z. Fattahi⁴, A. Mir⁴, M. Ayub^{6,7}, H. Najmabadi^{4,8}, J.B. Vincent^{1,3,9}.* 1) Center for Addiction & Mental Health, Toronto, ON, Canada; 2) Department of Pediatric Laboratory Medicine, Hospital for Sick Children, Toronto, Ontario, Canada; 3) Institute of Medical Science, University of Toronto, Toronto, Ontario, Canada; 4) Genetics Research Center, University of Social Welfare and Rehabilitation Sciences, Tehran 19857, Iran; 5) Human Molecular Genetics Lab, Department of Bioinformatics and Biotechnology, FBAS, International Islamic University, Islamabad, Pakistan; 6) Lahore Institute of Research & Development, Lahore, 51000 Pakistan; 7) Department of Psychiatry, Queen's University, Kingston, Ontario, K7L 3N6 Canada; 8) Kariminejad-Najmabadi Pathology and Genetics Center, Tehran 14667, Iran; 9) Department of Psychiatry, University of Toronto, Toronto, ON M5T 1R8, Canada.

PIDD1 encodes p53-Induced Death Domain protein (PIDD), which has a key role in DNA damage/stress response, acting as a cell-fate switch, either through interaction with pro-survival molecule, RIP1 and activation of NF- κ B signaling, or via interacting with a pro-death factor, CRADD, and activation of caspase-2 and apoptotic cell death. Biallelic truncating mutations in CRADD have been reported in intellectual disability (ID), and in a form of lissencephaly. Here, we identified four families with ID from Iran and Pakistan with three different biallelic mutations in PIDD, all potentially impacting the Death Domain, through which PIDD interacts with CRADD and other DD proteins such as RIP1 or CRADD/RAIDD. A nonsense mutation, Gln863*, in two apparently unrelated Pakistani families directly truncates the death domain (DD);, a missense mutation in DD, Arg815Trp, is also predicted to be damaging. A splice acceptor mutation in the fourth family is predicted to disrupt splicing upstream of the DD. In HEK293 cells, we showed that, while WT PIDD co-localizes with CRADD in the cytoplasm, both Gln863* and Arg815Trp not only prevent PIDD/CRADD co-localization, but also cause mis-localization of CRADD. Also both Gln863* and Arg815Trp fail to precipitate CRADD from co-expressing HEK293T cells. Together, this evidence along with identification of biallelic mutations in CRADD in ID points towards a significant role for PIDD-CRADD interactions, and the PIDDosome in general, in normal human neurodevelopment.

2067W

Characterizing the phenotypic presentation of an international sample of children with 48, XXY and the impact of testosterone replacement therapy (TRT). *P. Lasutschinkow¹, C. Samango-Sprouse^{1,2,3}, G. Harmon^{5,1}, F. Mitchell¹, M. McLeod¹, A. Gropman^{4,1}.* 1) The Focus Foundation, Davidsonville, MD; 2) George Washington University, Washington, DC; 3) Florida International University, Miami, FL; 4) Children's National Health System, Washington, DC; 5) Nemours Pediatrics, Wilmington, DE.

48, XXY(48s) occurs in 1:50,000 male births & is a more severe variant of 47, XXY (Klinefelter syndrome). Clinical features may include delayed speech, androgen deficiency, gait disturbance & moderate intellectual disability. Physical characteristics include hypertelorism, epicanthal folds & upslanting palpebral fissures. Previous studies indicated that IQ averaged 40-60 with a range between 20-79. We characterize the phenotypic profile of the largest cohort of 48, XXY to date & report on the influence of TRT. An international cohort of 23 48s under the age of six had multidisciplinary evaluations in neurocognition, motor & speech. A subset of 48s had previously been treated with TRT. Language development was assessed with the Preschool Language Scale (PLS-5). Intelligence was measured using the Weschler Intelligence Scale (WISC-IV) & Leiter International Performance Scale (LIPS-III). Behavior was assessed through questionnaires completed by caregivers, including the Behavior Rating Inventory of Executive Function (BRIEF-2) & Child Behavior Checklist (CBCL). 48s presented with an increase of childhood apraxia of speech, immunologic difficulties & motor planning deficiencies. Dysmorphic features included clinodactyly, cryptorchidism, tremors & radioulnar synostosis. In speech, untreated and treated 48s showed no significant differences in auditory comprehension with scores ranging from 85-118. In expressive communication, there was a significant difference between the groups ($P=0.045$) with an average of 82.17 in untreated boys and 98.67 in treated boys, with scores spanning from 67-117. On the LIPS-III, Brief IQs ranged from 49-98, with an average of 84. On the WISC-IV, there was only one untreated boy compared to 7 treated boys. The treated boys performed better in Verbal IQ (90.67 vs 81) and Performance IQ (104 vs 94) with respective ranges of 69-103 and 84-138. CBCL and BRIEF-2 showed increased difficulties in withdrawn, pervasive developmental delays, shift & flexibility. This study reports on a novel international sample of 48s, which is the largest cohort to date. The variability in the group demonstrates a large range of capabilities & outcomes and a possible improvement in the treated sample. 48s had deficits in planning, speech & motor function similar to but more complex than 47, XXY. These results show a much more capable developmental profile than previously appreciated & possible improvement in selected developmental domains with TRT.

2068T

Role of RIPK3-mediated necroptosis on the pathogenesis of incontinentia pigmenti. *M.V. Ursini¹, M.G. Cirigliano¹, A. El-Sharkawy¹, F. Fusco¹, M.B. Lio², A. Pescatore¹.* 1) CNR, Institute of Genetics and Biophysics, Naples, Italy; 2) Università della Basilicata, Potenza, Italy.

Incontinentia pigmenti (IP) is a X-linked neuroectodermal disease. In IP, dermatologic manifestations appearing at birth are essential for the clinical diagnosis. One third of patients show neurological symptoms. The only causative gene of IP is NEMO (also called IKBKG), encoding for the regulatory subunit of the I κ B-kinase (IKK) complex. The current mechanistic model suggests that NEMO functions as a crucial component mediating the recruitment of the IKK complex to the tumor necrosis factor receptor 1 (TNF-R1), thus allowing activation of the pro-survival NF- κ B response. Accordingly, NEMO deficiency impairs NF- κ B activation and induces the pro-apoptotic stimuli through the cytoplasmic FADD-containing complex. Recently, we discovered as an IP-NEMO mutant unable to activate NF- κ B in response to TNF, partially retains NEMO antiapoptotic function and licenses cells to die through necroptosis by inhibiting the recruitment of RIPK3 to the FADD-containing complex following TNF/zVAD stimulation. Importantly, targeting the kinase domain of RIPK1 with necrostatin 1 inhibits the interaction between RIPK1 and RIPK3 and blocks necroptosis in these settings. Indeed, we evidence a correlation between the absence of NEMO and the protein expression level of RIPK3 in cells and affected tissues derived from mice model of IP. Our study evidence as de-regulation of NEMO in different ways can have a profound impact on whether cells die by apoptosis or necroptosis, underlining the importance of address the role of TNF-induced cell death in the pathogenesis of IP.

2069F

TAF1's association with X-linked intellectual disability: Essential neuro-specific function established by zebrafish knockout. *S. Gudmundsson¹, M. Wilbe², B. Filippek-Górnio², A.M. Molin¹, S. Ekvall¹, J. Johansson¹, H. Gylje³, V.M. Kalscheuer⁴, J. Ledin², G. Annerén¹, M-L. Bondeson¹.* 1) Department of Immunology, Genetics and Pathology, Uppsala University, Science for Life Laboratory, Uppsala, Sweden; 2) Department of Organismal Biology, Genome Engineering Zebrafish, Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 3) Department of Paediatrics, Central Hospital, Västerås, Sweden; 4) Research Group Development and Disease, Max Planck Institute for Molecular Genetics, Berlin, Germany.

The TATA-box binding protein associated factor 1 (TAF1) protein is a key unit of the transcription factor II D complex that serves a vital function during initiation of transcription. The *TAF1* gene has repeatedly been highlighted in disease biology, recently associated with intellectual disability. *TAF1* has been detected in elevated levels in mice during the first three weeks of development suggesting an important function during early embryogenesis, but no suitable animal models of orthologues to mutant *TAF1* genes have so far been presented, hence its molecular function remains elusive. We present clinical, genetic, and molecular results of a five-generation family affected with a *TAF1* c.3568C>T, p.(Arg1190Cys) variant, which co-segregated with the phenotype. We report six hemizygous affected males that presented with intellectual disability and dysmorphic features such as long face, pointed chin, large hands, prominent forehead, short neck, low-set, protruding, and large ears, a long philtrum, prominent supraorbital ridge, deep-set eyes, and large feet. Five heterozygous females were asymptomatic and had completely skewed X-chromosome inactivation. We investigate *taf1* function further by creating the first complete knockout model of *TAF1* orthologue using CRISPR/Cas9 system in zebrafish. A crucial role of proper human *TAF1* function during early developmental processes can be inferred from the knockout model demonstrating that an intact *taf1* is essential for life from early embryonic development and onwards. Transcriptome sequencing was performed on *taf1*^{-/-} zebrafish embryos 3 days post fertilization. Pathway enrichment analysis of differentially expressed genes revealed enrichment for pathways associated with neurodevelopmental processes, such as chromatin and DNA assembly and ion transport. In conclusion, this work strengthens the molecular understanding of pathogenic *TAF1* variants causing syndromic X-linked intellectual disability in humans. We describe the clinical and genetic effect of the pathogenic *TAF1* c.3568C>T, p.(Arg1190Cys) in a large family. The zebrafish model suggests that *TAF1* is essential for life and that pathogenic *TAF1* variants cause syndromic X-linked intellectual disability in humans by dysregulation of genes essential for normal neurodevelopment.

2070W

TAF1 mutations in kinase domains associated with reduced RAP74 phosphorylation in patients with X-linked intellectual disability. M. Han¹, S. Srikanth², R. Wang¹, C. Skinner², T. Niranjan¹, R. Stevenson², C. Schwartz², T. Wand¹. 1) Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD; 2) Greenwood Genetic Center, Greenwood, SC.

TATA box binding protein-associated factor 1 (TAF1) is the largest subunit of transcriptional initiation complex, TFIID. *TAF1* mutations have recently been found in patients with X-linked intellectual disability and dystonia-parkinsonism syndromes. The mechanisms responsible for the cognitive and neurological phenotypes in the patients remain unknown. *TAF1* encodes a protein serine kinase that selectively phosphorylates the basal transcription factor, RAP74. Both N-terminal and C-terminal kinase domains are important for efficient transphosphorylation of serine residues. We report two novel *TAF1* missense mutations that segregate with intellectual disability phenotypes in two XLID families. One mutation, p.M21L, is located in the N-terminal domain and the other, p.Q1428P, in the C-terminal domain. Both mutations involve highly conserved amino acids residues and are predicted to have a deleterious effect on protein function. We studied RAP74 phosphorylation in lymphoblast cell lines from patients and normal controls using a phosphorylation site-specific antibody for RAP74. The results showed a significant reduction in RAP74 phosphorylation in lymphoblasts from one affected male carrying the p.M21L mutation and two affected males carrying the p.Q1428P mutation as compared to lymphoblasts from three normal males. RNAseq and quantitative RT-PCR analysis are being conducted on these cell lines to identify signaling proteins and neural pathways that are affected by the TAF1-mediated RAP74 phosphorylation defects. Our results suggest that both N- and C-terminal kinase domains are important for RAP74 phosphorylation and reduced RAP74 phosphorylation contribute to intellectual disability and congenital malformations in patients with *TAF1* deficiency.

2071T

Dual diagnosis of *CLTC* and *DNMT3A* mutation: DeMari syndrome and Tatton-Brown-Rahman syndrome (TBRS). Z. Mohamed¹, C. Mroske², S. Tang², J. Nimeh³, M. Byler⁴, N. Dosa³, J. Kalish⁴, R.R. Lebel¹. 1) Section of Medical Genetics, SUNY Upstate Medical University, Syracuse, NY; 2) Ambry Genetics Corporation, Aliso Viejo, California; 3) Department of Pediatrics, SUNY Upstate Medical University, Syracuse, NY; 4) Clinical Genetics Center, Children's Hospital of Philadelphia, Philadelphia, PA.

Our patient was born at 35 weeks of gestation to a 27-year-old Caucasian woman, after pregnancy complicated by pre-eclampsia. There was no exposure to a suspected teratogen, and the union is not known to be consanguineous. The father was also 27 years old at that time. Neonatal period was notable for hypoglycemia, apnea, bradycardia, hyperbilirubinemia, Grade I intraventricular hemorrhage, subdural hematoma, laryngomalacia and hypotonia with feeding difficulties. She had several minor dysmorphic features and evolved global developmental delays. Whole-exome sequencing (WES) revealed a *de novo* mutation (c.2737_2738dupGA; p.D913EFs*59) in *CLTC* gene. The *CLTC* gene encodes clathrin heavy chain 1, one of the subunits of clathrin. Clathrin participates in vesicle formation. The synapses of neurons rely on clathrin vesicles to release and recycle neurotransmitters. Further, during mitosis vesicle formation is suspended and clathrin then plays a role in stabilizing spindle fibers. The gene function in neuronal transmission is consistent with the patient's neurological features. Animal models with the *CLTC* gene disrupted show a defect in neurotransmission at the synapses. At 26 months of age, the patient was also found to have a ganglioneuroblastoma. Imaging revealed interval growth of the tumor. She underwent surgical debulking. Next generation sequencing of the tumor revealed a variant (c.2645G>A; p.R882H) in the *DNMT3A* gene. Reanalysis of the exome data confirmed the variant to be germline and likely pathogenic. Germline mutations in *DNMT3A* causes Tatton-Brown-Rahman syndrome (TBRS). TBRS is an autosomal dominant condition characterized by tall stature, intellectual disability, macrocephaly, round face. Neuroblastoma has not been reported previously in a patient with *DNMT3A* mutation. The present patient has been reported previously when we were conscious of only her *CLTC* mutation.

2072F

Correction of the reduced GluA2 and TIMP2 in *Fmr1* KO mice with adult restoration of *Fmr1*. R. Zong, D. Nelson. Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030.

Loss of *FMR1* function results in fragile X syndrome (FXS), a common form of intellectual disability. *FMR1* encodes an RNA binding protein that regulates translation of its targets. FXS appears to result from dysregulated translation, and elevated synthesis of several proteins has been described in the brains of mice lacking *Fmr1*. *Fmr1* protein (FMRP) is also involved in RNA transport and stability, regulation of the miRNA pathway and chromatin remodeling. Two Fmrp targets identified from the RNA and protein expression assays in brains of mouse *Fmr1* KO models showed greatly reduced expression, *GluA2*, the AMPA receptor subunit 2, and *Timp2*, the tissue inhibitor of MMP2. Both are interesting as potential therapeutic targets in fragile X syndrome. *GluA2* has a role in regulation of dendritic spine development, and *Timp2* has been shown to restore memory in aged mice. *GluA2* expression (mRNA and protein) is greatly reduced in *Fmr1* KO mouse brain. Fmrp interacts with miR-124 in a co-immunoprecipitation assay, and miR-124 directly targets *GluA2* mRNA. Consistent with the observation that *GluA2* expression level was reduced, the level of miR-124 was found to be increased in *Fmr1* KO mice in a quantitative PCR assay. We show that this reduction of *GluA2* can be rescued with adult restoration of Fmrp in *Fmr1* cON-CreER mice after tamoxifen induction indicating that reduction is closely associated with the Fmrp expression. In order to model potential gene therapy for Fragile X syndrome, we have developed a pAAV-*Fmr1* viral construct to restore Fmrp in adult *Fmr1* KO mice to determine whether reduced *GluA2* and TIMP2 levels in *Fmr1* KO mice can be corrected. The pAAV-*Fmr1* vector was shown to have a good expression of Fmrp in cell cultures, and the AAV viral particles have been produced for mouse injection. Mice will be assessed for neuronal and behavioral phenotypes after *Fmr1* restoration via viral gene transfer. Overall, we propose that a balance correction both up-regulated and down-regulated Fmrp targets could be a key to be successful in the therapeutics treatment for FXS.

2073W

Beyond a molecular diagnosis: Functional studies of AHDC1 in Xia-Gibbs syndrome. M.M. Khayat^{1,2}, J. Traynelis^{1,2}, Q. Meng^{1,2}, V. Chander^{1,2}, H. Li^{1,2}, A.C. Salvador⁶, Y. Jiang^{1,2}, M.F. Wangler^{2,3}, D.R. Murdock^{1,2}, A. Sabo^{1,2}, J.E. Posey², J.R. Lupski^{1,2,4,5}, R.A. Gibbs^{1,2}. 1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas; 3) Department of Pediatric Radiology, Texas Children's Hospital, Houston, TX, 77030; 4) Department of Pediatrics, Baylor College of Medicine, Houston, TX, 77030; 5) Texas Children's Hospital, Houston, TX, 77030; 6) Texas A&M University Health Science Center, College Station, TX, 77843.

The effect of mutations on gene function have been difficult to characterize in neurodevelopmental disorders (NDDs) due to underlying genetic and phenotypic heterogeneity. Thus, each individual case warrants thorough genotype/phenotype analyses and follow-up functional studies. Xia-Gibbs syndrome (OMIM: 615829: XGS) is a recently described severe developmental disorder characterized by hypotonia, limited speech and sleep apnea (Xia et al, 2014). The mechanism of pathogenesis in XGS is not known and may involve either a dominant gain of function or else a haploinsufficient model of disease. This disorder is emblematic of many NDDs where patients display a spectrum of phenotypes with a driving *de novo* autosomal dominant truncating mutation, in a gene where little is known about the normal function. To advance XGS studies we have recruited more than 30 XGS families and have verified their molecular diagnoses. The phenotypes and molecular data have been collected in an XGS Registry permitting deeper analysis. In addition, blood samples have been stored in a biobank for future research. Homology analyses have shown similarities to proteins (REV3L and KIAA2022) implicated in DNA repair and neurological development. Specifically, 4 domains share homology with AHDC1 and 3 AT-Hook DNA binding domains are observed. Protein expression has been achieved in insect cells and commercial antibodies have been validated for detection of AHDC1. Transient expression of cloned *AHDC1* in human embryonic kidney cells shows nuclear localization of the wildtype protein with increased expression in the nucleoli. Studies of deletion breakpoints in individual patients and patterns of RNA expression are underway. In addition to elucidating disease mechanism in XGS, these studies will inform our general understanding of autosomal dominant NDDs and their underlying causes.

2074T

Functional analysis of *MED13L* missense variants. J. Ghomid^{1,3}, M. Billot-tee¹, F. Frenois¹, R. Caumes^{1,3}, F. Petit^{1,3}, C. Thuillier², C. Roche-Lestienne², S. Manouvrier-Hanu^{1,3}, P. Volkel⁴, P.O. Angrand⁵, T. Smol^{1,5}. 1) EA 7364 RADEME - University of Lille, Lille, France; 2) Institut de Génétique Médicale, Hôpital Jeanne de Flandre, CHU Lille, Lille, France; 3) Service de Génétique Clinique, Hôpital Jeanne de Flandre, CHU Lille, avenue Eugène Avinée, Lille, France; 4) CNRS Lille, Inserm U908, Université de Lille, Bâtiment SN3, Cité Scientifique, 59655 Villeneuve d'Ascq, France; 5) Inserm U908, Université de Lille, Bâtiment SN3, Cité Scientifique, 59655 Villeneuve d'Ascq, France.

Mediator is a large complex conserved from yeast to humans. The complex is a master coordinator of cell lineage determination. In response to various stimuli, mediator undergoes conformational changes and creates a DNA loop between activated enhancer elements and promoter. Mediator physically bridges transcription factors bound at enhancer elements with the RNA polymerase II transcription machinery at core promoter regions. The complex is organized into four modules, namely the tail-, the middle-, the head-, and the CDK8-kinase module. In vertebrates, the latter module is composed of CCNT1 and three additional proteins: CDK8, MED12, and MED13; or their respective paralogs: CDK19, MED12L, and MED13L. Molecular anomalies in *MED13L*, leading to haploinsufficiency, have been reported in patients with moderate to severe intellectual disability and distinct facial features, with or without congenital heart defects. Until recently, clinical relevance of *MED13L* missense variants was unclear. We identified 7 de novo missense variants, clustered in mutational hot-spots, *i.e.* exons 15-17 and 25-31. We noted that patients with a variant located in exon 15 showed a more severe phenotype and were more at risk for epilepsy. In silico analyses suggested that variants located in exons 25 to 31 impair potential ubiquitination sites and that variants located in exon 15 to 17 impair potential phosphorylation sites. Through FBXW7-dependant ubiquitination assays, phosphorylation assays, MED13L subcellular localization analysis and RNA rescue assays in zebrafish, we propose to unravel molecular consequences of *MED13L* missense variants. This work gives clues to understand molecular basis of genotype-phenotype correlations observed in the patients.

2075F

A new *SMN1* variant (c.835G>C, p.Gly279Arg) in YG box domain causing exon 7 skipping and is responsible for spinal muscular atrophy. JL. Bai, YJ. Qu, F. Song, YY. Cao, MM. Chen, YW. Jin, H. Wang. Capital Institute of Pediatrics, Bei Jing, China.

Background: Spinal muscular atrophy (SMA) is an autosomal recessive neuromuscular disorder caused by deletion or subtle mutation of *SMN1* (survival motor neuron 1). The highly conserved YG box domain encoded by exons 6 and 7 is important for SMN binding to other proteins and for SMN self-association. **Methods:** We detected the *SMN1* subtle mutations in two non-homozygous patients by MLPA, genomic sequencing and T-A cloning on cDNA level. Then quantitative real-time PCR (qRT-PCR) was used to measure the effect of the *SMN1* variant on levels of full-length *SMN1* (fl-*SMN1*) transcripts in patient's blood and also in immortalized lymphocytes. The fragments of RT-PCR were sequenced to confirm whether it exits exon 7 skipping. To further evaluate its splicing effects on exon 7, we constructed mutant *SMN1* minigene with c.835G>C and a positive control splicing minigene including the classical splicing site c.835-1G>A in the *SMN1*-Short plasmid. After transfecting them in HEK293 cells respectively, we measured the percentage of fl-*SMN1* transcripts. **Results:** We identified two novel variants in two SMA patients: one with c.835G>C(p.Gly279Arg) in exon 7 and the other with c.211del A(p.Lys71Lysfs*78) in exon 2b of *SMN1* gene. For the c.835G>C, fl-*SMN1* transcripts was significantly decreased both in the peripheral bloods and in the immortalized lymphocytes of the patient ($p < 0.05$). And we observed two fragments by RT-PCR, and part exon 7 skipping was confirmed by direct sequencing. In ex vivo splicing analysis, it showed that positive control splicing minigene yielded 100% exclusion of *SMN1* exon 7 and the mutant minigene reduced its inclusion to 54%. **Conclusion:** We reported two novel *SMN1* mutations in two SMA patients. And our study clearly demonstrates that the c.835G>C(p.Gly279Arg) missense variant can lead to partly exon 7 skipping of *SMN1* transcripts. What's more, it emphasizes that the *SMN1* missense variant (c.835G>C) in the YG box domain does influence the correct splicing of exon 7 and is responsible for spinal muscular atrophy. **Acknowledgments:** This study was supported by National Natural Science Foundation of China (Nos. 81501850, 81100933, 81470056, 81500979), Beijing Municipal Administration of Hospitals' Youth Program (QML20161303), National Key Research and Development Program of China (2016YFC0901505), CAMS Initiative for Innovative Medicine (CAMS-I2M), Beijing Natural Science Foundation (No.7172039). *Corresponding author: F Song.

2076W

Complex genetic etiology of congenital diaphragmatic hernias, a common structural birth defect.

E. Bogenschutz¹, Z. Fox¹, R. Layer^{1,2}, A. Farrell^{1,2}, J. Wynn³, B. Moore^{1,2}, L. Yu³, G. Aspelund⁴, A. Quinlan^{1,2,5}, G. Marth^{1,2}, M. Yandell^{1,2}, Y. Shen^{6,7,8}, W. Chung^{3,9,10}, G. Kardon¹. 1) Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, UT; 2) USTAR Center for Genetic Discovery, University of Utah School of Medicine, Salt Lake City, UT; 3) Department of Pediatrics, Columbia University, New York, NY; 4) Department of Surgery, Columbia University Medical Center, New York, NY; 5) Department of Biomedical Informatics, University of Utah, Salt Lake City, UT; 6) Department of Systems Biology, Columbia University, New York, NY; 7) Department of Biomedical Informatics, Columbia University, New York, NY; 8) JP Sulzberger Columbia Genome Center, Columbia University, New York, NY; 9) Department of Medicine, Columbia University Medical Center, New York, NY; 10) Herbert Irving Comprehensive Cancer Center, Columbia University, New York, NY.

The diaphragm is an essential mammalian skeletal muscle, as it is critical for respiration and serves as a barrier between the thoracic and abdominal cavities. Defects in the diaphragm are the cause of a common (1 in 3000 births) birth defect, congenital diaphragmatic hernia (CDH). In CDH, weaknesses in the developing diaphragm allow abdominal contents to herniate into the thoracic cavity and impairs normal lung development. The resulting lung hypoplasia and pulmonary hypertension are the most significant contributors to the 50% neonatal mortality. The genetic etiology of CDH is complex and poorly understood. More than 100 genes have been implicated, but all genes are incompletely penetrant. Using mouse genetics, we have demonstrated (e.g. Merrell *et al.* 2015) that several CDH-implicated genes, including *GATA4*, are specifically required in the muscle connective tissue fibroblasts of the early developing diaphragm, and our data suggest that somatic mutations in these fibroblasts may play a role in CDH etiology. To test this hypothesis and probe more deeply into the etiology of CDH, we performed whole genome sequencing in multiple tissues from four CDH families: blood, skin, and diaphragm connective tissue from the proband, along with maternal and paternal blood. We have developed a pipeline to identify structural variants, insertions/deletions, and single nucleotide variants in coding and non-coding regions that could contribute to CDH etiology. Genes with variants identified with this pipeline were prioritized based on expression levels within mouse developing connective tissue fibroblasts, and we identified 18 novel candidate genes, none of which were recurrent across the families. No patients were found to harbor damaging somatic mutations. However, an inherited non-coding deletion was found in two unrelated patients, potentially impacting the function of a putative enhancer of *GATA4*. In summary, our in-depth analysis has revealed that the etiology of CDH is complex, with potentially each CDH patient harboring a unique set of mutations.

2077T

Near infrared light exposure uncouples electron transport from ATP synthesis in epithelial cells. *T.O Oliver, V.B Brown, D.C Coleman, A.J.S Johnson-Stewart.* Biology, Spelman College, Atlanta, GA.

Near infrared red (NIR) light exposure (2.88 J/cm², 632 nm) mediates cytoprotective effects in epithelial cells exposed to a pulse of 2 µm laser radiation. This is similar to the outcomes mediated by exogenous NO-donors. The overall goal of our research is to identify pathways by which red light exposures inhibit apoptosis. Light of various wavelengths and intensities has been found to affect various aspects of cellular respiration. This led us to examine the effects of NIR light on various aspects of cellular respiration, namely mitochondrial polarization, electron transport, and ATP synthesis. We also examined the effect of red light exposure on cGMP, due to its role in inhibiting apoptosis. In our approach, retinal pigmented epithelial (RPE) cells were exposed to NIR light (2.88 J/cm², 632 nm). The compounds TMRE, SNARF-1, and the CellTiter-Glo kit were used to assess various aspects of cellular respiration. The Direct cGMP ELISA by ENZO was used to quantify cGMP levels. Results from NIR light exposed and unexposed cells were compared. Near infrared light exposure was found to be associated with increased electron transport, and the increased pumping of protons into the mitochondrial inner membrane space, which, under normal circumstances are associated with ATP production. Interestingly, we did not find differences in ATP synthesis between exposed and unexposed RPE cells. cGMP levels were found to increase immediately after exposure. This difference was not detected three hours after exposure, but reappeared at the 6 hour time point. Differences in cGMP levels were not detected 24 hours after exposure. These data suggest that NIR light exposure affects cellular respiration by uncoupling electron transport from ATP synthesis, a novel finding.

2078F

Polymorphism rs1052536 in base excision repair gene is a risk factor in a high-risk area of neural tube defects in China. X. Wang, G. Li, X. Wang, Z. Guan, J. Guo, F. Wang, J. Zhang, B. Niu, T. Zhang, J. Yang, J. Wang. Capital Institute of Pediatrics, Beijing, China.

Background: DNA Base Excision Repair Gene-DNA LigaseIII (LIG3) is an important repair gene in the repair pathway and plays an important role in maintaining the integrity of mitochondria. Rs1052536 and rs3135967 polymorphisms of the gene are associated with lung cancer, keratoconus, and Fuchs endothelial corneal dystrophy. There is no previously published report on the relationship between the polymorphisms and neural tube defects (NTDs). **Methods:** Mass ARRAY iPLEX was used to determine the distribution of the polymorphisms in the case group of 108 NTD pregnant women and a control group of 233 normal healthy pregnant women to examine the relevance of their polymorphisms and NTD occurrence. **Results:** The homozygotes of rs1052536 TT were associated with an increased risk for NTDs than CC ($P=0.014$, OR=2.31, 95%CI [1.17–4.54]), and variants of rs1052536 T were associated with an increased risk of NTDs ($P=0.024$, OR=1.50, 95%CI [1.06–2.13]). The stratified analysis showed that TT genotype of rs1052536 increased the risk of anencephaly ($P=0.016$, OR=2.69, 95%CI [1.18–6.10]) and the T allele significantly increased the risk of cranial NTDs ($P=0.033$, OR=1.56, 95%CI [1.04–2.35]). **Conclusions:** Rs1052536 in LIG3 gene might be a potential genetic risk factor in a high-risk area of NTDs in China.

2079W

MA-43: Potential drug for mitochondrial diseases. A. Miyauchi¹, T. Kouga¹, E. Jimbo¹, T. Matsushashi², T. Abe², T. Yamagata¹, H. Osaka¹. 1) Pediatrics, Jichi Medical University, Shimotsuke, Tochigi, Japan; 2) Clinical Biology and Hormonal Regulation, Tohoku University Graduate School of Medicine, Sendai, Japan.

Introduction: Leigh syndrome (LS) and MELAS (myopathy encephalopathy lactic acidosis and stroke-like episodes) are common phenotype in children. However, there is no curative treatment and effective treatment options are highly expected. We examined a chemical library using fibroblasts from patients with LS and MELAS, and identified MA-43 as a potential therapeutic drug for mitochondrial disease. **Methods & Results:** Our study was carried out in four fibroblast cell lines from LS and MELAS patients after obtaining their informed consent. We examined firstly cell viability assay under oxidative stress condition induced by L-buthionine (S, R)-sulfoximine (BSO), a glutathione synthesis inhibitor. Idebenone, an analogue of coenzyme Q10, was used as a positive control. In our chemical library, 3 chemicals were more effective greater than idebenone on all kind of fibroblasts and MA-43 was the most effective among them. Subsequently, we examined mitochondrial bioenergetic functions in LS patient fibroblast under treatment of effective 3 chemicals using Extracellular Flux Analyzer Seahorse XFe 96. Among 3 chemicals, only MA-43 significantly improved all kinds of the oxygen consumption rate (OCR) values; basal OCR, ATP-linked OCR and maximal OCR. In addition, to investigate the mechanism of action, we performed a microarray analysis of ~63,000 genes in MA-43-treated fibroblasts from a patient with Leigh syndrome and we found 475 MA-43-responsive genes. **Conclusion:** We identified that MA-43 rescued fibroblasts from cell death under oxidative stress. MA-43 also improved the mitochondrial respiratory activity. While further studies are needed in order to reveal mechanism of action, MA-43 appears to be potential drug for mitochondrial diseases with the effects of anti-cell death and increase in mitochondrial respiratory activity.

2080T

Transcriptomic analysis of mammalian cell cultures deficient for enzymes involved in the TCA cycle. A.C. Santos de Medeiros, B.H. Graham. Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN.

The tricarboxylic citric acid (TCA) cycle is the central metabolic pathway for energy production and oxidation of carbohydrates, fatty and amino acids. Pathogenic variants in several of the genes encoding enzymes of the TCA cycle have been associated with recessive genetic disorders featuring metabolic and mitochondrial dysfunction. These disorders can manifest a spectrum of neurological, muscular and other phenotypes and the mechanism(s) of pathogenicity are not fully understood. Previous work from our lab demonstrated energy metabolism defects and mitochondrial DNA (mtDNA) depletion in fibroblasts from patients with mutations in either *ACO2* or *SUCLG1* that encodes mitochondrial Aconitase or α -subunit of Succinyl-CoA (SCS), respectively. In addition, our lab developed a mouse deficient for *Sucla2* encodes the ADP-specific β -subunit of SCS. *Sucla2* deficient embryonic fibroblasts also exhibit energy metabolism defects and mtDNA depletion. To further characterize these disorders, our lab performed RNAseq using three genotypes (controls, mutant and mutant rescue lines) for each gene. The main goal of this project is to determine differential expression patterns for each cell line group and compare and contrast among the cell line groups to understand the global transcriptomic effects of these TCA cycle defects. Our preliminary data revealed 3331 significant changes associated with *Sucla2* deficiency where 56% of transcripts were downregulated. In the *SUCLG1* deficient line, 1799 significant changes (37% downregulated transcripts) were detected. In the *ACO2* deficient line, 1947 significant changes (54% downregulated transcripts) were detected. The transcript level of the mutated locus in each cell line was reduced, a result in alignment with the reduced protein levels previously reported. For initial bioinformatics analysis, we used DAVID Software and selected the 5 most enriched functional clusters for each cell line. The majority of the functional groups detected were specifically associated to each line and common functions included secreted pathways, signaling pathways, glycoprotein, and glycosylation. Additional bioinformatics analyses are underway to narrow down the list of candidate genes and unravel pathways either common to all disorders or specific to each disorder. Ultimately, the pathways identified will be used in our models for better understanding of the pathophysiological mechanism(s) of disorders associated with components of the TCA cycle.

2081F

Polymorphism in the *POU4F2* influences the severity of normal tension glaucoma. J. Lee^{1,2}, J. Jeoung³, S. Oh⁴, D. Kim³, J. Ahn⁵, Y. Oh⁵, M. Kim¹, M. Seong¹, S. Park^{1,5}, J. Kim⁵. 1) Department of Laboratory Medicine, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, Republic of Korea; 2) Department of Laboratory Medicine, Hallym University Sacred Heart Hospital, Anyang, Republic of Korea; 3) Department of Ophthalmology, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, Republic of Korea; 4) Department of Biostatistics SMG-SNU Boramae Medical Center, Seoul, Republic of Korea; 5) Biomedical Research Institute, Seoul National University Hospital, Seoul, Republic of Korea.

Introduction: Normal tension glaucoma (NTG) occurs despite normal intraocular pressure. Neurodegeneration of retinal ganglion cells (RGCs) and the ischemic condition are thought to be the main causes increasing the risk of NTG development. Genes mediating RGC development have been recently identified. Here, aimed to investigate the association of polymorphisms in RGC development genes with NTG and clinical characteristics. **Methods:** A total of 435 NTG patients and 419 normal controls were evaluated. We genotyped four single nucleotide polymorphisms (SNPs) in genes responsible for RGC development including *POU4F1* (rs9601092), *POU4F2* (rs13152799 and rs1504360) and *ISL1* (rs2288468). NTG patients were evaluated on visual field defect severity (mean deviation, MD; pattern standard deviation, PSD) and presence of optic disc hemorrhage (DH). The influence of four SNPs on risk of NTG and DH development was analyzed using Pearson's χ^2 test or Fisher's exact test, and multiple logistic regressions. The associations for each SNP with MD and PSD were assessed using multiple linear regression analysis. **Results:** No significant association was observed between candidate SNPs and NTG development. However, minor allele T at rs1504360 in *POU4F2* was strongly associated with lower MD and higher PSD after adjustment for covariates (MD, coefficient -1.363, $P = 0.012$; PSD, coefficient 1.416, $P = 0.001$). In addition, major allele T at rs9601092 in *POU4F1* was significantly associated with DH development: The frequency of the major allele T was 1.65 times higher in the DH group, when compared to no DH group ($P = 0.005$). **Conclusions:** rs1504360 and rs9601092 in RGC development genes, *POU4F2* and *POU4F1*, were associated with the clinical severity of NTG. Carriers of these risk alleles are thought to be predisposed to severe visual field defect and DH development. These findings suggest that *POU4F2* and *POU4F1* variants would be important genetic markers for predicting clinical progression of NTG.

2082W

Ocular biometry measurements using SD-OCT in zebrafish models of myopia. *W.H. Quint¹, A.I. Iglesias^{1,2,3}, M. Meester-Smoor^{1,2}, R. Willemsen³, C.W. Klaver^{1,2,4}.* 1) Department of Ophthalmology, Erasmus Medical Center, Rotterdam, The Netherlands; 2) Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands; 3) Department of Clinical Genetics, Erasmus Medical Center, Rotterdam, The Netherlands; 4) Department of Ophthalmology, Radboud University Medical Center, Nijmegen, The Netherlands.

Purpose: In this study, we aim to establish a method to perform high-resolution eye biometry measurements in zebrafish in order to study myopia-associated candidate genes. We evaluated changes in axial length (AL) using spectral-domain optical coherence tomography (SD-OCT) in a *gjd2* and *lrp2* knock-out (KO) model. In humans, *GJD2* influences the risk of complex myopia while *LRP2* causes syndromic myopia. **Methods:** The *lrp2* and *gjd2* KO fish zebrafish were obtained from the Medical College of Wisconsin and the University of Oregon, respectively. In total, three experimental groups were studied, wild-type (WT), *gjd2*^{-/-} and *lrp2*^{-/-}. Each group consisted of n=40 eyes (i.e., 20 fish per group). SD-OCT was performed in adult zebrafish at 110 days post-fertilization. Fish were anesthetized by standard methods, and eyes were measured using a 900nm Ganymede OCT system (Thorlabs). AL was measured from the front of cornea to the most anterior part of the retina pigmented epithelium. Further, body length (BL) was measured from the tip of the head to the end of the trunk. To normalize the eye metrics, the ratio between the AL and BL (AL:BL ratio) was calculated. Wilcoxon rank sum test (95% CI) was used to evaluate statistical differences between WT and KO fish (i.e., *gjd2* and *lrp2*). Inter-observer variability was tested in a sub-set of *lrp2* and WT fish (n=20 eyes). Furthermore, we used a 7Tesla MRI system to validate the SD-OCT measurements. **Results:** The AL of the *lrp2* mutants was significantly larger than in WT (1.83mm vs 1.36mm, p=9.88e-12). In the *gjd2* mutant we also found a bigger AL (1.41mm vs 1.36mm, p=7.83e-4). Analysis of the AL:BL ratio, confirmed the larger eye phenotype of the *lrp2* (p=2.34e-13). However, in the *gjd2* we observed a smaller AL:BL ratio compared to WT, suggesting a relatively smaller eye (p=1.24e-3). Inter-observer correlation was high (r=0.96, p<2.2e-16), and the difference between OCT and MRI measurements was < 1.5%. **Conclusions:** SD-OCT allows consistent, high-resolution assessment of *in-vivo* eye biometry in zebrafish. We found differences in AL of *lrp2* and *gjd2* mutants compared to WT. Further analyses need to be performed to confirm the effect of *gjd2* in eye size. Given the observed variability in body length among fish lines, careful control of husbandry conditions is essential. Future studies will include other myopia KO models, additional time points and an increased sample size.

2083T

Gjb2, encoding the gap junction protein Connexin26, is required in supporting cells of the organ of Corti in a developmental stage- and cell type-specific manner. *D.M Martin, J.M Skidmore, L.A. Beyer, D.L. Swiderski, L. Kabara, D. Dolan, Y. Raphael.* University of Michigan, Ann Arbor, MI.

Mutations in *GJB2*, encoding Connexin26 (Cx26), are the most common cause of autosomal recessive hereditary deafness. Despite this high prevalence, specific pathogenic mechanisms leading to *GJB2*-related hearing loss are not understood, and there are no curative treatments. Individuals with *GJB2*-related hearing loss exhibit normal organ of Corti morphology and spiral ganglion neurons (SGNs), suggesting the mechanisms of hearing loss do not involve major disruptions in morphology of auditory epithelia or nerves. In contrast, mice with conditional loss of *Gjb2* in auditory epithelial supporting cells (*Sox10Cre;Gjb2^{fllox/lox}*) exhibit profound loss of hair cells and SGNs. Here we used inducible *Sox10Cre^{ERT2}*-mediated loss of *Gjb2* to test whether auditory epithelial integrity and function are sensitive to *Gjb2* loss in supporting cells at various developmental stages (i.e. embryonic, early postnatal, and adult). To delete *Gjb2*, tamoxifen was given to *Gjb2^{fllox/lox}* and *Gjb2^{fllox/lox};Sox10Cre^{ERT2}* littermate mice by gavage at postnatal day 1 (P1) or by intraperitoneal injection at P14. Mice injected at P1 were allowed to mature until P21, and mice injected at P14 were allowed to mature until P21 or P35 to determine if Connexin26 is required for the maintenance of hearing. Hearing was analyzed using ABR audiometry and Distortion Product Otoacoustic Emissions (DPOAE), and ears were collected and analyzed histologically. Deletion of *Gjb2* was confirmed by reduced anti-Connexin26 staining in ears from mice injected at P1 and P14. Interestingly, *Gjb2* deletion at P1 led to coat hypopigmentation, severe to profound hearing loss, and no major changes in hair cell organization. Similarly, *Gjb2* deletion at P14 resulted in no changes in hair cell organization and progressively decreased hearing, with prolonged *Gjb2* loss in supporting cells at later stages (analysis at P35 vs P21). In contrast, *Gjb2* deletion at P14 had no effects on coat color. These results indicate that reduced *Gjb2* dosage has important developmental stage- and cell type-specific effects on both coat color and on auditory epithelial morphology and function. These findings shed light on specific developmental stages and cell types that are critical for Cx26 function. This unique mouse model has a phenotype that more closely mimics human *GJB2*-related hearing loss than germline mutant mice, and may be suitable for designing therapies to correct *GJB2*-related hearing loss in humans.

2084F

Rhesus macaques provide new spontaneous animal models of human inherited retinal diseases. *M. Raveendran¹, R.A. Harris¹, J. Wang¹, S. Thomas², A. Moshiri³, J.A. Roberts⁴, R. Wiseman⁵, N. Artemyev⁶, T. Stout¹, R. Chen¹, J. Rogers¹.*

1) Baylor College of Medicine, Houston, TX; 2) School of Veterinary Medicine, University of California, Davis, CA; 3) School of Medicine, University of California-Davis, Sacramento, CA; 4) California National Primate Research Center, University of California, Davis, CA; 5) University of Wisconsin-Madison, Madison, WI; 6) University of Iowa, Iowa City, IA.

Rhesus macaques (*Macaca mulatta*) are widely used as animal models for the study of many questions regarding human health and disease, including cardiovascular, metabolic and neurological disorders. Here we describe the development of novel rhesus macaque models of inherited retinal diseases (IRDs). IRDs constitute a major cause of progressive vision loss and visual impairment. One gene-based therapy has been FDA-approved for IRD, but much more work is needed to develop effective therapies for the broad scope of IRDs. Most prior animal studies have used mice as the model organism, but results from mouse models often fail to translate to the human situation. The retina and overall visual system of nonhuman primates is much more similar to the human than are non-primate mammals that lack a cone-rich macula. We have sequenced the whole genomes of >550 rhesus macaques and screened another 1200 for a panel of 225 genes known to cause human IRDs. We identified predicted damaging mutations in multiple IRD genes, and have validated with Sanger sequencing damaging mutations in macaque *NPHP4*, *CRB1*, *USH2A*, *ATF6* and *EYS*. In one case, we performed extensive phenotyping (electroretinography, retinal imaging, etc.) on four visually impaired rhesus macaques from the California National Primate Research Center and validated a naturally occurring macaque model of achromatopsia. All four affected animals are homozygous for the same missense mutation in *PDE6C*, a gene known to cause achromatopsia in humans. The macaque mutation changes an arginine to glutamine in the catalytic domain of the protein, and *in vitro* assays confirm that the altered enzyme has dramatically reduced activity. The *PDE6C* model of achromatopsia illustrates the translational value of nonhuman primate models of IRD and creates new opportunities for the testing of gene- or cell-based therapies. On-going studies of identified mutations in other IRD genes suggest that this approach (i.e. screening of primates for damaging mutations in known human disease loci) will produce valuable new genetic models of disorders for which non-primate models are unsatisfactory.

2085W

SPECC1L regulates oral epithelial cell integrity downstream of IRF6 in palatogenesis. *E.G. Hall¹, N.R. Wilson¹, L.W. Wenger¹, J.P. Goering¹, Y. Kousa², D.S. Acevedo³, L. Pitstick³, M. Dunnwald⁴, B.C. Bjork³, B.C. Schutte², I. Saadi¹.*

1) Department of Anatomy and Cell Biology, University of Kansas Medical Center, Kansas City, KS; 2) Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI; 3) Department of Biochemistry, Midwestern University, Downers Grove, IL; 4) Department of Anatomy and Cell Biology, University of Iowa, Iowa City, IA.

Craniofacial anomalies constitute a third of all congenital malformations. Cleft lip with or without cleft palate (CL/P), one of the most common craniofacial birth defects, is a multifactorial disease that manifests in affected families with incomplete penetrance. Pathogenic *SPECC1L* variants identified in patients with rare atypical clefts and syndromic CL/P suggest the gene plays a primary role in face and palate development. To elucidate the role of *SPECC1L* in palatogenesis, we generated null and hypomorphic *Specc1l* deficient mouse alleles. Null mutants are lethal at embryonic day 9.5 (E9.5) with defective neural tube closure and cranial neural crest cell delamination. Hypomorphic *Specc1l*^{Δ200} mutants die perinatally, but do not show cleft palate. We argued that further deficiency of *SPECC1L* dosage may be required. Indeed, *Specc1l*^{Null/Δ200} compound mutants exhibited transient oral adhesions, subepidermal blebbing, and a delay in palate elevation in 60% of embryos at E14.5, when the palate is adhered in all controls. By E15.5, approximately 15% of embryos showed permanent cleft palate. We hypothesized that defects in the oral epithelium are responsible for the ectopic oral adhesions, which result in delayed palate elevation. Immunostaining of mutant palatal shelves revealed abnormalities in the periderm layer, which serves to prevent adhesion between oral epithelia during early embryonic development. We observed increased actin filaments and ectopic apical distribution of adherens junction markers in the oral adhesions of mutant palate periderm. Heterozygous mouse mutants for known clefting genes, *Irf6*, *Grhl3* and *Arhgap29*, also show periderm defects and ectopic oral adhesions. Accordingly, we found that *SPECC1L* expression in the oral epithelium requires IRF6. Loss of IRF6 has been shown to cause increased ROCK1 activation and subsequent actin stress fiber formation through a reduction in Rho-GTPase ARHGAP29. *Specc1l* mutants also show an increase in actin filaments, but with no change in ARHGAP29 and a reduction in ROCK1 levels, indicating that *SPECC1L* acts downstream of IRF6, but in a ROCK1-independent manner. Together, our data show that *SPECC1L* regulates embryonic palate epithelial integrity, and that delayed palate elevation may represent a predisposing factor in this complex disease. Our results extend the understanding of human CL/P etiology by placing *SPECC1L* within the IRF6/GRHL3 pathway, which plays a major role in lip and palate development.

2086T

Warsaw breakage syndrome: Further clinical and genetic delineation. *E. Alkhunaizi*^{1,2}, *R. Shaheen*³, *SK. Bharti*⁴, *AM. Joseph-George*⁵, *K. Chong*¹, *G. Abdel-Salam*⁶, *M. Alowain*⁷, *S. Blaser*⁸, *B. Papsin*⁹, *M. Hashem*³, *N. Martin*¹, *R. Godoy*¹, *R. Brosh Jr.*⁴, *F. Alkuraya*^{3,10,11}, *D. Chitayat*^{1,2}. 1) The Prenatal Diagnosis and Medical Genetics Program, Department of Obstetrics and Gynecology, Mount Sinai Hospital, University of Toronto, Toronto, Ontario, Canada; 2) Division of Clinical and Metabolic Genetics, Department of Pediatrics, The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada; 3) Department of Genetics, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia; 4) Laboratory of Molecular Gerontology, National Institute on Aging, National Institutes of Health, NIH Biomedical Research Center, 251 Bayview Blvd, Baltimore, MD 21224, USA; 5) Cytogenomics Laboratory, Division of Genome Diagnostics, Department of Pediatric Laboratory Medicine, The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada; 6) Department of Clinical Genetics, Human Genetics and Genome Research Division, National Research Centre, Cairo, Egypt; 7) Department of Medical Genetics, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia; 8) Division of Neuroradiology, Department of Diagnostic Imaging, The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada; 9) Department of Otolaryngology - Head & Neck Surgery, The Hospital for Sick Children, University of Toronto, Toronto, ON, Canada; 10) Department of Anatomy and Cell Biology, College of Medicine, Alfaisal University, Riyadh, Saudi Arabia; 11) Saudi Human Genome Program, King Abdulaziz City for Science and Technology, Riyadh, Saudi Arabia.

Warsaw breakage syndrome (WBS) is a recently recognized *DDX11*-related rare cohesinopathy, characterized by severe prenatal and postnatal growth restriction, microcephaly, developmental delay, cochlear anomalies and sensorineural hearing loss. Only seven cases have been reported in the English literature, and thus the information on the phenotype and genotype of this interesting condition is limited. We provide clinical and molecular information on five additional unrelated patients carrying novel bi-allelic variants in the *DDX11* gene, identified via whole exome sequencing. One of the variants was found to be a novel Saudi founder variant. All identified variants were classified as pathogenic or likely pathogenic except for one which was initially classified as a variant of unknown significance (VOUS) (p.Arg378Pro). Functional characterization of this VOUS using heterologous expression of wild type and mutant *DDX11* revealed a marked effect on protein stability, thus confirming pathogenicity of this variant. The phenotypic data of the seven WBS reported patients were compared to our patients for further phenotypic delineation. Although all the reported patients had cochlear hypoplasia, one patient also had posterior labyrinthine anomaly. We conclude that while the cardinal clinical features in WBS (microcephaly, growth retardation and cochlear anomalies) are almost universally present, the "breakage" phenotype is highly variable and can be absent in some cases. This report further expands the knowledge of the phenotypic and molecular features of WBS.

2087F

An iPSC approach towards understanding the molecular mechanisms underlying SRCAP mutations in Floating-Harbor syndrome. *R.L. Hood*^{1,2}, *K.M. Boycott*^{3,4}, *M.J.M. Nowaczyk*⁵, *W.L. Stanford*^{1,6,7}, *D.E. Bulman*^{1,2,4,7}. 1) Department of Biochemistry, Microbiology and Immunology, University of Ottawa; 2) CHEO Research Institute, University of Ottawa, Ottawa, Ontario, Canada; 3) Department of Genetics, Children's Hospital of Eastern Ontario; 4) Department of Pediatrics, University of Ottawa; 5) Department of Molecular Medicine and Pathology, McMaster University, Hamilton; 6) The Sprott Centre for Stem Cell Research, Regenerative Medicine Program, Ottawa Hospital Research Institute; 7) Department of Cellular and Molecular Medicine, University of Ottawa.

Floating-Harbor Syndrome (FHS; OMIM# 136140) is characterized by short stature, delayed osseous maturation, cognitive deficit, and unique dysmorphic facial features. We previously employed an exome-sequencing strategy which identified heterozygous truncating mutations clustered in the 3' end of *SRCAP* as the genetic cause underlying FHS. More recently, using a high-resolution genome-wide DNA methylation analysis method we identified a specific methylation "epi-signature" in the peripheral blood of FHS individuals. *SRCAP* has been described as having coactivator roles in CREB and CBP-mediated and nuclear (steroid) hormone receptor signaling pathways. It has also been shown to co-precipitate as part of a large chromatin remodeling complex which catalyzes ATP-dependent displacement of the histone variant H2A by H2A.Z. Based on the described roles of *SRCAP*, the clustered nature of truncating *SRCAP* mutations in FHS, and the knowledge that these mutations result in global methylation alterations, we expect that mutations in *SRCAP* to result in widespread gene dysregulation. The existence of individuals with deletions encompassing *SRCAP* who do not have FHS leads us to postulate that the mechanism of disease is not due to haploinsufficiency. To examine the functional consequences of *SRCAP* mutations in cell lineages implicated in the FHS phenotype, we have utilized an induced pluripotent stem cell (iPSC) disease-modeling approach. FHS patient and gender-matched control fibroblast cells were reprogrammed into iPSCs using an episomal vector system and subsequently characterized by pluripotency cell surface marker expression, in vitro embryoid body differentiation, and in vivo teratoma formation assays. iPSCs were plated into Aggrewell™ plates and differentiated into neural crest cells using a dual SMAD inhibition approach. Expression of neural crest markers p75, HNK-1, and SOX10 were confirmed by immunofluorescence staining. Lines were FACS sorted (p75⁺/PI⁻) to isolate pure neural crest cell populations. A combination of RNA-seq and ATAC-seq and will be employed to examine differences in global transcript expression and genomic architecture. Taken together, these analyses will identify downstream targets and pathways which are perturbed within the developing neural crest of FHS patients. These results will provide valuable insight into FHS disease mechanism and the dysmorphology seen in FHS in a precise tissue specific manner.

2088W

Hypomorphic mutations in the replication complex member *GINS3* result in a Meier-Gorlin phenotype. K. Ahmed¹, M. McQuaid¹, K. Kernohan⁵, S. Tran⁶, P. Grover³, L. Dupuis³, R. Mendoza-Londono³, A. Howard², J. Stimec⁸, K. Boycott⁵, F. Boisvert⁸, R. Shaheen¹⁰, Z. Alhassnan¹⁰, F. Alkuraya¹⁰, H. Wurtele⁴, E. Campos⁶, P. Kannu^{1,2,3}. 1) Developmental and Stem Cell Biology, Hospital for Sick Children, Toronto, ON, Canada; 2) Institute of Medical Science, University of Toronto, Toronto, ON, Canada; 3) Division of Clinical and Metabolic Genetics, Hospital for Sick Children, Toronto, ON, Canada; 4) Université de Montréal, Maisonneuve-Rosemont Hospital Research Center, Montreal, QC, Canada; 5) Department of Genetics, Children's Hospital of Eastern Ontario, Ottawa, Canada; 6) Genetics & Genome Biology, Hospital for Sick Children, Toronto, ON, Canada; 7) Division of Orthopaedics, Hospital for Sick Children, Toronto, ON, Canada; 8) Body Imaging Division, Hospital for Sick Children, Toronto, ON, Canada; 9) Département d'anatomie et de biologie cellulaire, Université de Sherbrooke, Sherbrooke, QC, Canada; 10) Department of Genetics, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia.

To ensure efficient duplication of the genome, numerous factors promoting unperturbed DNA replication have evolved. The evolutionarily conserved CDC45/MCM2-7/GINS1-4 (CMG) protein complex is an essential helicase that unwinds DNA during DNA replication. Interestingly, a causative link was recently made between germline mutations in CMG-encoding genes and Meier-Gorlin syndrome (MGS). MGS is clinically defined by absent or hypoplastic patella, microtia and pre- and postnatal growth restriction leading to short stature. Here we report the identification of four individuals from two unrelated families presenting a Meier-Gorlin-like phenotype, characterized by unreported recessive hypomorphic mutations in the *GINS3* gene, which encodes a subunit of the CMG helicase. Affected individuals were all characterized by global growth failure. We found that mutations in *GINS3* limit cellular proliferation, providing a potential mechanism for the dwarfism phenotype. Furthermore, the mutations seem to destabilize the GINS subcomplex, and the entire CMG helicase altogether. Cells with the *GINS3* mutations accumulate in the S and G2-phases of the cell cycle. We also show that the loss of *GINS3* was associated with replication-associated DNA damage. In summary, we have identified mutations in *GINS3* as a new cause of Meier-Gorlin syndrome and established *GINS3* as a critical component required for efficient DNA replication and genome stability.

2089T

A 61 base-pair intronic deletion in *Snrpb* causes abnormal phenotypes in mice. S. Alam¹, J. Majewski^{1,2}, L.A.J. Majewska^{1,2,3}. 1) Human Genetics, McGill University, Montreal, Quebec, Canada; 2) McGill University Health Centre at Glen Site, Montreal, QC, Canada; 3) Department of Pediatrics, McGill University, Montreal, Quebec, Canada.

SNRPB, which codes for a core component of the spliceosomal small nuclear ribonucleoproteins (snRNPs) is mutated in patients with a rare autosomal dominant disorder known as Cerebrocostomandibular Syndrome (CCMS). Patients mostly have craniofacial abnormalities and rib defects. Exome sequencing has revealed mutations mostly in the highly conserved alternative exon 2 (AE2) of the gene, where it plays an instrumental role in autoregulating SNRPB. Using Crispr/Cas9, we have generated a mutant mouse line with a 61 base-pair (bp) intronic deletion upstream of the AE2 of *Snrpb*. Heterozygous and homozygous mice and embryos for this intronic deletion show phenotypic abnormalities that are seen in CCMS patients. Intriguingly, heterozygous and homozygous embryos show reduced SNRPB protein levels, suggesting that this 61 bp sequence plays a critical role in regulating SNRPB levels. Our preliminary data suggests that the splicing of the gene is changed due to the 61-bp intronic deletion where the AE2 is preferentially included in the transcript. Future studies will elucidate the mechanism by which this intronic sequence regulates SNRPB. The importance of this 61-bp intronic sequence near the AE2 might shed light on CCMS studies in cases where the patients do not have mutation in the exons.

2090F

Craniofacial development requires glycosphosphatidylinositol biosynthesis. M. Lukacs, P. Chaturvedi, R. Stottmann. Cincinnati Children's Hospital Medical Center, Cincinnati, OH., USA.

Post-translational protein modification via glycosphosphatidylinositol (GPI) attachment is a multi-step process resulting in the plasma membrane anchoring of nearly 150 proteins. Patients with mutations in GPI biosynthetic pathway genes display an array of phenotypes including brain malformations and neurocristopathies. There is virtually no mechanism to explain the pathology or therapy for patients with germline GPI deficiency. Through a forward genetic screen we identified a mouse mutant (*cleft lip/palate, edema, and exencephaly; Clpex*) with a hypomorphic mutation in *Post-Glycosphosphatidylinositol Attachment to Proteins-2 (Pgap2)*. *Pgap2* is required for GPI anchor lipid maturation by promoting the addition of stearic acid to the lipid portion of the anchor in the golgi apparatus. We hypothesized the *Clpex* mutation results in a lack of surface expression of a critical GPI-anchored protein, Folate Receptor 1 (FOLR1), which is required for neural tube closure and neural crest cell survival. *In vitro* overexpression of tagged FOLR1 in *PGAP2*^{-/-} cells demonstrates an inability to present FOLR1 on the plasma membrane. We found the cleft lip/palate phenotype in *Clpex* mutants is due to apoptosis of neural crest cells in the first branchial arch. *In utero* dietary supplementation with folic acid can rescue the cleft lip phenotype and partially rescue the neural crest cell apoptosis in *Clpex* but not the exencephaly or cleft palate. Surprisingly, RNA sequencing of *Clpex* mutants identified decreased expression of anterior patterning genes and a reciprocal increase in posterior patterning genes. Defects in GPI-anchored Cripto may underlie these changes in gene expression as Cripto regulates anterior-posterior patterning through Nodal. Similar defects in A/P patterning have been identified in other GPI biosynthesis mutants. Loss of anterior patterning gene expression may underlie the folic acid resistant neural crest cell apoptosis in *Clpex*. These findings advance our understanding of the pathology of GPI deficiency and provide potential targets for therapy.

2091W

New candidate gene for CAKUT: Exome sequencing in human bladder exstrophy and knockdown studies in zebrafish implicate SLC20A1 as disease gene and major regulator of urinary tract development. J.M. Schmidt^{1,2}, O. Yilmaz², R. Zhang¹, A.S. Japp³, M. Pleschka², H.M. Reutter^{1,4}, B. Odermatt².

1) Institute of Human Genetics, Faculty of Medicine, University of Bonn, Bonn, Germany; 2) Institute of Anatomy, Faculty of Medicine, University of Bonn, Bonn, Germany; 3) Institute of Neuropathology, Faculty of Medicine, University of Bonn, Bonn, Germany; 4) Department of Neonatology and Pediatric Intensive Care, Faculty of Medicine, University of Bonn, Bonn, Germany.

Introduction: The bladder exstrophy-epispadias complex (BEEC) represents the severe end of human congenital anomalies of the kidney and urinary tract (CAKUT). Previously exome sequencing in eight case-parent trios with cloacal exstrophy identified one novel *de novo* variant (p.G237R) in *SLC20A1* in a single patient. Re-sequencing of 526 BEEC patients identified one additional novel *de novo* variant (p.V298A) and one novel maternally transmitted variant (p.K441Q) from an affected mother. The amino acid change p.G237R is highly conserved and localizes to a transmembrane domain of SLC20A1. p.V298A and p.K441Q localize to cytosolic domains and are less conserved among species. *SLC20A1* encodes for a Na⁺/PO₄ cotransporter known to play a role in proliferation and TNF-induced apoptosis. We investigated the developmental function of *slc20a1a* in developing zebrafish larvae (zfl) performing Morpholino (MO) knockdown (KD) experiments. The urinary tract in zfl consists of two pronephric ducts, similar in segmentation to human nephrons, that fuse at the cloaca. *Slc20a1a* is frequently used as a pronephric in situ hybridization (ISH) marker. **Methods:** *Slc20a1a* KD in zfl was done by injecting antisense oligonucleotides with a morpholino[®] backbone into 1-2 cell staged eggs, blocking the gene's translation. Specificity of the MO was shown by Western Blot and rescue experiments, co-injecting human mRNA transcripts of *SLC20A1*. We characterized the MO phenotypes using different assays such as sulforhodamine 101 excretion, ISH with several urogenital marker probes, immunohistochemistry (IHC) staining and the transgenic zfl reporter line *Tg(wt1b:GFP)*. **Results:** MO KD of *slc20a1a* in zfl results in a severe lethal phenotype that affects multiple organ systems. Focusing on urinary tract, we see formation of pronephric cysts and disorganization of the cloaca. Excretion assay uncovers severe opening defects of the cloaca for the intestine, similar to our index patient with cloacal exstrophy. Yolk endocytosis defects resemble abdominal wall fusion deficiencies. IHC shows defects in proliferation and apoptosis. **Discussion:** In conclusion, exome sequencing in human bladder exstrophy and developmental biology studies in zfl implicate *SLC20A1* as disease gene for human BEEC and therefore CAKUT and as major regulator of urinary tract development. Severe MO phenotypes present with a high lethality among zfl, underlining the overall importance of *slc20a1a* during embryonic development in zfl.

2092T

Loss-of-function mutations of *TMEM260* cause autosomal recessive polycystic kidney, cerebral atrophy and cardiac malformation. T.M. Keszthelyi^{1,2}, M. Varga^{1,3}, D. Ralbovsk³, L. Ablonczy⁴, C. Bole⁵, C. Antignac⁵, K. Tóty^{1,2}.

1) MTA-SE Lendület Nephrogenetic Laboratory, Budapest, Hungary; 2) Semmelweis University, 1st Department of Pediatrics, Budapest, Hungary; 3) Department of Genetics, Eötvös Loránd University, Budapest, Hungary; 4) Gottsegen György National Institute of Cardiology, Budapest, Hungary; 5) INSERM, UMR 1163, Imagine Institute, Paris, France.

Association of polycystic kidney, cerebral atrophy and truncus arteriosus had not been reported in the medical literature at the time of consulting a child with this phenotype. We aimed to identify the causal gene and to characterize the encoded protein. Whole exome sequencing was performed to identify the causal mutations. Gene expression was studied in PBMC. The localization of the *TMEM260* protein was studied by immunofluorescence staining under confocal microscopy in transiently transfected IMCD3 cells after the induction of ciliogenesis. Mutant zebrafish lines were generated by the CRISPR/Cas9 system. We identified two heterozygous truncating mutations in the *TMEM260* gene: c.592_593delTT, p.Leu198Valfs44* and c.1854C>A (p.Tyr618*). They were in trans as verified by Sanger sequencing of the parental samples. We found both mutated alleles to be equally expressed in the parental samples, indicating that none of them leads to nonsense-mediated decay. Out of the two *TMEM260* isoforms, only the long is affected by the p.Tyr618* mutation, allowing the short isoform to be properly translated. We generated several *tmem260* loss-of-function mutations and found the adult zebra danio mutants with biallelic loss-of-function mutations in a mendelian ratio with no obvious phenotype. Interestingly, in contrast to the proteins implicated in ciliary kidney disorders, the N-terminal myc-tagged *TMEM260* protein showed a cytoplasmic localization, but was not detectable in the primary cilium. The affected boy is currently 7-year-old with extremely enlarged polycystic kidneys (+9SD), preserved renal function, severe developmental delay, epilepsy, muscle hypotonia, strabismus, ptosis, cystic liver and bilateral bronchial stenosis. His phenotype is consistent with the phenotype of four children recently reported with *TMEM260* mutations from two families. In conclusion, we confirm that mutations of *TMEM260* cause a severe autosomal recessive disorder with severe neurological, renal, cardiac, hepatic and bronchial involvement. *TMEM260* is the first protein mutated in a cystic kidney disease which is not localized to the primary cilium. The preserved translation of the short isoform in the two previously published and the current family suggest that the dysfunction of the short isoform is intrauterine lethal. Supported by MTA-SE Lendület Research Grant (LP2015-11/2015), NKFIA/OTKA K109718, FK124230 and KH125566.

2093F

Elucidating molecular mechanisms for the expanding phenotype of *POLR1A*-related disorders. K.N. Weaver¹, K. Baltrunaite¹, C. Brunswick¹, M.P. Adam², A. Begtrup³, D. Bertola⁴, M. Cho³, L. Demmer⁵, E. Demo⁶, O. Devinsky⁷, E. Gallagher⁸, M.J. Guillen Sacoto³, J. Keller-Ramey⁹, R. Ladda⁸, H. McLaughlin³, K. McWalter⁹, K. Menzer⁹, R. Person³, R. Schnur⁹, S. Sell⁸, J.D. Symonds⁹, R. Willaert³, T. Wenger⁹, S.M. Zuber⁹, R.W. Stottmann¹.

1) Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 2) University of Washington, Seattle, WA; 3) GeneDx, Gaithersburg, MD; 4) University of São Paulo, Brazil; 5) Levine Children's Specialty Center, Charlotte, NC; 6) Sibley Heart Center, Atlanta, GA; 7) Department of Neurology, Comprehensive Epilepsy Center, New York University School of Medicine, New York, NY; 8) Hershey Medical Center, Hershey, PA; 9) Pediatric Neurosciences Research Group, Royal Hospital for Children, University of Glasgow, Glasgow, UK.

Pathogenic variants in *POLR1A* were identified as the cause of autosomal dominant Acrofacial dysostosis, Cincinnati type (AFD-CIN) in 2015. *POLR1A* encodes a subunit of RNA polymerase I, which transcribes ribosomal RNA. Zebrafish lacking *polr1a* exhibit deficient ribosome biogenesis. Thus, AFD-CIN is a ribosomopathy, a congenital malformation syndrome caused by ribosome dysfunction. The 3 index patients with AFD-CIN only had craniofacial and limb anomalies. Other ribosomopathies are associated with pleiotropic phenotypes (including craniofacial, limb, heart, and neurodevelopmental abnormalities) as well as with reduced penetrance and variable expression. We describe expansion of phenotypes beyond the face and limbs in a growing cohort of individuals with suspected pathogenic variants in *POLR1A*. Among 13 patients, there is an increased incidence of developmental delay, congenital heart defects, and seizures. Novel craniofacial phenotypes include cleft lip/palate, craniosynostosis, ptosis, and vocal cord paralysis. Novel limb anomalies include distal contractures, duplicated thumb, and short radii and ulnae. Three individuals inherited variants from a mildly affected or unaffected parent, consistent with reduced penetrance and/or variable expressivity. In this cohort there are 9 missense variants which affect different domains of the *POLR1A* protein. The third cleft domain is the most commonly affected (3 variants). It is thought to be important for transformation of RNA Polymerase I between active and inactive conformations. To test pathogenicity for new variants and elucidate molecular mechanisms for observed phenotypes, we are pursuing *in vitro* studies to characterize each mutant protein as well as *in vivo* studies with a new *Polr1a* loss-of-function mouse model. Preliminary results demonstrate that none of the missense variants lead to decay of mutant *POLR1A* protein. Mice with conditional deletion of *Polr1a* in neural crest cells, an embryonic cell population essential for facial development, exhibit severe deficiency of craniofacial primordia and die mid-gestation. Therefore, our data suggest that variable phenotypes observed among patients with *POLR1A* variants are due to dominant-negative or gain-of-function effects of specific variants rather than haploinsufficiency. Our studies will enhance understanding of the disease pathogenesis and molecular mechanisms associated with *POLR1A*-related disorders and ultimately will improve the care of affected individuals.

2094W

Ubiquitous expression of *Akt1* p.(E17K) results in vascular defects and embryonic lethality in mice. M.J. Lindhurst¹, W. Li², J.J. Shweta¹, H.C. Kondolf¹, M.R. Yourick¹, Y. Mukoyama², L.G. Biesecker¹. 1) Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, NIH, Bethesda, MD; 2) Laboratory of Stem Cell and Neuro-Vascular Biology, Genetics and Developmental Biology Center, National Heart, Lung, and Blood Institute, NIH, Bethesda, MD.

Heterozygous mosaic expression of the activating variant, c.49G>A, p.(E17K) *AKT1*, results in Proteus syndrome (PS), a rare disorder characterized by disproportionate, progressive, asymmetric overgrowth that can affect any tissue in the body. In 1986, Happle hypothesized that if the PS mutation were present in the germ line it would be lethal. Using knock-in technology, we created a conditional *Akt1* p.(E17K) allele and developed an inducible mouse model of PS. To test the Happle hypothesis, we activated the variant ubiquitously by crossing *Akt1* conditional and ACTB-cre expressing mice. No live born mutant offspring from this cross were identified (n=62). Timed matings were performed and expected numbers of mutant embryos were seen through E12.5, though some mutants were clearly dead by this time point. Mutant embryos from E13.5-E17.5 were pale, had fewer visible blood vessels and increased hemorrhage. At each time point, there were reduced numbers of mutant embryos with the phenotype varying from mild to severe. Angiogenesis defects were seen in the skin of mutant embryos as they had fewer mural cells and failed to undergo proper arteriovenous remodeling. Histological evaluation of mutant embryos revealed angiectasia in the liver and mild structural defects in the heart. Because the phenotype of the ACTB-Cre mutant embryos suggested that the lethality was due to a vascular defect, we restricted *Akt1* p.(E17K) expression to specific cell types using three additional Cre lines, Tek-cre (endothelial cells), Tagln-cre (smooth muscle cells), Nkx2.5-cre (cardiac cells). Live born offspring heterozygous for both Cre and the *Akt1* conditional allele were found for each line. Evaluation of skin vasculature, liver and heart structures in double heterozygous embryos is underway, but to date only mild liver angiectasia has been found in some of the Tek-Cre embryos. These data show that activation of the PS mutation in endothelial cells, smooth muscle cells or heart is not sufficient for the lethality and vascular defects seen with ubiquitous expression. This suggests that another component of the vascular system is responsible, or alternatively, that activation of the mutation in multiple cell types is necessary to disrupt vascular development and cause the lethality seen with ubiquitous expression of the PS mutation.

2095T

Mutation of mouse *Snap29* models CEDNIK and 22q11.2 deletion syndrome. L.A. Jerome-Majewska^{1,3,5}, V. Keser¹, S.S. Shameen¹, Y. Lim², E. Scarlata³, C. O'Leary⁴. 1) Department of Human Genetics, McGill University, Montreal, Quebec, Canada; 2) Department of Pathology, Brigham and Women's Hospital, Harvard Medical School Boston, Massachusetts, United States of America; 3) McGill University Health Centre at Glen Site, Montreal, QC, Canada; 4) Departments of Surgery (Urology Division); 5) Department of Pediatrics, McGill University, Montreal, Quebec, Canada.

Synaptosomal-associated protein 29 (SNAP29) is a member of the SNARE family of proteins and is implicated in a number of intracellular protein trafficking pathways. *SNAP29* maps to human 22q11.2 and mutations in this gene are associated with two developmental syndromes: Cerebral dysgenesis, neuropathy, ichthyosis, and palmoplantar keratoderma (CEDNIK) and 22q11.2 deletion syndrome (22q11.2DS). Unfortunately, existing *Snap29* mutant mouse lines only model skin abnormalities found in CEDNIK patients. Our goals were to describe expression of *Snap29* during embryogenesis and to create a mutant mouse line that can be used to uncover the etiology of malformations shared between CEDNIK and 22q11.2DS patients. *In situ* hybridization was used to characterize expression of *Snap29* throughout embryogenesis, and CRISPR/Cas9 was used to generate a mouse line with deletion in *Snap29*. Morphological abnormalities in *Snap29* mutant embryos and mice were analyzed using standard histological and molecular methods. Herein, we report that *Snap29* was widely expressed in the precursors of organs affected in CEDNIK and 22q11.2DS patients, in addition to many other tissues that were not reported to be affected. Mice and pups with homozygous mutation in *Snap29* survived until adulthood and showed skin, motor, and craniofacial abnormalities, enabling us to study the consequences of mutation in this gene in the neonatal and the postnatal period. In addition, our studies revealed a previously unknown requirement for *Snap29* in the reproductive system. Our study indicates that this novel *Snap29* mutant mouse line models most abnormalities observed in CEDNIK patients and suggest that deletion of *SNAP29* contributes to the phenotypic spectrum of abnormalities found in 22q11.2DS.

2096F

Loss of function mutation of *Eftud2*, the gene responsible for mandibulo-facial dysostosis with microcephaly (MFDM), leads to pre-implantation arrest in mouse. *M. Beauchamp*¹, *A. Djedid*², *K. Daupin*¹, *K. Clokie*¹, *J. Majewski*², *L. Jerome-Majewska*^{1,2,3}. 1) Research Institute of McGill University Health Center, Montreal, Quebec, Canada; 2) Department of Human Genetics, McGill University; 3) Department of Anatomy and Cell Biology, McGill University.

Haploinsufficiency of *EFTUD2* is associated with MFDM (mandibulo-facial dysostosis with microcephaly), but how mutations of *EFTUD2* cause abnormalities associated with this syndrome remains unknown. In the present study, we first aimed to characterize the expression of *Eftud2* during mouse development and to generate a mutant mouse line carrying *Eftud2* mutation. *In situ* hybridization revealed expression of *Eftud2* throughout embryonic development, with tissue-specific expression in ectodermal and mesodermal components of the developing head and craniofacial region. The CRISPR/Cas9 system was used to delete exon 2 of *Eftud2*. *Eftud2* heterozygous animals were viable and fertile despite a 38% and a 30% reduction of *Eftud2* mRNA and protein expression, respectively. Furthermore, heterozygous embryos showed delayed development during embryogenesis, which disappears by organogenesis. Noticeably, RNA sequencing revealed that *Eftud2* was the only transcript affected in heterozygous mice. After collection of embryos from embryonic day (E)3.5 to E18.5 from matings between *Eftud2* heterozygous males and heterozygous females, homozygous mutant embryos were found only at E3.5 stage, indicating pre-implantation arrest. Further *ex vivo* analysis of E3.5 embryos from *Eftud2*^{+/−} inter se crosses showed a 32% embryonic death compared to embryos collected from wild-type crosses. Those embryos never grew beyond the blastocyst stage and were unable to hatch. Our observations indicate that reduced levels of *Eftud2* is associated with developmental delay at the embryonic stage, and that *Eftud2* homozygous mutant embryos do not survive post-implantation. Thus, a conditional *Eftud2* knock-out mouse model is necessary to specifically study the requirement of this splicing factor in the developing neural crest cells during craniofacial development.

2097W

Genome-wide chromatin accessibility and transcriptome profiling identifies dysregulated hedgehog signaling in Danforth's short tail mice. *C.E. Keegan*, *J.S. White*, *P. Orchard*, *P.E. Thomas*, *A. Mychalowych*, *A. Kiseleva*, *J. Hensley*, *B. Allen*, *S.C.J. Parker*. University of Michigan, Ann Arbor, MI.

Caudal birth defects are a group of clinically well-recognized disorders affecting proper development of the spine, kidneys, reproductive organs, intestines and lower limbs, occurring in 1:10,000 pregnancies. The Danforth's Short Tail (*Sd*) mouse exhibits a fully penetrant, semi-dominant phenotype characterized by complete lack of tail, a severely shortened spinal column, persistent cloaca and anal atresia in homozygous *Sd/Sd* mutants. The *Sd* mutant phenotype is caused by an endogenous retrovirus (ERV) insertion, resulting in ectopic expression of the transcription factor *Ptf1a* at E9.5 in the developing tailbud. The *Sd* mutant phenotype has substantial overlap with human caudal dysgenesis disorders, including caudal regression, VACTERL association, and persistent cloaca, thus providing a model to investigate developmental signaling pathways that are disrupted in human caudal dysgenesis. The goal of this project was to investigate epigenomic and transcriptomic changes resulting from ectopic *Ptf1a* expression in embryonic tailbuds from *Sd* mutant embryos to identify the downstream signaling pathways that are disrupted in *Sd* embryos. We performed ATAC-seq and mRNA-seq on tailbuds from E9.5 wildtype and *Sd/Sd* mutant embryos. ATAC-seq analysis identified a region of significantly ($P=4.5 \times 10^{-7}$, Log2FC=0.90) increased chromatin accessibility in *Sd* mutants at the lncRNA *Gm13344* adjacent to the ERV insertion near the *Ptf1a* locus, which corresponds to a previously identified autoregulatory enhancer of *Ptf1a*. *Ptf1a* and *Gm13344* were significantly upregulated by mRNA-seq (*Ptf1a* $P=4.7 \times 10^{-11}$, Log2FC=8.4; *Gm13344* $P=2.7 \times 10^{-16}$, Log2FC=8.4). In total, 49 genes were significantly differentially expressed (5% FDR), including known targets of *Ptf1a*, such as *Kirrel2*, *Nphs1* and *Cpa*. KEGG pathway analysis identified the Hedgehog signaling pathway as being significantly ($P=1.1 \times 10^{-4}$) downregulated, which was confirmed by whole mount *in situ* hybridization with *Shh* and crossing to *Gli1-LacZ* reporter mice. In addition, we found reduced expression of *Shh* and *Foxa2* in the notochord and floorplate and aberrant patterning of the neural tube at E10.5 as a consequence of reduced *Shh* signaling. More generally, we have shown how a single non-coding regulatory mutation propagates through a molecular cascade to perturb a critical signaling pathway. This work can help provide insight into how regulatory changes in gene expression underlie caudal structural birth defects.

2098T

De novo insertion in the *CHD7* gene causing a CHARGE syndrome. L. Pranckieniene¹, E. Preiksaitiene¹, V. Kucinskas¹, A. Reymond², L. Gueneau². 1) Vilnius university, Vilnius, Lithuania; 2) Center for Integrative Genomics, Faculty of Biology and Medicine, Lausanne University, Switzerland.

CHARGE syndrome is an autosomal dominant development disorder with a broad phenotype that can involve almost all organ and sensory systems. CHARGE syndrome is characterized by a pattern of congenital anomalies including choanal atresia and malformations of the heart, inner ear, and retina. Variants in the *CHD7* gene on chromosome 8q12.1 cause CHARGE syndrome in most patients. We report a novel *de novo* pathogenic p.Y2114delinsYHILNDH insertion in *CHD7* gene causing severe features of CHARGE syndrome. In this research whole exome sequencing (WES) of a Lithuanian trio (affected individual and healthy parents) was performed. WES was generated by sequencer Illumina™ HiSeq 2500 and primary analysis was performed by GATK™. Exomes were mapped according to the human reference genome build 38. *De novo* variants (DNVs) called by VarScan. To discard a false positive *de novo* calls when it was not known whether all of the individuals in the trio are identified correctly conservative filters by SnpSift software on detected DNMs quality parameters were applied: 1) genotype quality of the individual ≥ 50 ; 2) number of reads at each site > 35 . Called potentially DNVs manually reviewed by the IGV and validated by Sanger sequencing by ABI PRISM 3130xl Genetic Analyzer. Functional annotation of all identified DNMs were performed using ANNOVAR and hg38 reference human genome sequence. The insertion c.6340_6341insATCACATCCTCAATGACC identified in patient is in a region of *CHD7* which is highly conserved between species. According *in silico* analysis this insertion causes an important structural change on the protein level: from polar tyrosine into a chain of histidine, isoleucine, leucine, asparagine which forms a short alpha helix instead the linear chain. Considering the size of p.Y2114delinsYHILNDH nonframeshift insertion, which possibly forms a new structure of protein domain, we provide evidence, that insertion in this region causes severe phenotype of CHARGE syndrome. This study was supported by the Lithuanian–Swiss cooperation program under UNIGENE project agreement no. CH-3-□MM-01/04.

2099F

Analysis of molecular pathway in primary fibroblasts and myoblasts of Wolf-Hirschhorn patients (WHS) by RNA-seq. P.N. Doronzio, G. Marangi, S. Lattante, S. Frangella, M. Zollino. Istituto di Medicina Genomica, Università Cattolica del Sacro Cuore, Roma, RM, Italy.

Wolf-Hirschhorn Syndrome (OMIM: 194190) is a contiguous gene syndrome caused by partial 4p deletion, which core phenotype include intellectual disability (ID), the typical "greek warrior helmet profile", growth delay and seizures. All these signs are regarded as minimal diagnostic criteria for WHS and they all map within the terminal 1.9 Mb on 4p, where the WHS critical region (WHSCR-2) was described. Pathogenesis of WHS appears to be multigenic, being *WHSC1* the major candidate gene for growth delay and the typical facial dysmorphism, and *LETM1* and *CPLX1* the major synergic genes for seizures. We aimed to investigate by RNA Seq impaired molecular pathways at a whole transcriptome level in fibroblast and myoblast cell lines obtained from 4 WHS patients and from 3 healthy controls. Extent of the 4p deletion was of 6.5 Mb in two patients, 3.1 Mb and 16.5 Mb in individual patients. RNA samples has been used to prepare RNA-Seq libraries for ION Proton platform. The results has been analyzed with "Cufflinks" and "Salmon" workflows, while the DAVID functional annotation tool has been used for pathway enrichment analysis. We observed a differential expression of genes related to cell growth / proliferation and to cell adhesion pathways both in fibroblasts and myoblasts. With respect to these subset of transcripts, results were to large extent comparable to each other independently from the 4p deletion size.

2100W

Multiple malformations associated with loss of function variants in *ZNF462*: A study of humans and zebrafish. T. Hu¹, P. Kruszk², J.A. Martinez-Agosto^{2,3}, R. Signer², J. Jurgens⁴, E.C. Engle⁴, T. Hartley⁵, K. Boycott⁶, M. Fannemel⁷, N. Beck⁸, N. Ghali⁹, B. Anderlid¹⁰, J. Lundin¹⁰, S. Ghedia¹¹, C.M. Bartley¹², R.C. Spillmann¹³, K. Wigby^{14,15}, J.C. Giltay¹⁶, K.L.I. van Gassen¹⁶, K.W. Brigatti¹⁷, S. Mazzola¹⁸, E.H. Zackai^{18,19}, S. Hong¹, A. Hamosh⁸, M.C. Jones^{14,15}, K. Weiss²⁰, M. Muenke¹. 1) Medical Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA; 2) Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA, USA; 3) Department of Pediatrics, David Geffen School of Medicine, University of California, Los Angeles, CA, USA; 4) Department of Neurology, Boston Children's Hospital, Harvard University, Boston, MA, USA; 5) Children's Hospital of Eastern Ontario Research Institute, Ottawa, Ontario, Canada; 6) Department of Genetics, Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada; 7) Department of Medical Genetics, Rikshospitalet, Oslo University Hospital, Oslo, Norway; 8) Department of Pediatrics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA; 9) North West Thames Regional Genetics Service, Northwick Park Hospital, London North West Healthcare NHS Trust, Harrow, UK; 10) Department of Molecular Medicine and Surgery, Center of Molecular Medicine, Karolinska Institute, Karolinska University Hospital, Stockholm, Sweden; 11) Department of Clinical Genetics, Royal North Shore Hospital, St. Leonards, New South Wales, Australia; 12) Department of Psychiatry, University of California, San Francisco, CA, USA; 13) Department of Pediatrics, Duke University School of Medicine, Durham, NC, USA; 14) Department of Pediatrics, University of California, San Diego, CA, USA; 15) Rady Children's Hospital, San Diego, CA, USA; 16) Genetics Department, University Medical Center Utrecht, Utrecht, Netherlands; 17) Clinic for Special Children, Strasburg, PA, USA; 18) Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA, USA; 19) Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; 20) Rambam Medical Center, Haifa, Israel.

Background: We recently reported on 8 individuals from 6 families with loss of function variants in *ZNF462* who presented with metopic craniosynostosis, ptosis, agenesis of the corpus callosum, and developmental delay. Subsequently, Cosemans et al. reported on an individual with a *de novo* balanced translocation affecting *ZNF462*; this individual presented with 46,XY,t(9;13)(q31.2;q22.1) and intellectual disability, metopic craniosynostosis, dysgenesis of the corpus callosum, and ptosis. Little is known about the function of *ZNF462* in humans. Chang et al. showed that *Zfp462*, an ortholog of *ZNF462*, is expressed in the developing forebrain, brainstem, and spinal cord in mice. To further understand the pathogenicity of loss of function of *ZNF462*, we have continued phenotyping patients and are studying loss of function of *znf462* in zebrafish. **Methods:** Additional cases of *ZNF462* variants were collected from correspondence associated with the Weiss et al. paper published by our group and GeneMatcher. Zebrafish morpholino oligonucleotide (MO) and in situ hybridization experiments are being conducted. An MO was designed to target the *znf462* translation start site. Anti-sense RNA probes are being used to target *znf462* in wild type zebrafish and dorsal and ventral markers, including *chrd* and *bmp2b*, in morphants. **Results:** In addition to the 9 published reports, we have identified 10 new individuals from 9 families with variants in *ZNF462*; six of these variants are truncating variants with predicted loss of function and one variant is a missense variant. Thirteen of the 19 (68%) total cases (including previously published cases) exhibit dysgenesis of the corpus callosum, 12 of 19 (63%) have ptosis, and 9 of 19 total are affected with intellectual disability or autism spectrum disorder. Preliminary results in our zebrafish experiments have shown that morphant embryos display varying severities of a ventralized phenotype, which involves a reduction in the size of dorsally derived structures such as the neural plate and somites. The morphant phenotype first appears with injections of 2.5 ug of MO. In situ hybridization experiments show that morphant embryos have decreased *chrd* expression and increased *bmp2b* expression at the 6 hpf and 8 hpf stages, respectively. **Conclusions:** Loss of function in *ZNF462* results in a distinctive phenotype in both humans and zebrafish. Preliminary data in the zebrafish model suggest that *znf462* may play a role in dorsoventral patterning.

2101T

Impaired pro-angiogenesis in mesenchymal stem cells of Werner syndrome. H.H. Cheung, J. Tu, W.Y. Chan. School of Biomedical Sciences, The Chinese University of Hong Kong, Hong Kong, China.

Werner syndrome (WS) is an autosomal recessive genetic disorder characterized by premature aging. Progressive degeneration is found in a variety of tissues indicated by symptoms such as premature graying and loss of hairs, osteoporosis, arteriosclerosis, early onset of bilateral cataracts, type 2 diabetes, skin atrophy and ulcers. Arteriosclerosis, malignancies and the associated abnormalities result in potentially life-threatening complications. Premature arteriosclerosis, caused by thickening, hardening and loss of elasticity of artery walls, gives rise to a number of vascular diseases such as myocardial infarction, ischemic legs, skin atrophy and ulcers. These clinical symptoms suggest a predisposed vascular disorder, whereas the mechanistic origin remains largely unclear. Our current work employs reprogramming and gene editing technologies to generate patient-specific mesenchymal stem cells (MSCs) and the isogenic MSCs with the causative mutation (*WRN*) corrected. The gene-corrected MSC has an enhanced function in promoting angiogenesis of endothelial cells (ECs). We compared the pro-angiogenic trophic factors secreted by these MSCs, and found hepatocyte growth factor (HGF) diminished in WS MSC. HGF plays a central role in supporting angiogenesis, morphogenesis, tissue regeneration and homeostasis. Knockdown of HGF in MSC decreased angiogenesis of ECs. Therefore, we suggest that deficit of HGF engenders the impaired pro-angiogenesis function of stem cells, signifying a mechanism for the pathogenesis of vascular aging in WS.

2102F

The Impact of prenatal diagnosis and early hormonal therapy (EHT) on autism spectrum traits (AST) in boys with 47,XXY (Klinefelter syndrome). S. Chea¹, C. Samango-Sprouse^{1,2,3}, T. Sadeghin¹, A. Gropman⁴. 1) The Focus Foundation, Davidsonville, MD; 2) Department of Pediatrics at George Washington University, Washington, DC; 3) Department of Human and Molecular Genetics at Florida International University, Miami, FL; 4) Children's National Medical Center, Washington, DC.

Background: 47,XXY is the most frequently occurring X&Y chromosomal variation (1:650), yet 75% of cases remain unidentified in their lifetime. The social-cognitive phenotype of these boys includes language-based learning disorders, speech and motor delays, and executive dysfunction. Autism Spectrum Disorder (ASD) has been reported to be increased in this population, but research has been confounded by many salient factors. This study investigated the incidence of Autism Spectrum Traits (AST) within the social-cognitive pathway, as well as the effects of Early Hormonal Therapy (EHT) and prenatal diagnosis on AST in boys with 47,XXY. **Method:** 74 boys with 47,XXY between the ages of 2-7 years underwent neurodevelopmental evaluations, including the Gilliam Autism Rating Scale, 2nd Edition (GARS-2) and Social Responsiveness Scale, 2nd Edition (SRS-2). 22 boys (30%) with 47,XXY received EHT; 52 (70%) did not. EHT was administered based on patients' pediatric endocrinologist's assessment of phallus size compared to neurotypical boys of the same age. 57 of the 74 boys (77%) were prenatally diagnosed (PRE); 17 (23%) were postnatally diagnosed (POST). **Results:** On the GARS-2, boys in the EHT group showed significant improvement in their social interaction compared to the non-EHT group ($P=.026$). There was also a significant difference in the probability of ASD between the two ($P=.038$), where the non-EHT group had a 30.1% likelihood for ASD traits compared to 13.6% in the EHT group. Positive effects of EHT were seen in the social cognition domain of the SRS-2 ($P=.041$) for the treated group. Within the non-EHT group, PRE boys with 47,XXY had greater communication ($P=.033$) and social interaction ($P=.002$) than POST on the GARS-2. The probability of ASD was also significantly less in the non-EHT PRE boys compared to POST ($P=.004$). Untreated PRE boys had significantly better social awareness ($P=.046$) and motivation ($P=.004$), and less autistic mannerisms ($P=.026$) on the SRS-2. **Conclusion:** This study provides further evidence that this subset of boys with 47,XXY present with an increased risk for disturbances on the social-pragmatic pathway within ASD. The etiology of this vulnerability is unknown but could be derived from family history of ASD, parental origin of the additive X, timing of diagnosis, EHT, and/or other unknown factors. This origin of the vulnerability warrants further investigation to identify those infants at greatest risk who then need more specialized support.

2103W

The genetics of situs inversus totalis without primary ciliary dyskinesia. M.C. Postema¹, A. Carrion-Castillo¹, S.E. Fisher^{1,2}, G. Vingerhoets³, C. Francks^{1,2}. 1) Department of Language & Genetics, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands; 2) Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, Netherlands; 3) Department of Experimental Psychology, Ghent University, Belgium.

Situs inversus totalis (SIT), a left-right mirror reversal of the viscera, sometimes occurs with recessive Primary Ciliary Dyskinesia (PCD). However, most people with SIT do not have PCD, and the etiology of their condition remains poorly studied. Those without PCD may have an elevated rate of left-handedness, implying developmental mechanisms linking brain and body laterality. We sequenced the genomes of 15 people with SIT, of which six had PCD, and 15 controls. The SIT subjects with PCD all had likely recessive mutations in genes already known to cause PCD and SIT. Two non-PCD cases also had recessive mutations in known PCD genes, suggesting reduced penetrance for PCD in some SIT cases. One non-PCD SIT case had a recessive mutation in *PKD1L1*, which has previously been linked to SIT. However, six of the nine non-PCD SIT cases, including most of the left-handers, had no obvious monogenic basis for their condition. Environmental influences, or possible random effects in early development, must be considered.

2104T

The rare diseases clinical research network (RDCRN): Translational research in action. *T.K. Urv¹, L. Reineck², L. Griffith³, J. Witter⁴, J. Wan⁵, T. King⁶, W. Merritt⁷, C. Roy⁸, J. Morris⁹.* 1) NCATS/ORDR, NIH, Bethesda, MD., USA; 2) NHLBI, NIH, Bethesda, MD., USA; 3) NIAID, NIH, Bethesda, MD., USA; 4) NIAMS, NIH, Bethesda, MD., USA; 5) NIDCR, NIH, Bethesda, MD., USA; 6) NICHD, NIH, Bethesda, MD., USA; 7) NCI, NIH, Bethesda, MD., USA; 8) NIDDK, NIH, Bethesda, MD., USA; 9) NINDS, NIH, Bethesda, MD., USA.

A Rare Disease is one that affects no more than 200,000 individuals in the United States. Collectively there are an estimated 7,000 diseases that fall into this category; cumulatively there are approximately 25 million people in the United States who are affected by rare diseases indicating that 'Rare Diseases' are a significant public health concern. Despite advances in our understanding of the causes and mechanisms of many rare diseases, effective treatments are available for fewer than 5%. Although advances in technologies have recently led to promising and potentially transformative treatments, there are challenges in bringing effective treatments to individuals living with rare diseases. First, making a diagnosis can be challenging with many patients experiencing a "diagnostic odyssey" of months or years because of limited knowledge of the range of disease manifestations. Second, there are often no high quality natural history datasets documenting how the disease affects patients' functioning, and how it progresses over time. Third, there are often no adequate clinical or biological markers to support the clinical development of new therapeutics. Fourth, the small number of patients lead to challenges in the design and implementation of studies. Finally, the resources available for therapeutics development are limited, making it critical to establish partnerships among patient groups, industry, academic investigators and federal funding agencies. To address these challenges the Rare Diseases Act of 2002 directed the NIH to support consortia for advancing prevention, diagnosis, clinical readiness, and treatment of rare diseases. Since its inception, numerous NIH institutes and centers have partnered to support the Rare Diseases Clinical Research Network (RDCRN). The RDCRN is a collaborative network of investigators and patient groups committed to the investigation of rare diseases working in partnership with leaders in technology to enhance communication and sharing of resources through a multidisciplinary approach. To date this program has successfully supported 31 individual consortia that conducted research on 238 individual disorders. This presentation will outline the successes and challenges of this translational research network for rare diseases. It will also highlight the goals of the latest iteration of the network and will highlight opportunities to build bridges from basic to clinical research and beyond research to therapy.

2105F

The Birth Defects Research Laboratory biorepository of human fetal tissue. *D. O'Day¹, K. Aldinger², B. Nelson², K. Millen², D. Doherty¹, I. Glass¹.* 1) University of Washington, Seattle, WA; 2) Seattle Children's Research Institute, Seattle, WA.

The Birth Defects Research Laboratory (BDRL), a NIH funded repository of both normal and abnormal human fetal tissue collected from elective terminations and spontaneous loss, provides research material to academic and non-profit organizations. We systematically undertake assessments of tissue and organ preparation and storage protocol's for histological analyses, bulk nucleic acid or protein analyses, cultured cell lines and single cell/nuclei analyses. We are currently evaluating protocols to optimize collection and processing of fetal tissues to facilitate entry into whole tissue and single cell/nuclei protocols. Methods vary by starting material, dissociation method, and storage conditions, as well as downstream applications. For example, we performed RNA-sequencing of whole tissue and single cell RNA-sequencing of fixed cells from the same specimen for several organ systems that were isolated either immediately or after storage in Hibernate E at 4°C for 24 hours. While few differences were detected between whole tissue RNA, using several next-generation single cell approaches, we observed decreased representation of certain cell types, including progenitor and neuronal subpopulations, in the 24 hour brain sample possibly due to increased storage time prior to processing. Our results highlight the importance of immediate processing for brain tissue for single cell RNA-sequencing analysis to minimize the potential loss of certain cell populations. In addition to providing bulk tissues, we collaborate with recipients to optimize the storage conditions of tissues and processing methods based on their biological questions and techniques. All tissues, cultured cells, and single cell/nuclei preparations that are not distributed to recipients are stored for future use. Tissues and their derivatives provided by BDRL have been instrumental for major impact research projects such as ENCODE-Roadmap, BrainSpan Atlas of the Developing Human Brain, and over 60 additional NIH-funded investigators over the past 2 years.

2106W

Non-coding mutations causing NCMD in a new family: Multimodal imaging provides insight into the development of the human macula. K.W. Small^{1,2}, F.S. Shaya^{1,2}, L.R. Small³, N. Udar^{1,2}, E.M. Stone⁴. 1) Macula & Retina Institute, Glendale, CA; 2) Molecular Insight Research Foundation, Glendale and Los Angeles, CA; 3) Bascom Palmer Eye Institute, University of Miami, Miami, FL, United States; 4) University of Iowa, Iowa City, IA, United States.

Purpose: To describe multimodal imaging and corresponding functional studies in a new family with North Carolina Macular Dystrophy (NCMD/MCDR1) incorporating OCT angiography (OCT-A) for the first time to our knowledge. **Methods:** A descriptive, retrospective study of a family with NCMD. Diagnostic multimodal imaging and functional testing of the retina included color fundus photographs, fundus auto fluorescence, fluorescein angiography (IVFA), spectral domain OCT (SD-OCT), OCT angiography (OCT-A), multifocal ERG (mf-ERG), full field ERG, and microperimetry. DNA sequencing was performed using Sanger sequencing. IRB approval was obtained. **Results:** Three subjects, representing 3 generations of a single family each demonstrating a different grade of NCMD, underwent clinical and genetic testing. Multimodal imaging helped to demonstrate the developmental nature of the retinal and choroidal lesions in each and the extent of visual function that was suggestive of the stage of development that is affected by methylation at the 6q16 locus. Genetic testing demonstrated the V2 point mutation (French / German mutation) in the DNASE 1 hypersensitivity binding site on chromosome 6q16 causing overexpression of the retinal transcription factor PRDM13. The oldest family member has many similarities to foveal hypoplasia / foveal plana. **Conclusion:** NCMD has great phenotypic variability, which can only be appreciated by examining multiple family members. This is the first report to our knowledge showing a correlation of functional studies with multimodal imaging including OCT-A. All layers of the retina and choroid demonstrate mal-development and varying degrees of malfunction. Although PRDM13 is over expressed in the amacrine cells, we have yet to demonstrate an abnormality specific to this cellular layer. The retinal vasculature seems surprisingly well preserved / intact by OCT-A when compared to the IVFA. OCT-A suggests foveal hypoplasia to be a phenocopy of grade 1 NCMD. OCT-A does reveal vascular abnormalities not found with standard IVFA. We propose that the grade of severity is determined by the embryonic stage that methylation occurs at the DNASE1 hypersensitivity binding site which regulates the expression / overexpression of the retinal transcription factor PRDM13. We propose that there is a similar control of expression at the MCDR3 / IRX1 site.

2107T

Genetics of cortical dysplasias and epilepsy treated surgically. C.A.B. Garcia¹, H.R. Machado¹, W.A. Junior¹, T.L. Teixeira¹, M.S. Volponi¹, L.L. Ball², K.N. James², J.J. Gleeson². 1) University of São Paulo, Ribeirão Preto, Brazil; 2) Howard Hughes Medical Institute (HHMI).

Focal cortical dysplasias are a group of cerebral cortical developmental malformations characterized by changes in the microarchitecture of the cerebral cortex, whether or not accompanied by abnormal cells, mainly dysmorphic (giant, dysplastic) neurons and balloon cells, is considered the most common cause of epilepsy refractory to drug treatment in the pediatric population. Hemimegalencephaly is a clinically devastating malformation of cortical development in children. Studies demonstrated that genetic factors are involved in the HME. Ten patients diagnosed with HME were selected, with an indication for surgical treatment in the Neuropediatrics outpatient clinic of the Hospital das Clínicas of the Medical School of Ribeirão Preto - USP. The surgical protocol consists of a biopsy of the damaged tissue, being fractioned in two parts, one for the morphological analysis and another frozen fraction in liquid nitrogen for analysis through the exome sequencing. The objective of this study is to identify the specific genetic alterations leading to HME by analyzing the genes of mTOR signaling pathways as well as other parallel tracks. The microscopic exam of most patients show histological sections of the neocortex stained by H&E show extensive areas of neuronal loss and neuronal bodies grouped, resulting in architectural distortion, mainly of the laminar distribution. Additionally, there are some rare foci of giant neurons, with large nuclei and wide cytoplasm. In some areas, the cortex presents reduction of the number of layers (four layers), with fusion of the gyri. The analysis of the results identified somatic mutations in the same PIK3CA gene, in two cases, being considered one of the genetic causes for patients diagnosed with HME. For case 1, with a rate of 22%, c.1258T> C (p.C420R) and case 2, with a rate of 11%, c.1624G> A (p.E542K). The selection of 50 most expressed genes, with several types of mutations, and their impact on the phenotype were also performed in these patients. We conclude that these mutations are associated with increased mTOR signaling in affected brain regions. These results may contribute to the identification of the genes responsible for HME and thus to understand the mechanisms associated with the evolution of the pathology. In addition, to contribute to the discovery of new therapies for epileptic seizures caused by these developmental lesions, with modulators for the mTOR pathway.

2108F**Optimization of CRISPR/Cas9-mediated gene editing in monkey embryos.**

J. Jiang^{1,2}, X. Luo^{1,2}, B. Su¹. 1) State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China; 2) Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming 650204, Yunnan, China.

The RNA-guided Cas9 nuclease from the microbial clustered regularly interspaced short palindromic repeats (CRISPR) adaptive immune system have been used to facilitate efficient genome editing in many model and non-model animals. However, its application in non-human primates is still at early stage, and the efficiency of gene modification in primate embryos remains rather low. Here, using zygotic injection of Cas9 mRNA and multiple single-guide RNAs with an optimized protocol, we achieved a high knockout efficiency of 86.6% for MCPH1- a human autosomal recessive primary microcephaly gene. Furthermore, through homology directed repair (HDR), we obtained monkey embryos with humanized MCPH1, and the efficiency is 21.4%. The improved gene editing efficiency in monkeys may serve as a useful tool in human disease and human evolution studies.

2109W**Pharmacogenomics role in tuberculosis treatment: A view in Peruvian patients.**

H. Guio¹, K. Levano¹, D. Tarazona¹, C. Sanchez¹, L. Jaramillo¹, S. Capristano¹, R. Zegarra¹, L. Solari¹, A. Soto², A. Mendoza³, J. Cabrera³, C. Rojas¹. 1) Instituto Nacional de Salud, Lima, Peru; 2) National Hospital Hipolito Unanue; 3) Peruvian National Health Strategy for Prevention and Control of Tuberculosis.

INTRODUCTION: N-acetyltransferase 2 (NAT2), cytochrome P4502E1 (CYP2E1) and arylacetamide deacetylase (AADAC) are the enzymes responsible for metabolizing isoniazid (INH) and rifampicin (RIF), first-line anti-tuberculosis drugs. Variation in the genes responsible for these enzymes could affect the metabolism of INH and RIF and thus its efficacy in tuberculosis. This study is the first to analyze NAT2, CYP2E1 and AADAC genotypes in a Peruvian population with tuberculosis. 500 drug-sensitive tuberculosis patients were genotyped for NAT2, CYP2E1 and AADAC gene polymorphisms by sanger sequencing. **RESULTS:** We found seven haplotypes consisting of 6 SNPs and 14 NAT2 genotype variations in Peruvian population. According to bimodal distribution, the predicted phenotype was composed of 53% rapid and 47% slow acetylator, CYP2E phenotype was composed of 37% rapid and 63% slow acetylator, together 25.2% was slow acetylator and could have risk of developing hepatotoxicity to isoniazid. On the other hand, AADAC phenotype was composed of 82% rapid and 18% slow acetylator, however present similar CLint values to rifampicin. We did not find any association between slow acetylator to isoniazid or rifampicin and hepatotoxicity ($p > 0.05$). In addition, we found hepatotoxicity to isoniazid in patients with higher cholesterol levels and received their anti-tuberculosis treatment in a specific health region ($p < 0.05$). We continue to evaluate the genotypes of glutathione S-transferase (GST) as an important phase II detoxification enzyme and be able to complete the metabolic cycle. **CONCLUSION:** Different genotypes identification and association with different comorbidity would help to determine a new treatment scheme adjusting frequency and dose, foresee MDR-TB development in Peruvian population and allow future proposals to personalized medicine that represents a challenge in environments with limited resources.

2110T

Cohen syndrome-associated gene *VPS13B* is required for male fertility and spermiogenesis in mice. *M. Bordessoules¹, M. Guilleman², V. Carmignac-Quere¹, V. Lhussiez¹, C. Bruno², P. Fauque², C. Choux², J. Barberet², C. Thauvin-Robinet^{1,3}, L. Faivre^{1,3}, L. Duplomb¹, R. Da Costa¹.* 1) INSERM UMR 1231 GAD team, Genetics of developmental disorders, University of Bourgogne Franche-Comté, F-21000 Dijon, France; 2) Service de biologie de la reproduction et centre d'étude et de conservation des œufs et du sperme humains (CECOS), F-21000 Dijon, France; 3) Centre de Référence maladies rares "Anomalies du Développement et syndromes malformatifs" de l'Est, Centre de Génétique, FHU TRANSLAD, CHU Dijon, F-21000 Dijon, France.

Cohen syndrome is a polymalformative autosomic recessive disorder due to mutations and chromosomal rearrangements affecting the *VPS13B* gene. Individuals with CS display microcephaly, intellectual disability, progressive retinopathy, facial dysmorphism, truncal obesity and intermittent neutropenia. In order to comprehend the cellular and molecular basis causing those symptoms, we generated a novel *Vps13b*^{Δex3/Δex3} mouse model. Among the first and most striking phenotypes, we observed male infertility. Intellectual disability does not permit CS individuals to care for children and become parents, and therefore no fertility issues have yet been reported in patients. In *Vps13b*^{Δex3/Δex3} mice, while dissections showed a fully developed reproductive tract, epididymal dilacerations revealed a low count of non-motile and structurally abnormal spermatozoa. Histological analysis of testicular sections demonstrated a reduction in late-stage spermatids compared to controls. TUNEL assay was performed to confirm spermatid death. Periodic Acid Schiff staining failed to evidence acrosomes in all spermatid stages. As in patient fibroblasts and cellular models for CS, electron microscopy showed a disorganized Golgi apparatus. In addition, no pro-acrosomal vesicle was found docking onto the spermatid nuclear membrane. Altogether, our data supports the putative function of *VPS13B* in cellular trafficking and suggests that (1) male individuals with CS are likely to be affected with oligoasthenoatozoospermia and (2) other spermiogenesis factors may be hidden among genes associated with rare syndromic forms of intellectual disability for which there is no data on the patients' fertility.

2111F

Identification of novel transcription factor in mid-brain development using alternative transcription factor binding site prediction method. *H.S. Lee^{1,2,3}, H.J. Lee^{1,3}, Y.H. Shin², S.H. Bae², J.M. Kang², Y.E. Lee², J. Moon².* 1) Biomarker Center, Biomedical Research Institute, Seoul National University Hospital, Seoul, Republic of Korea.; 2) General Medical Research Institute, Bundang Medical Center, CHA University, Seoul, Republic of Korea.; 3) co-author.

Embryonic development in mammals involves function of a plethora of transcription factors. Cell fate specification relies on the action of critical transcription factors that become available at distinct stages of embryonic development. The development of mid-brain Dopaminergic neurons (mDAn) is controlled by a combination of between cell extrinsic pathway (SHH-FGF8 or Wnt1) and intrinsic signals (FOXA2, LMX1A). Both Nurr1 and Pitx3 are required for mDAn differentiation and survival. Human neural progenitor cells were more efficient differentiation into dopamine neurons in the treatment to the small chemical compound X (sCCX) than other conventional differentiating factors such as SHH&FGF8, confirmed by in vitro results. To understand the developmental patterns of gene expression in the dopamine differentiation of human neural progenitor cells, we performed transcriptome analysis of targeting neurons at different treatment time points by in vitro culture with or without small molecules or SHH or negative control condition using next-generation RNA sequencing (RNA-seq). We showed that distinct gene expressions shown by different factor-treated cells (negative, SHH & FGF8 and sCCX) and most genes were down-regulated and four genes uniquely upregulated by sCCX and predicted to be related to neurogenesis by bioinformatics data analysis. Furthermore, that genes uniquely up-regulated by sCCX, Nurr1 and TH act in a coordinated manner for dopamine differentiation and they have motifs in its promoters which novel transcription factor X (nTFX) binds by alternative transcription factor binding site prediction method. We found that nTFX expressed midbrain of E10 embryonic mouse but not midbrain of over E12 embryonic mouse brain. The nTFX protein merged with TH+ cells and bind to the promoter of TH gene. Our results suggest that a nTFX is one of the required factor for mDAn specification during brain development.

2112W

Ret loss of function leads to changes in developmental trajectories of single cells in the developing enteric nervous system of a mouse model of Hirschsprung disease. E.A. Vincent¹, S. Chatterjee², G. Cannon¹, D. Auer¹, H. Ross¹, A. Chakravarti², L.A. Goff^{1,3}. 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD; 2) Center for Human Genetics and Genomics, NYU Langone Medical Center, New York, NY; 3) Department of Neuroscience, Johns Hopkins School of Medicine, Baltimore, MD.

During development enteric neural crest cells (ENCC) migrate rostro-caudally along the intestinal tract where these progenitors eventually differentiate into the neurons and glia that form the enteric nervous system (ENS). In Hirschsprung disease (HSCR) the ENS fails to form in the distal colon, leading to pseudo-obstruction and severe distension. The majority of HSCR cases are attributed to loss of function (LOF) of *RET*, which encodes a tyrosine kinase receptor expressed by all ENCC, yet the precise mechanisms by which *RET* LOF leads to HSCR remain unknown. We hypothesize that *RET* LOF leads to precocious differentiation of ENCC, thereby depleting the progenitor pool prior to reaching the distal colon, and leading to changes in acquired cell fate. To address our hypotheses, we performed single-cell RNA sequencing on FACS-enriched ENCC in the developing gut of *Ret*^{+/+} and *Ret*^{-/-} mice. These cells include uncommitted progenitors, maturing glial, and neuronal cell types. In developing glial cells, differential gene expression analysis with respect to genotype identified genes associated with cell cycle regulation and neuronal differentiation. *Ret*^{-/-} cells were enriched for mature glial markers, consistent with a model of precocious differentiation. Pseudotemporal analysis of the neuronal lineage identified several developmental trajectories associated with maturation of distinct neuronal cell types. Within this context, *Ret* LOF leads to restriction in the lineage potential of developing neurons. One lineage associated with decreased expression of *Ret* remains accessible to *Ret*^{-/-} cells, while another, in which *Ret* expression is maintained, is inaccessible to *Ret*^{-/-} cells. Importantly, while neural trajectories that require *Ret* also express the canonical Ret co-receptor Gfra1, trajectories that do not require *Ret* express other, noncanonical co-receptors in the Gfra family. Additional analyses identified differential genes with respect to genotype, sex, developmental age, and cell type. These results provide key insights into how the organization and development of the ENS are affected in HSCR, and identify sex-dependent changes in differentiation rate and cell fate commitment in a *Ret* LOF context. These data yield insights into the complex developmental processes involved in establishment of the mammalian ENS, and demonstrate the specific roles for Ret signaling in regulation of ENS progenitor proliferation, differentiation, and cell fate commitment.

2113T

Biallelic inactivation of HEPHL1 impairs ferroxidase activity and causes abnormal hair phenotype. P. Sharma¹, M. Reichert¹, T.C. Markello^{1,2,3}, G. Golas¹, D.R. Adams^{1,2,3}, P.J. Steinbach⁴, Y. Lu⁵, B. Fuqua⁶, C.D. Vulpe⁶, G.J. Anderson⁵, W.A. Gahl^{1,2,3}, M.C.V. Malicdan^{1,2,3}. 1) NIH Undiagnosed Diseases Program, Common Fund, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA; 2) Medical Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA; 3) Office of the Clinical Director, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA; 4) Center for Molecular Modeling, Center for Information Technology, National Institutes of Health, Bethesda, MD, USA; 5) Iron Metabolism Laboratory, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia; 6) Center for Environmental and Human Toxicology, Department of Physiological Sciences, University of Florida, Gainesville, Florida, USA.

The maintenance of the correct redox status of iron is functionally important for several critical biological processes. At the molecular level, multicopper ferroxidases play an important role in oxidizing ferrous iron, released from cells, into ferric iron, which is subsequently distributed by transferrin. There are two well-characterized ferroxidases, ceruloplasmin (CP) and hephaestin (HEPH) that facilitate this reaction in diverse tissues. Recently, a novel ferroxidase has been identified called Hephaestin like 1 (HEPHL1), also known as Zyklopen. Here we report a child with compound heterozygous mutations in *HEPHL1* (NM_001098672), who presented with abnormal hair (pili torti and trichorrhexis nodosa), combined-type ADHD, speech articulation disorder, severe heat intolerance, and chronic leg pain. The maternally inherited mutation (c.809_1063del; p.ala270_ala355del) affects mRNA splicing and eliminates exon 5, causing an in-frame deletion of 85 amino acids. The paternally inherited mutation (c.3176T>C; p.Met1059Thr) changes a highly conserved methionine that is part of a typical type I copper binding site in HEPHL1. We further demonstrate that HEPHL1 has ferroxidase activity and that the two mutants found in this patient exhibit loss of this activity, likely due to loss of copper binding. These results suggest that the biallelic mutations we identified in our proband are loss-of-function mutations. We subsequently generated a loss-of-function *Heph1* model in mice. *Heph1* knockout mice are viable, and have curly whiskers throughout their life, consistent with the hair phenotype in the patient with loss-of-function *HEPHL1* mutations. Overall, our results enhance our understanding of the function of HEPHL1 and implicate altered ferroxidase activity in hair growth and hair disorders.

2114F

aPKC modifier genetic screen for suppressors of epithelial junctional defects in *C. elegans*. J.G. Montoyo-Rosario¹, J. Nance^{1,2}. 1) Helen L. and Martin S. Kimmel Center for Biology and Medicine at the Skirball Institute of Biomolecular Medicine, NYU School of Medicine, New York, New York 10016, USA; 2) Department of Cell Biology, NYU School of Medicine, New York, New York 10016.

Epithelial cells are specialized cells that line our organs that possess adhesive junctions enabling them to function as a barrier to the outside environment. Loss of junctions compromises tissue structure and has been linked to human diseases such as kidney disease, cancer, and birth defects. The scaffolding protein PAR-6 is a conserved regulator of junction maturation. PAR-6 is known to function together with the kinase aPKC, but the targets of aPKC that are important for junction formation are largely unknown. We found that junctions in the *C. elegans* spermatheca can break in animals carrying a temperature-sensitive allele of *pkc-3* (aPKC), leading to sterility. We also used a PKC-3 protein degradation strategy and found that junctions in embryos failed to condense into continuous junctions. To find genes that may function with *pkc-3* in regulating junctions, we performed a large-scale suppressor screen to find *pkc-3* mutant animals that are now fertile. In addition to intra-genic mutations in the *pkc-3* gene, we identified a new allele of a known *pkc-3* regulator (*lgf-1*). In addition, we identified two different alleles of *F22F4.1*, an uncharacterized gene that encodes a secreted protein expressed in a subset of embryonic epithelia. We will report on our characterization of these *pkc-3* suppressors and their function in junction maturation.

448W

Fetal echocardiographic features and whole exome sequencing results of ventricular non-compaction cardiomyopathy. X.Y Hao¹, X.W Liu², H.R Sun³, Y.H He⁴. 1) Echocardiography Department, Beijing An Zhen Hospital, Capital Medical University, Beijing, China; 2) Echocardiography Department, Beijing An Zhen Hospital, Beijing, China; 3) Beijing An Zhen Hospital, Capital Medical University, Beijing, China; 4) Beijing Key Laboratory of Maternal-Fetal Medicine and Fetal Heart Disease, Beijing Anzhen Hospital, Capital Medical University, Beijing, China.

Objective: To investigate the echocardiographic diagnosis for fetal ventricular non-compaction cardiomyopathy (VNCC) and to analyze the genetic associations with VNCC. **Methods:** A total of 13 fetuses were diagnosed as VNCC from 30009 fetal echocardiographic studies in Beijing Key Laboratory of Maternal-Fetal Medicine and Fetal Heart Disease from October 2010 to March 2018. Fetal umbilical cord tissue samples and the parental peripheral blood were collected after the families opted for termination of the pregnancy. Whole exome and Sanger sequencing were performed using the biological specimens.

Results: In this cohort, 6 fetuses were diagnosed with biventricular, 5 were with left ventricle, and 2 were with right ventricle and the gestational weeks of the fetuses range from 20 to 34 years (median week, 25 weeks). Among the 13 fetuses with VNCC, 11 cases were terminated after a detailed prognosis consultation during pregnancy, 1 fetus was born, 1 case lost follow-up. In addition, 8 cases were associated with congenital heart disease (CHD), and 4 cases with fetal arrhythmia, and the average cardiovascular profile score of fetuses was 7.15. There were VNCC mostly involved the apical and lateral wall of the LV. The non-compacted layer to compaction layer ratio of RV was more than LV. For genomic study, 1 case had LDB3 gene mutation (c. T964C / p. S322P), and 1 had PRKAG2 gene mutation (c. G1592A / p. Arg531Gln), both of the two mutations were de novo. In addition, 3 cases came from one family were associated with pathologically significant mutation which inherited from their mother who didn't combine any representation. **Conclusion:** Fetal VNCC can involve both the left and the right ventricles. Prenatal echocardiography combined with WES could play an important role in the timely diagnosis, prognosis and parental counseling of patients with VNCC.

449T

Comparison of the prevalence of pathogenic copy number variation in tetralogy of fallot and associated congenital heart defect diagnosed between fetuses and children. T. Man¹, Y. He², X. Hao¹, H. Sun¹, S. Ge². 1) Beijing Anzhen Hospital and Capital Medical University, Beijing, China; 2) Pediatrics, St. Christopher's Hospital for Children, Blue Bell, PA, USA.

Objectives: To ascertain the use of whole genome low coverage sequencing in TOF and associated defects in fetuses and children. **Methods:** A total of 143 with TOF (70 fetuses, and 73 children) confirmed by pathological anatomy were analyzed from the data base at the Beijing Key Laboratory of Fetal Heart Defect/Maternal Fetal Medicine. Demographic, echocardiography and whole genome low coverage sequencing (×1) data were reviewed to investigate the association between genetic abnormalities and TOF. **Results:** 21 cases (21/70; 30%) in fetuses had pathogenic CNVs. 1 with Smith-Magenis syndrome, 12 with 22q11.2 microdeletion syndrome, 1 with 22q11.2 microduplication, 1 with 1p36 microdeletion syndrome, 1 with 17p12 microdeletion, 2 with Down syndrome, 2 with Edward syndrome and 1 case of Patau syndrome. While 8 cases (8/73; 11%) in fetuses had pathogenic CNVs. 5 with 22q11.2 microdeletion syndrome, 1 with 22q11.2 microduplication, and 1 with Xp22.2 microduplication. There was significant difference in pathogenic CNVs between fetuses and children ($p < 0.01$). In this cohort, 38 cases (38/70; 54%) in fetuses and 39 cases (39/73; 53%) in children were complicated with other intracardiac malformations. For the cases with pathogenic CNVs, 13 cases (13/21; 62%) in fetuses were complicated with other intracardiac malformations. 7 with right aortic arch, 5 with absent ductus arteriosus, 4 with aberrant left subclavian artery, and 3 cases of persistent left superior vena cava. And 5 cases (5/8; 63%) in children were complicated with other intracardiac malformations. 3 with right aortic arch, 3 with intramuscular ventricular septal defect, 1 with atrial septal defect, 1 with patent ductus arteriosus, and 1 case of persistent left superior vena cava. There was no significant difference in associated other intracardiac malformations between fetuses and children, with both the whole cohort and pathogenic CNVs ($p > 0.05$). **Conclusions:** There was a significant yield of pathogenic CNVs by whole genome low coverage sequencing in fetuses and children with TOF. The prevalence of CNVs in fetuses was higher than children with TOF.

450F

A gene-by-gene interaction associated with the risk of conotruncal heart defects. C. Lyu¹, D.M. Webber², X. Tang³, S.L. MacLeod⁴, C.A. Hobbs³, M. Li¹, the National Birth Defects Prevention Study. 1) School of Public Health, Indiana University Bloomington, Bloomington, IN; 2) Department of Pathology and Immunology, Washington University at St. Louis, St. Louis, MO; 3) Department of Pediatrics, University of Arkansas for Medical Sciences, Little Rock, AR; 4) Center for Translational Pediatric Research, Arkansas Children's Research Institute, Little Rock, AR.

Congenital heart defects (CHDs) are the most common birth defects and the leading cause of infant mortality attributable to birth defects. Conotruncal heart defects (CTDs) are a subtype of CHDs accounting for 20 ~ 30% of CHD cases. It is widely believed that the cause of disease involves a complex relationship among genetic variants and parental lifestyle factors. Nevertheless, most of the existing studies have evaluated each single genetic variant for association with the disease risk. In this article, we focused on the interactions between 13 candidate genes within folate, homocysteine, and transsulfuration pathways for potential association with CTD risk. The discovery phase includes 328 case-parental trios enrolled in the National Birth Defects Prevention Study (NBDPS). To evaluate the interaction of two genes, we applied a conditional logistic regression model to estimate the interaction effect for all possible SNP pairs within two respective genes by contrasting the affected infants with their pseudo-controls. The findings were replicated with an independent replication phase of 86 additional case-parental trios enrolled in the NBDPS. The results of two phases were further integrated by a fixed-effect meta-analysis. One SNP pair (i.e. rs56219526 and rs11892646) located in gene *MTRR* and *DNMT3A*, respectively, involving in homocysteine pathway, were found to be associated with CTD risk (nominal p-value of 6.55e-06). After multiple testing adjustment, the association was marginally significant with an adjusted p-value of 0.06. Further studies with larger sample sizes are needed to confirm and elucidate this potential interaction between two genes.

451W

Risk of congenital anomalies is associated with birth weight in infants of diabetic mothers. S. Ramanathan¹, B. Becerra², N. Kraus¹, R.D. Clark¹. 1) Pediatrics, Loma Linda University Medical Center, Loma Linda, CA, CA; 2) Allied Health Studies, Loma Linda University.

Maternal diabetes is well-known to increase the risk for birth defects. The purpose of this study was to determine the cumulative incidence of infants of diabetic mothers (IDMs) and of congenital anomalies in these infants over a five year period at Loma Linda University Children's Hospital. This was a retrospective chart review of all infants with the International Classification of Diseases (ICD)-9 code 775.0 (Infant of Diabetic Mother- IDM), admitted to the nursery or neonatal intensive care unit (NICU) between 2010 and 2014. There were 465 infants with this code among their diagnoses. Of these, 418 met inclusion criteria (excluded for insufficient clinical data, other teratogenic exposures, genetic syndromes or chromosome abnormalities). Of these, 65 mothers had type 1 diabetes mellitus (T1DM), 125 have type 2 diabetes (T2DM) and 217 had gestational diabetes (GDM). About 21% of the infants (86/418) identified as IDMs at our institution had at least one birth defect over this time period which is double the 10% risk that is reported in the literature. Of these, 48 were born to women with pregestational diabetes mellitus (PGDM): 35 to women with T2DM (41%) and 13 to women with T1DM (15%). Thirty-seven were born to women with GDM (43%) The risk for birth defects was significantly increased in infants who were small or appropriate for gestational age (SGA/AGA), when compared to large for gestational age (LGA) babies. 26% of SGA/AGA babies had birth defects, while 13% of LGA babies had birth defects (Chi-square test =10.919, p<0.05). While only 6% of the IDMs were SGA (25/418), 40% of them had congenital anomalies (10/25). While 50% of the IDMs were AGA (205/418), 25% of had congenital anomalies (51/205). About 45% of the IDMs were LGA (188/218); however, only 13% (35/188) had congenital anomalies. **IDMs with and without congenital anomalies and type of maternal diabetes:**

All (n= 418)	T1DM	T2DM	GDM	Not known
Average for gestational age	25	54	121	5
Small for gestational age	1	9	15	
Large for gestational age	39	62	82	5
With congenital anomalies (n= 86)				
Average for gestational age	7	21	23	
Small for gestational age	1	4	5	
Large for gestational age	5	8	11	1

In our patient population, the risk for congenital anomalies is highest for SGA infants and is lowest for LGA infants. This suggests that the factors that drive the disruption of the embryonic development during the first trimester may be different from those that drive diabetic fetopathy in the second and third trimesters in IDMs.

452T

Targeted cfDNA analysis using DANSR assays for determination of fetal RHD status. S. Saini, R. Foley, P. Hartley, C. Kingsley, E. Wang, P. Bogard, M. Schmid. Roche Sequencing Solutions, 5945 Optical Court, San Jose, CA 95138.

Objective: Cell-free DNA (cfDNA) testing provides high sensitivity for common trisomies and a lower false positive rate than traditional prenatal screening methods. Here we aim to develop a targeted cell-free (cfDNA) test, the Harmony® prenatal test, enhancement that allows determination of fetal RhD status in RhD-negative pregnant women. **Methods:** 12 simulated pregnancy plasma samples with known RHD genotype were prepared by titrating non-pregnant, RHD-positive cfDNA (fetal source) into non-pregnant, female RHD-negative cfDNA (maternal source) to simulate fetal fractions of 5%, 10% and 15%. A 0% sample served as a negative control. Digital Analysis of Selected Regions (DANSR) assays targeting exons 2, 3, 4, 5 and 7 of the RHD gene were added to existing DANSR assays and the generated DANSR products were hybridized onto a custom DNA microarray for analysis of fetal fraction and determination of fetal RHD status using the fetal fraction optimized algorithm FORTE. **Results:** In all 12 simulated pregnancy samples, RHD sequences were detected. The mean RHD signal in each case correlated with fetal fraction and was therefore consistent with an RHD-positive fetal source on the background of an RHD-negative maternal source. As expected, no RHD sequences were detected for samples with 0% fetal fraction. **Conclusions:** Targeted cfDNA testing using DANSR assays has the potential to determine fetal RHD status and be used as a noninvasive screening method to identify pregnancies at increased risk for RhD immunization. .

453F

Prenatal diagnosis of unbalanced XY translocation following a sex discordance between prenatal cell-free DNA screening and ultrasound examination results. L. Lai, A. Wiita, Z. Qi, J. Yu. Laboratory Medicine, University of California San Francisco, San Francisco, CA.

With the general acceptance of prenatal cell-free DNA (cfDNA) screening, more women have access to fetal DNA information. Gender discordance is observed in 1 in 1500 to 2000 pregnancy in patients who undertake prenatal cfDNA screening. The phenotypic females with male cfDNA screening result have been reported in one patient with 9p24 deletion including *DMRT1* gene which encodes a male-specific transcriptional regulator and one patient with mosaic monosomy X/XY (90%/10%), but not yet in patient with unbalanced XY translocation. We report a case involved a 37-year-old G1 woman who was seen at 21 weeks GA for genetic counseling because of female gender on ultrasound with male gender reported on cfDNA screening. She underwent cfDNA screening (MaterniT21 PLUS test, Sequenom Laboratories, San Diego, CA, USA) at 12 weeks gestation based on advanced maternal age. Her cfDNA screening was negative for trisomy 21, 18, and 13 and positive for Y chromosomal material. The fetal fraction was reported at 10%. Subsequent amniocentesis with chromosome analysis and SNP microarray revealed an unbalanced XY translocation with the breakpoints at Xp22 and Yq11, in addition to a normal X chromosome (ISCN nomenclature: 46,X,der(X)t(X;Y)(p22.33;q11.21).arr[GRCh37] Xp22.33(60814_2697868)x1,Xp22.32q28(2834723_154927199)x2,Yq11.21q12(14604583_28817458)x1). This unbalanced XY translocation resulted in one copy loss within the terminal pseudoautosomal region 1 (PAR1), two copies of non-PAR1 region of X chromosome, and one copy of Yq11qter excluding *SRY* gene. Further metaphase FISH analysis confirmed the loss of *SRY* gene. XY translocations are relatively rare and females with t(X;Y)(p22;q11) are typically phenotypically normal except for short stature due to haploinsufficiency of *SHOX* gene, while the males may have various abnormalities including intellectual disability, infertility, or more severe anomalies. However, a recent long-term follow-up study suggests that additional phenotypes, including intellectual disability, facial dysmorphism, eye abnormalities, or skeletal defects, may be present in a minority of female t(X;Y) carriers. To our knowledge, this is the first reported case of a rare unbalanced XY translocation with prenatal sex discordance between cfDNA screening and ultrasound results and demonstrating the importance of chromosomal/microarray analysis in detecting rare structure rearrangements relevant to the diagnosis of ambiguous genitalia in fetus.

454W

Case report: Heterozygous FGF10 deletion associated with severe pulmonary hypoplasia causing neonatal deaths in a family with Lacrimo-auriculo-dento-digital (LADD) Syndrome. *T.M. Bartell, M.H. Lipson.* Genetics, Kaiser Permanente, Sacramento, CA.

Lacrimo-auriculo-dento-digital (LADD) syndrome results from pathogenic variants in FGFR2, FGFR3 and FGF10 genes. It is associated with abnormalities of the lacrimal and salivary glands, teeth, ears, digits, limbs and genitourinary system. LADD Syndrome demonstrates variable expressivity and autosomal dominant inheritance with a few sporadic cases. We report a family with LADD syndrome who lost two neonates in two generations to severe pulmonary hypoplasia. The family has a large, interstitial deletion of 2.08 Mb at 5p11p12 which contains the entire FGF10 gene. FGF10 is responsible for the branching morphogenesis of many organs including the lungs. Homozygous and heterozygous mutations of the FGF10 gene are reported to cause severe pulmonary hypoplasia in mice. There are no cases of isolated severe pulmonary hypoplasia in humans with heterozygous FGF10 mutations in the literature. The occurrence of severe pulmonary hypoplasia is underreported in LADD Syndrome caused by pathogenic FGF10 variants, and is an important aspect for prenatal genetic counseling.

455T

Preimplantation genetic diagnosis for nephrotic syndrome. *M.T. Akbari^{1,2}, F. Khordadpoor Deilamani².* 1) Tarbiat Modares University Faculty of Medical Sciences, Tehran, Iran; 2) Tehran Medical Genetics laboratory, Tehran, Iran.

Introduction: *NPHS2* is a causative gene for autosomal recessive steroid-resistant nephrotic syndrome. This syndrome is characterized by proteinuria, hypoalbuminaemia and oedema. Carrier couples are at 25% risk for producing an affected child. Here we report the first successful pregnancy in Iran in a family with a previous affected child following single cell embryo biopsy and PGD for the detection of c.156del (p.(Thr53Profs*46)) mutation in the *NPHS2* gene. **Materials and Methods:** Two different tests (STR based linkage analysis and Cycle sequencing of exon 1 of *NPHS2* gene) were used to study the transmission of c.156del (p.(Thr53Profs*46)) mutation in *NPHS2* gene to the embryos. The PGD workup started with DNA sample analysis of the prospective parents and their affected son. STR haplotyping for family members was performed, in order to identify the most informative STR markers linked to the disease causing gene, to be used in the following clinical PGD cycle. Blastomeres from 6 embryos were received from IVF center. Cell lysis, DNA extraction and whole genome amplification (WGA) were performed on each blastomere. The WGA products were then analyzed using the above mentioned tests in order to determine the mutation status. **Results:** WGA was not successful for two embryos. One embryo was homozygous for mutant allele and one another embryo was homozygous for normal allele which was transferred to the uterus. Two embryos showed discrepancy between sequencing and linkage analysis which may be due to allele drop out. **Conclusions:** Preimplantation genetic diagnosis of nephrotic syndrome resulted in successful pregnancy.

456F

Development of a polycystic ovarian syndrome algorithm to identify cases in electronic health records. K. Atkins^{1,3}, D. Hucks^{2,3}, P. Straub^{2,3}, M. Wellons⁴, L. Davis^{2,3}. 1) Department of Microbiology, Immunology, and Physiology, Meharry Medical College, Nashville, TN; 2) Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN; 3) Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN; 4) Division of Diabetes, Endocrinology, and Metabolism, Vanderbilt University Medical Center, Nashville, TN.

Polycystic ovarian syndrome (PCOS) is one of the most common endocrine disorders among reproductive-aged women and one of the leading causes of infertility^{1,2}. It has been characterized by polycystic ovaries, excess androgens, excessive hair growth, and irregular menses. Although a small number of genetic loci contributing to PCOS have been identified, the exact molecular mechanisms underlying the disease are still unknown. Women who are diagnosed with PCOS often demonstrate features of metabolic disorders such as increased body mass index (BMI), insulin resistance, dysglycemia, and dyslipidemia which are also risk factors for cardiovascular diseases (CVD) including, but not limited to, type 2 diabetes, hypertension, and heart disease^{1,3}. Furthermore, PCOS is associated with greater CVD risk suggesting there could be similar underlying mechanisms regulating CVD and PCOS^{3,4}. Even though PCOS does not demonstrate higher prevalence among minority populations, the comorbid metabolic dysfunction is more often observed in African American and Hispanic populations. Moreover, evidence has shown that Hispanic women tend to have a more severe PCOS phenotype due to hyperandrogenism and dysmetabolism². To better understand PCOS, greater attention should be focused on the racial and ethnic differences in risk factors and comorbidities including metabolic syndrome. To achieve this goal, we developed an algorithm designed to identify PCOS cases across a diverse hospital population using the information in the de-identified electronic health records (EHRs) at Vanderbilt University Medical Center. Using a combination of structured and unstructured data, we developed two PCOS algorithms. The first algorithm relied solely on International Classification of Disease codes, Ninth and Tenth Revision, to identify a set of PCOS cases whereas the second algorithm included text mining for PCOS keywords. A positive predictive value (PPV) of 40% was obtained for the first dataset and a PPV of 78% was obtained for the second dataset. Inclusion of specific PCOS keywords increased the accuracy of the algorithm and prevented the inclusion of false positives in the dataset. The PCOS algorithm will be validated through an eMERGE collaboration using Northwestern EHRs.

457W

Prenatal characterization of osteogenesis imperfecta type V caused by *IFITM5*. T.K. Ha¹, D. Belefard¹, J.T. Shieh^{1,2}, S. Chetty², M.E. Norton³, *Both authors contributed equally to this work. 1) Department of Medical Genetics, University of California San Francisco, San Francisco, CA; 2) Institute for Human Genetics, University of California, San Francisco; 3) Department of Obstetrics, Gynecology, and Reproductive Sciences, University of California, San Francisco.

Osteogenesis imperfecta (OI) is heterogeneous connective tissue condition with a variable phenotype ranging from perinatal lethal to subclinical fractures. About 95% of cases are caused by autosomal dominant or *de novo* variants in *COL1A1/2*. Recently, an interferon induced transmembrane protein 5 (*IFITM5*)-related OI Type V (OMIM 610967) was characterized by distinctive findings. We describe, to our knowledge, the earliest prenatal diagnosis of OI Type V at 18 weeks gestation caused by a paternally inherited c.119C>T (p.40S>L) missense variant in *IFITM5*. The patient was a G4P0A4 woman with 2 prior spontaneous abortions. During her third and fourth pregnancies at 22 and 18 weeks, respectively, ultrasound showed bilateral bowed femurs, fractured tibias, and decreased skull ossification. Autopsies after termination revealed blue sclerae, generalized osteopenia, and fractures thought to be consistent with perinatal lethal osteogenesis imperfecta Type II. The father had an unremarkable family history and normal stature; however, imaging noted subclinical bone lesions that were nondiagnostic on biopsy. High resolution chromosomal analysis, chromosomal microarray, captured gene panel sequencing of known OI genes, and deletion/duplication analysis of *COL1A1/2* on genomic DNA extracted from cultured amniocytes were nondiagnostic. Sequence analysis of *IFITM5* on fetal amniocytes and paternal blood showed a pathogenic missense variant c.119C>T (p.40S>L). Maternal testing was negative. We report the earliest prenatal diagnosis of *IFITM5*-related OI Type V at 18 weeks gestation caused by a paternally inherited missense variant. While Glorieux *et al.* 2000 first described OI Type V in patients with fractures, distinctive radiographic findings, and mesh-like pattern in lamellae, the full phenotypic spectrum is ongoing. The paternal subclinical lesions and severe prenatal findings in two fetuses suggest markedly variable expressivity. As the phenotype of OI Type V continues to be delineated, we recommend prenatal sequence analysis for *IFITM5* in *COL1A1/2*-negative cases where radiographic evidence is highly suggestive of OI. Our findings expand the characterization of OI type V and the recurrence in this case illustrates the need for suspicion for this condition and consideration of prenatal diagnosis.

458T

Clinical exome sequencing identifies a *COL1A1* splice altering mutation in a prenatally diagnosed skeletal dysplasia patient. M. Mulatinho¹, H. Lee¹, D. Krakow², UCLA Clinical Genomics Center. 1) Department of Pathology and Laboratory Medicine, University of California, Los Angeles, Los Angeles, CA; 2) Department of Obstetrics and Gynecology, University of California, Los Angeles, Los Angeles, CA.

Prenatal imaging scans performed by ultrasound (US) have been successfully used to identify fetal anomalies; depending on the findings, parents are offered invasive procedures followed by genetic tests, in order to identify their clinical significance. Recently, clinical exome sequencing (CES) is being introduced as a prenatal diagnostic tool because it can provide results with higher resolution down to the single base-pair level, compared to traditional methods (karyotype, FISH and CMA), with its turnaround time improving to be reasonable for prenatal cases. Nevertheless, CES data interpretation and clinical relevance in the prenatal setting remain complex due to limited phenotype. Here we describe a 24 weeks old female fetus with possible skeletal dysplasia (SD) due to bowing of the bilateral femurs identified by US. These findings were suggestive of a moderate form of osteogenesis imperfecta (OI) and were discussed with both parents that declined invasive diagnostic test or termination of pregnancy. The infant was born by spontaneous vaginal delivery at 39 weeks gestation [7 lb 1.6 oz (3.22 kg)], and after birth a complete skeletal survey was obtained. The radiographs revealed poor mineralization of the calvarium, appendicular skeleton - thin cortices, mild bowing/upper extremities, and more severe bowing in the lower extremities. Cortical irregularity of the left proximal humerus was also noted. Widened squamosal and coronal sutures were noted in the skull. The SD clinical features present in the newborn is more severe than seen by US. CES performed on the patient's cord blood and the parents blood samples revealed a novel *de novo* heterozygous c.2451+1G>C splice site variant in the *COL1A1* gene. Although this exact variant has not been reported in other patient, c.2451+1G>A and c.2451+1G>T have been reported to cause OI type I, one of the most common clinical phenotypes associated with *COL1A1*. Although our patient has more severe skeletal findings than OI type I, it is not uncommon to see a spectrum of phenotypes with the same mutation. Here, we show how we successfully interpreted the US prenatal findings, and their genotype-phenotype correlation was confirmed postnatally by CES. In addition, this study shows that even though CES is available for prenatal diagnostic, not all families are willing to make decisions during pregnancy but how the US findings still played an important role on the immediate clinical management postnatally.

459F

Non-invasive prenatal sequencing for pregnancies with fetal skeletal dysplasia using cell-free fetal DNA. Y. Jiang¹, G. Wang¹, L. Lu¹, L. Wang¹, J. Xiao¹, J. Yang¹, Y. Feng¹, D. Chen¹, S. Mullegama^{1,2}, X. Wang^{1,2}, Z. Chen^{1,2}, Y. Wang^{1,2}, I. Veyver^{2,3}, C. Eng^{1,2}, J. Zhang^{1,2}. 1) Baylor Genetics Laboratories, Houston, TX; 2) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 3) Obstetrics and Gynecology, Baylor College of Medicine, Houston, TX.

We developed a cell-free fetal DNA based non-invasive prenatal sequencing test for multiple Mendelian monogenic disorders that includes 30 genes associated with common dominant disorders. A total of 638 samples were tested between January 2017 and June 2018 and 40 of them had positive molecular findings. Pathogenic or likely pathogenic variants were detected in 28 cases affected with achondroplasia (n=3), thanatophoric dysplasia (n=12), osteogenesis imperfecta (n=9), Cornelia de Lange syndrome (n=2) and craniosynostosis syndromes (n=2). For these 28 cases, the mean maternal age was 32.4 years (SD=6.3, range: 21-44) and the mean paternal age was 36.0 years (SD=6.0, range 22-48). The average gestational age at diagnosis was 23.7 weeks (SD=5.8, range: 13.0-36.6). 218 fetuses (35.4%) had abnormal ultrasound findings and 85 (13.3%) had at least one of the following skeletal dysplasia-related features on fetal ultrasound screening: shortened long bones, bowed and/or fractured bones, small thorax and craniosynostosis. While the overall positive detection rate was 6.3% (40/638), it was 16.5% (36/218) for fetuses with any abnormal ultrasound finding and increased to 25.9% (22/85) with ultrasound findings suggesting fetal skeletal dysplasia. Our test had the highest positive detection rate (66.7%, 10/15) in fetuses suspected with a single etiology based on the sonographic findings consistent with achondroplasia, thanatophoric dysplasia or osteogenesis imperfecta. Six fetuses harbored pathogenic variants in *FGFR3* and four in *COL1A1*. In comparison, the positive detection rate for *FGFR3* mutations was 55.3% in a targeted non-invasive NGS study for pregnancies affected with achondroplasia or thanatophoric dysplasia. In another recent study of non-invasive prenatal testing for achondroplasia using high-resolution melting PCR, the positive detection rate was 21.9%. Additionally, in four cases pathogenic variants inherited from an affected father were detected and clarified that these fetuses inherited the paternal variant at an earlier gestational age than would be possible with other approaches. In conclusion, we found that our non-invasive test adds valuable molecular information for suspected diagnoses based on sonographic findings that can help guide physicians for evaluation of fetal skeletal abnormalities. This test also has potential as a screening test for pregnancies at risk of such conditions.

460W

Lethal skeletal dysplasias (SD): A clinical-epidemiological study based on 131 Brazilian fetuses including molecular investigation. D.P. Cavalcanti¹, T.Y. Kanazawa¹, C.A. Moreno¹, K.C. Silveira¹, G.V.A. França². 1) Medical Genetics, University of Campinas, Campinas, Campinas, Brazil; 2) Secretariat of Health Surveillance, Brazilian Ministry of Health, Brasília, Brazil.

Epidemiological information regarding lethal-SD is very limited and outdated. The aim of this work was to conduct a descriptive clinical-epidemiological study based on a Brazilian series of 131 fetuses registered over 26 years (1990–2016) by the Campinas Skeletal Dysplasia Group. While 75 patients were born in our Hospital, 31 were evaluated during prenatal, but were born in neighborhood cities, and the remaining (25) were evaluated after birth in our site (www.ocd.med.br). For the analysis of risk factors, a healthy control sample of 41,186 live births born in the same period was used. Considering the 75 fetuses born among 79,745 births from the local hospital, the birth prevalence was 9.4 per 10,000 births (95% CI = 7.4–11.8). Although most cases (61.1%) had been detected during prenatal US examination, the specific diagnosis required a postnatal evaluation. Most fetuses born alive (75; 57%), however, 60% of them died in the 1st day. Among the stillbirths (56; 43%), 29% were due to pregnancy termination. Parental ages (father 33.1±7.2 and mother 28.0±5.7), and parental consanguinity (12.8%) were higher in cases than in controls ($p < 0.001$). Likewise, cases presented lower birth weight, and lower gestational age than controls. The most common groups (71%) were the FGFR3 group (43; 32.8%), mainly represented by DT-I (36; 83.7%); the OI group (33; 25.2%) with OI-IIA as the most common (22; 66.7%); and the collagenopathies type 2 group (17; 13%), being the achondrogenesis type II (ACG-II) the most frequent condition (10; 58.8%). Molecular investigation was performed in most of the cases (75; 57.3%) and confirmed the radiological diagnosis in 63 of them (84%). Among the negative cases, two fetuses presented an identical and novel clinical-radiological pattern, however, the WES failed to show an associated pathogenic variant. In conclusion, the epidemiological evaluation of the present series of fetuses shows a high prevalence of lethal-SD in our hospital due to its reference role in the region. The main SD found were DT-I (FGFR3 group), OI-IIA (OI group) and ACG-II (type 2 collagenopathies group). Although most cases of lethal-SD were suspected during pregnancy, a specific diagnosis was only reached after postnatal evaluation. The main risk factors were high paternal age and high parental consanguinity. Finally, a novel lethal-SD was found in this series. Financial Support: CNPq #590148/2011-7 / Faepex #402/15 / FAPESP # 98/16006-6; 2015/22145-6.

461T

Clubfoot as a prenatal clue to the diagnosis of the 22q11.2 deletion syndrome (22q11.2DS). J.F. Homans¹, T.B. Crowley², E. Chen², D.E. McGinn³, V.F.X. Deeney³, R.M. Castelein¹, R.S. Davidson³, D.M. McDonald-McGinn². 1) Department of Orthopedic Surgery, University medical Center, Utrecht, Utrecht, Netherlands; 2) Division of Human Genetics, The 22q and You Center, Children's Hospital of Philadelphia, Philadelphia, PA, USA; 3) Department of Orthopaedic Surgery, The University of Pennsylvania School of Medicine, The Children's Hospital of Philadelphia, Philadelphia, PA, USA.

Background: 22q11.2DS is the most common microdeletion syndrome occurring in ~1:2000 - 4000 live births and ~1:1000 fetuses. The presentation is complicated by significant intra and inter familial variability across the lifespan. Patients diagnosed neonatally generally present with severe features such as Tetralogy of Fallot +/- pulmonary atresia. However, patients without severe classically associated congenital malformations can either be diagnosed later in life, sending their families on a diagnostic odyssey, or not diagnosed at all, despite significant structural anomalies such as a solitary kidney. Thus, features less typically associated with 22q11.2DS may be overlooked or may not lead to appropriate diagnostic testing. Orthopedic manifestations are a prime example. Clubfoot has a general population prevalence of ~1:1000 and has been occasionally described in association with 22q11.2DS. However, this has not been systematically investigated. We hypothesize the prevalence of clubfoot is higher in patients with 22q11.2DS and suggest diagnosis of clubfoot prenatally should evoke suspicion of 22q11.2DS. **Methods:** We performed a retrospective review under IRB approval on the world's largest database of patients with 22q11.2DS to determine the prevalence of clubfoot. Clubfoot is a clinical diagnosis and can be mistaken for other malformations. "True clubfoot" requires treatment (either conservative or surgical), therefore we only included those patients with proof of treatment. **Results:** Records were reviewed on 1339 patients. Of these, 43 (3.2%) had confirmation of clubfoot (95% Confidence Interval: 2.4–4.3): 20 (46%) had bilateral clubfoot, nine (21%) left clubfoot, and 14 (33%) right clubfoot. Twelve patients were noted to have clubfoot on parental ultrasound, including one child whose father had a history of repaired VSD and cleft palate, but the diagnosis of 22q11.2DS was not considered in the differential. The latter child was discharged to home from the outside hospital and transferred emergently in cardiac extremis due to a previously unrecognized diagnosis of an interrupted aortic arch type B. **Conclusion:** We report the prevalence of club foot in 22q11.2DS to be 30 times higher than that observed in the general population. This suggests the diagnosis of clubfoot, especially in the face of other typically associated abnormalities, should provoke consideration of 22q11.2DS as an underlying diagnosis in particular in the prenatal setting.

462F

Combining familial genetic diagnosis with prenatal genetic testing for Mendelian disorders reduces recurrence risk. Z. Schlachetzki¹, M.Y. Issa², V. Stanley¹, R.D. George¹, M. Zaki², J.G. Gleeson¹. 1) Departments of Neurosciences and Pediatrics, Rady Children's Institute for Genomic Medicine, Howard Hughes Medical Institute, University of California San Diego, La Jolla, California 92093; 2) Clinical Genetics Department, Human Genetics and Genome Research Division, National Research Centre, Cairo 12311, Egypt.

Congenital malformations and birth defects are a leading cause of infant mortality. About half of them have a genetic cause and are therefore potentially preventable. Prenatal genetic screening is the current method of choice for detecting carriers of known pathogenic variants. However, the variants tested generally include only those with population allele frequencies greater than 1:100, and omit more rarely observed variants. Rare disease-causing variants can be identified with high certainty in about 50% of families with recessive Mendelian disorders that undergo whole-exome sequencing (WES). This genetic diagnosis can then be used to genotype subsequent pregnancies and to counsel families based on the results of the prenatal testing. However, there are no clear guidelines or large-scale studies of such practices, and each prenatal center follows its own best practice. We designed a retrospective study to determine the effect of prenatal genetic testing on disease recurrence risk in families with an established genetic diagnosis. Of 1172 families with severe recessive disease that were counseled between 2009 and 2015, 526 families received a conclusive genetic diagnosis following WES. This prompted fetal genetic testing through amniocentesis for 84 subsequent pregnancies. Of 59 fetuses that genotyped negative for the pathogenic variant and were carried to term, all were phenotypically normal for the tested disease at clinical follow-ups, which resulted in a 100% negative predictive value for our test. Of 24 fetuses that genotyped positive for the pathogenic variant, 8 were carried to term and developed a disease concordant with their affected siblings. The remaining 16 pregnancies were terminated and this reduced the disease recurrence from expected 25% to the observed 12% (95% Confidence Interval [CI], 0.04 to 0.20; $p=0.011$). This is the first comprehensive report that 1) demonstrates how next generation sequencing and prenatal testing can be combined to significantly reduce disease burden and positively impact family planning, and 2) offers an accessible and effective workflow for medical professionals to provide genetic counseling to families with severe and rare recessive disorders, thus ensuring informed reproductive decisions.

463W

SCN1A mutations in a SIDS cohort. C.A. Brownstein, I.A. Holm, R.D. Goldstein, A.H. Poduri. Boston Children's Hospital, Boston, MA.

Sudden infant death syndrome (SIDS), the death of an infant <1 year of age that remains unexplained after complete autopsy and death scene investigation, is the leading cause of postneonatal infant mortality in the United States. We identified *SCN1A* variants in 2 infants who died of SIDS with hippocampal abnormalities from an exome sequencing study of 10 cases of SIDS but no history of seizures. One harbored *SCN1A* G682V, and the other had 2 *SCN1A* variants in cis: L1296M and E1308D, a variant previously associated with epilepsy. Functional evaluation in a heterologous expression system demonstrated partial loss of function for both G682V and the compound variant L1296M/E1308D. This novel association of *SCN1A* with SIDS supports further intense efforts to understand epilepsy-related mechanisms into sudden death across the age spectrum in individuals with and without an overt history of seizures or epilepsy.

464T

The genetic landscape of sperm mosaicism: Implications for the recurrence risk of neutral and autism-causing mutations. M.W. Breuss^{1,2}, D. Antak^{3,4,5}, R.D. George^{1,2}, M. Kleiber^{3,4,5}, K.N. James^{1,2}, L.L. Ball^{1,2}, O. Hong^{3,4,5}, I. Mitra^{6,7}, S.A. Wirth^{1,2}, J. Gu^{1,2}, C.A.B. Garcia^{1,2}, W.M. Brandler^{3,4,5}, D. Musaev^{1,2}, A. Nguyen^{1,2}, J. McEvoy-Venneri^{1,2}, R. Knox^{1,2}, E. Sticca^{1,2}, O. Devinsky⁶, M. Gymrek^{6,7}, J. Sebat^{3,4,5}, J.G. Gleeson^{1,2,3}. 1) Department of Neurosciences, Howard Hughes Medical Institute, University of California, San Diego, La Jolla, CA; 2) Rady Children's Institute for Genomic Medicine, San Diego, CA; 3) Beyster Center for Genomics of Psychiatric Diseases, University of California, San Diego, La Jolla, CA; 4) Department of Psychiatry, University of California, San Diego, La Jolla, CA; 5) Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, CA; 6) Department of Medicine, University of California San Diego, La Jolla, CA; 7) Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA; 8) Department of Neurology, Epilepsy Division, New York University School of Medicine, New York, NY.

De novo genetic mutations represent a major contributor to pediatric disease, including autism spectrum disorders (ASD), congenital heart disease, and muscular dystrophies, but there are currently no methods to prevent or predict them. These mutations are classically thought to occur either at low levels in progenitor cells or at the time of fertilization and are often assigned a low risk of recurrence in siblings. Here, we directly assessed the presence of *de novo* mutations in paternal sperm and discovered abundant, germline-restricted mosaicism. In 8 families with 14 children we evaluated the allelic fractions of 912 *de novo* single nucleotide variants (dSNVs) in paternal blood and sperm, employing deep (200x) whole genome sequencing (WGS). 23 of these variants showed evidence of mosaicism of up to 17% allelic fraction; and almost two thirds were either exclusively found or highly enriched (>3 times) in sperm. Unbiased analysis using MuTect 2 and Strelka 2 uncovered >50 additional, mosaic variants at a similar range of allelic fractions. We discovered germline-restricted mosaicism for structural variants (SVs) at similar rates and allelic fractions, demonstrating that germline mosaicism is present across different classes of *de novo* mutations (DNMs). Employing single-molecule genotyping, we found that this phenomenon was also observed in likely-pathogenic mutations in ASD, where 4/20 variants showed similar germline mosaicism and, consequently, an increased risk of recurrent transmission. This included one family, for which the mosaicism of a *GRIN2A* variant reached close to 15% in paternal sperm, but only around 1% in saliva. The disease allele was recurrently inherited by all three children, a fact which was missed before due to the previously reported, variable expressivity of this mutation. Our results demonstrate that germline mosaicism is an underestimated phenomenon with important implications for clinical practice and our understanding of the basis of human disease. Genetic analysis of sperm, rather than of blood or saliva, can accurately assess the individualized recurrence risk following the birth of a child with a *de novo* disease, as well as the risk for any male planning to have children.

465F

Rapid detection of Trisomy 13, 18, 21 and sex chromosomes aneuploidies – experience of a clinical hospital laboratory. K. Tan, K.S. Poon, L. Chiu, S.S.Y. Ho, E.S.C. Koay, M. Choolani. National University Hospital Singapore.

Background: Aneuploidies of chromosomes 13, 18, 21, X and Y account for more than 80% of the clinically significant chromosomal abnormalities diagnosed prenatally. Rapid tests for prenatal diagnosis of these chromosomal aneuploidies can improve pregnancy management. We describe our experience with prenatal testing of chromosomal aneuploidies in a clinical genetics laboratory of a tertiary hospital in Singapore. Methods: Rapid detection of trisomy 13, 18, 21 and sex chromosomes aneuploidies was performed using an in-house developed Quantitative Fluorescent Polymerase Chain Reaction (QF-PCR) assay (Baig et al, Singapore Med J 2010; 51(4): 343-348). This assay contains 19 microsatellite markers located on chromosomes 13, 18, 21, X and Y. The turnaround time is 2 working days. Results: In our molecular diagnostics laboratory, 1226 QF-PCR assays were performed on 935 amniocentesis samples and 291 chorionic villus samples since June 2007. Majority of samples were obtained at gestational week 16+. Out of the 1226 cases tested, 124 cases of Trisomy 21 (10.1%), 42 cases of Trisomy 18 (3.4%), 6 cases of Trisomy 13 (0.5%), 10 cases of Monosomy X (Turner's syndrome) (0.8%) and 4 cases of XXY (Klinefelter's syndrome) (0.3%) were identified. Trisomy 21, 18 and 13 cases were 100% concordant between karyotyping and QF-PCR (n = 150). Since the availability of non-invasive prenatal screening (NIPS) testing around 2014, the number of test requests have decreased by approximately 50%. Co-testing of NIPS was performed in 39 cases. Out of these, 21 were reported as high risk for Trisomy 21 and 18 of these were confirmed Trisomy 21 by QF-PCR (86%). Nine were reported as high risk for Trisomy 18 and 7 of these were confirmed Trisomy 21 by QF-PCR (78%). One was reported by NIPS to be high risk for Trisomy 13, 18, and 21 but was not confirmed by QF-PCR. Conclusion: Rapid detection of trisomies 13, 18, and 21 by QF-PCR on amniocentesis and chorionic villus samples is a useful test in the management of high risk pregnancies. More importantly it serves as a confirmatory test to ascertain the trisomy cases predicted by NIPS, in which a relatively high rate of false-positives had been encountered.

466W

A component of sperm fibrous sheath has major effect on testis morphology and functionality. *N. Li.* Guangzhou Women and Children's Medical Center, Guangzhou, Guangdong, China.

The fibrous sheath is a unique cytoskeletal structure located in the principle piece of the sperm flagellum with more than 13 protein components. AKAP4 is the most abundant protein in the fibrous sheath, which interacts with at least 3 other proteins. The molecular structure and functionality of the fibrous sheath are largely unknown. We collected a clinic sample of sperms featured by dysplasia of fibrous sheath (DFS), leading to a failure of natural conception. The sperm donor is an offspring of consanguineous family, and we identified an inherited homozygous truncating mutation in his genome by whole-exome sequencing. The affected protein is one of the components of fibrous sheath, and this mutation caused a shortened protein which lost part of the original functions. The mice model with this mutation introduced by CRISPR-Cas9 technique showed similar phenotype to the human, with sperms of reduced number and lower motility due to flagellum malfunction. Strikingly, we observed that, unlike AKAP4 knock-out mice model, the knock-out mice model we constructed for the novel gene exerted major effect on testis, manifested by significant size/weight reduction and azoospermia. Microscopic observation of testis slices and in-situ hybridization showed abnormal cross-section of the seminal vesicles and disorganized progression from spermatogonia to spermatid, when comparing the knock-out mice model with the control ones. Therefore, we concluded that this gene is of critical importance to normal spermatogenesis and testis development.

467T

De novo mutations in caudal type homeo box transcription factor 2 (CDX2) in patients with persistent cloaca. *J.S.J Hsu^{1,4}, M.T. So², C.S.M. Tang^{2,4}, A. Karim², R. Porsch^{1,4}, C. Wong², M. Yu², F. Yeung², H. Xia³, R.Z. Zhang³, S.S. Cherny^{1,4}, P.H.Y. Chung², K.Y. Wong², P.C. Sham^{1,4}, N.D. Ngo⁵, P.K.H. Tam^{2,4}, V.C.H. Liu⁴, M.M. Garcia-Barcelo^{2,4}.* 1) Department of Psychiatry, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong; 2) Department of Surgery, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong; 3) Department of Pediatric Surgery, Guangzhou Women and Children's Medical Center, Guangzhou, Guangdong, China; 4) Centre for Genomic Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China; 5) Department of Human Genetics, National Hospital of Pediatrics, Hà Nội, Vietnam.

The cloaca is an embryonic cavity that is divided into the urogenital sinus and rectum upon differentiation of the cloacal epithelium triggered by tissue-specific transcription factors including Caudal Type Homeo Box Transcription Factor 2 (*CDX2*). Defective septation anomalies during the development lead to persistent cloaca in humans (PC), a phenotype recapitulated in *Cdx2* mutant mice. PC is linked to hypo/hypervitaminosis A. Although no gene has ever been identified, there is a strong evidence for a genetic contribution to PC. We applied whole-exome sequencing and copy-number-variants analyses to 21 PC patients and their unaffected parents. The damaging p.Cys132* and p.Arg237His *de novo CDX2* variants were identified in two unrelated patients. Both variants were novel among hundred thousand whole exome sequencing sample in the gnomAD database ($N > 123,136$). Not only the predictive model from Phyre2 Investigator indicated the p.Arg237His is located on DNA binding domain with higher mutation sensitivity compared with adjacent regions, but also SWISS-MODEL indicated a secondary structural change in the DNA binding domain. Moreover, these variants functionally altered the expression of *CYP26A1*, a direct *CDX2* target encoding the major retinoic acid (RA)-degrading enzyme. Other RA genes, including the RA-receptor alpha, were also mutated. Genes governing the development of cloaca-derived structures were recurrently mutated and over-represented in the basement-membrane components set ($q\text{-value} < 1.65 \times 10^{-6}$). Joint analysis of the patients' profile highlighted the extracellular matrix–receptor interaction pathway (MsigDBID: M7098, FDR: $q\text{-value} < 7.16 \times 10^{-9}$). This is the first evidence that PC is genetic, with genes involved in the RA metabolism at the lead. Given the *CDX2 de novo* variants and the role of RA, our observations could potentiate preventive measures. For the first time, a gene recapitulating PC in mouse models is found mutated in humans. Despite the severity, the rarity of PC, establishing disease causality for any given gene is extremely difficult. As patients with the identical rare disease are limited, data sharing and re-analysis should be fully considered and should be conducted under proper regulations.

468F

Oculocerebrocutaneous (Delleman) syndrome. Prenatal manifestations and failure to identify tumors' somatic mutation using whole exome sequencing. N. Damseh^{1,2}, D.J. Stavropoulos³, K. Chong⁴, S. Blaser⁵, I. Miron⁴, S. Carmona⁴, D. Chitayat^{1,4}. 1) Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, University of Toronto, Toronto, ON, Canada; 2) Medical Genetics Residency Training Program, University of Toronto, Toronto, ON, Canada; 3) Department of Paediatric Laboratory Medicine, Division of Genome Diagnostics, The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada; 4) The Prenatal Diagnosis and Medical Genetics Program, Mount Sinai Hospital, University of Toronto, Toronto, ON, Canada; 5) Department of Pediatrics, Division of Radioimaging, The Hospital for Sick Children, University of Toronto, Toronto, ON, Canada.

Oculocerebrocutaneous syndrome (Delleman syndrome) is characterized by orbital cysts, microphthalmia/anophthalmia, focal skin defects, facial skin appendages and a variety of cerebral malformations. All cases so far have been sporadic and the condition is thought to be due to somatic mutation of an autosomal dominant lethal gene important in early development. We report the prenatal and postnatal findings in a female newborn with Delleman syndrome. Our patient was born to a 29-year-old primigravida mother who presented at 26 weeks gestation with fetal ultrasound finding of a solid mass on the left cheek inferior to the left orbit. Follow up ultrasounds showed an abnormally shaped cerebellum and two fetal facial tags with no blood flow. A fetal MRI confirmed a left cheek subcutaneous, finger-like tissue lesion, hypoplastic cerebellar vermis and dysplastic cerebellar hemispheres. Delivery of the infant at term confirmed the prenatal findings and the diagnosis of Delleman syndrome. In an attempt to find the causative gene mutation we performed WES on DNA extracted from the tumor but no gene mutations was identified. Delleman syndrome is a rare disorder characterized by eye, cutaneous, and central nervous system anomalies. Typically, neonates present with microphthalmia or orbital cysts, orbital skin tags, dermal aplasia, or hypoplasia. Brain abnormalities including such as agenesis of the corpus callosum, hydrocephalus, enlarged ventricles, and cerebellar hypoplasia have been reported. In most cases the periorbital skin tags are unilateral and as in our patient, most frequently left-sided. Some of the patients experience seizures and intellectual disability. This is the second reported patient with Delleman syndrome diagnosed prenatally based on the finding of skin appendages in the periorbital area and cerebellar abnormalities. To our best knowledge this is the only case where whole exome sequencing was done on DNA extracted from the facial tumor and unfortunately it failed to identify the genetic basis of this condition.

469W

Prenatal detection and clinical outcomes of trisomy 16 using massive parallel sequencing of total cell-free DNA (cfDNA). A. McFaddin¹, A. Jacobson¹, C. Lockwood¹, B. Shirts¹, B. Colbert¹, E. Oehler², J. Rietzler², S. Paolucci², S. Ball², C. Albright², J. Hitt². 1) Laboratory Medicine, University of Washington Medical Center, Seattle, WA; 2) Maternal Fetal Medicine, University of Washington Medical Center, Seattle, WA.

Trisomy for chromosome 16 (T16) is the most common autosomal trisomy detected in spontaneous abortions, and is predicted to be present in as many as 1.5% of all conceptions. The majority of pregnancies with full T16 are spontaneously lost by week 12 of pregnancy, but cases of fetal mosaicism as well as confined placental mosaicism have been reported in later gestational ages as well as live births. Historically, maternal serum screening and ultrasound were the only available options for noninvasive aneuploidy screening; however, these techniques are not optimized for rare trisomies such as T16. Non-invasive screening using cell-free DNA (cfDNA) from maternal blood has greatly improved the reliable detection of autosomal aneuploidies in recent years. Many clinically available cfDNA screening tests only report findings for the more common trisomies of 13, 18, and 21 and have not reported rare autosomal aneuploidies, such as T16. If present, these rare aneuploidies may be of clinical importance and have impact on pregnancy outcomes and decisions. The University of Washington Genetics laboratory performs a cfDNA prenatal screen that assesses aneuploidy for all chromosomes using massively parallel sequencing of total cfDNA from maternal plasma. Since clinical implementation of this screen in 2017, 1110 cfDNA samples have been tested. Of these, 4 cases of T16 have been detected and reported, with varying clinical outcomes. Case A was detected in the setting of abnormal ultrasound findings that ended in an elective pregnancy termination. Case B was detected in the setting of a mother of advanced maternal age (AMA) and normal first trimester ultrasound, with delivery of growth restricted infant at 27 weeks with negative peripheral blood testing and confined placental mosaicism for T16 confirmed postnatally. Case C was detected in the setting of a mother of AMA and normal first trimester ultrasound with subsequent fetal demise at 16 weeks and confirmed mosaicism for T16 in fetal tissue post-demise. Case D was detected in the setting of a positive serum screen indicating a 1/20 risk for trisomy 21 with abnormal ultrasound findings of a ventricular septal defect and shortened femur bones. The family declined confirmatory testing at time of submission and pregnancy is currently continuing. This data highlights the variability in clinical presentation for T16 detected by cfDNA and the potential utility of reporting this information in the prenatal setting.

470T

The yield of chromosomal microarray analysis among 5750 fetuses with various sonographic anomalies. S. Ben-Shachar^{1,2}, I. Maya³, A. Reches^{1,2}, A. Frumkin⁴, J. Grinshpun-Cohen⁵, R. Segele⁶, E. Manor⁷, M. Khayat⁸, T. Tenne^{9,2}, E. Banne¹⁰, A. Shalata¹¹, H. Yonath^{12,2}, R. Berger¹³, T. Naiman¹, A. Singer¹⁴, L. Sagi-Dain¹⁵. 1) Genetics Institute, Tel Aviv Sourasky Medical Center,; 2) Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel; 3) Recanati Genetics Institute, Beilinson Hospital, Rabin Medical Center, Petach Tikva, Israel; 4) Department of Genetics and Metabolic Diseases, Hadassah, Hebrew University Medical Center, Jerusalem, Israel; 5) Genetic Institute, Assaf Harofeh Medical Center, Zerifin, Israel; 6) Medical Genetics Institute, Shaare Zedek Medical Center and the Hebrew University School of Medicine, Jerusalem, Israel; 7) Genetic Institute, Soroka University Medical Center, Faculty of Health Sciences Ben-Gurion University of the Negev, Israel; 8) Institute of Human Genetics, Haemek Medical Center, Afula, Israel; 9) Medical Genetics Institute, Meir Medical Center, Israel; 10) Genetics Institute, Kaplan Medical Center, Rehovot, Israel, affiliated to the Hebrew University and Hadassah Medical School, Jerusalem, Israel; 11) Simon Winter Institute for Human Genetics, Bnai Zion Medical Center, Haifa, Israel; 12) Gertner Institute of Human Genetics, Sheba Medical Center, Israel; 13) Cytogenetic Maccabi Health Care, Israel; 14) Community Genetics, Public Health Services, Ministry of Health, Jerusalem, Israel; 15) Genetics Institute, Carmel Medical Center, Haifa, Israel.

OBJECTIVE: To examine the role of chromosomal microarray analysis (CMA) in pregnancies with various sonographic anomalies, and to characterize the copy number variants (CNVs) in diverse fetal phenotypes. **METHODS:** We retrospectively examined CMA analyses performed nationwide due to fetal sonographic anomalies (structural defects, intrauterine growth restriction and polyhydramnios) between 1/2013 and 09/2017. The rate of abnormal CMA findings was compared between the different phenotypes. **RESULTS:** Clinically significant CMA aberrations were detected in 272 of 5750 cases (4.7%): 115 (2%) karyotype-detectable and 157 (2.7%) submicroscopic. Unlike the submicroscopic findings, the prevalence of karyotype-detectable findings varied notably between different phenotypes (0.2%-9.7% vs. 2.2-3.7%, respectively). Most commonly detected CNVs were 22q11.21 deletions (0.4%), followed by 22q11.21 gain of copy number (0.2%). Specific CNVs detected among pregnancies with abnormal sonographic findings were up to 20-fold more prevalent compared to low-risk pregnancies. Some CNVs were associated with specific phenotypes (e.g., 22q11.21 and 17q12 microdeletions with cardiovascular and genitourinary defects, respectively). **CONCLUSION:** Our data demonstrate the importance of performing CMA in pregnancies with non-major anomalies and non-structural sonographic abnormal findings. Thus, CMA in pregnancies with sonographic aberrations should not be limited only to major anomalies.

471F

Whole exome sequencing improves the genetic diagnosis in fetuses with increased nuchal translucency. H. Wang^{1,2}, Z. Dong^{3,4}, H. Zhang¹, R. Zhang², H. Lan^{1,2}, X. Yang², J. Lei¹, X. Xu¹, L. Xiong^{1,2}, Y. Zhu², K. Choy^{3,4}. 1) Department of Central Laboratory, Bao'an Maternity and Child Healthcare Hospital, Jinan University, Shenzhen, China; 2) Department of prenatal diagnosis center, Bao'an Maternity and Child Healthcare Hospital, Jinan University, Shenzhen, China; 3) Department of Obstetrics and Gynaecology, The Chinese University of Hong Kong, Hong Kong; 4) Shenzhen Research Institute, The Chinese University of Hong Kong, Shenzhen, China.

Background: High nuchal translucency (NT >3.5mm) detected in the first trimester is a well-known and widely accepted indicator for aneuploidy or genomic disorders in prenatal care. Our pilot study reported ~8.1% fetus with high NT carries pathogenic copy number variants (CNVs), there are still more than 90% of cases without a genetic diagnosis. **Objectives:** We aim to investigate the detection yield of combination of whole-exome sequencing (WES) and CMA in fetuses with high NT. **Methods:** The study comprised 247 singleton pregnancies with increased fetal NT (NT>3.5mm) and DNAs from amniotic fluid were used for genetic testing. Parental DNAs were used for further confirmation of inheritance. Karyotyping and CMA were carried out for each case and classification of CNVs follow the guidelines of American College of Medical Genetics and Genomics (ACMG). Ten cases with negative findings from karyotyping and CMA testing were further subjected for whole-exome sequencing with their parental DNAs. SNV interpretation also followed ACMG guidelines. **Results:** The maternal age was 32.1 on average and the median fetal NT thickness was 4.3 mm. Karyotyping/CMA identified 33 cases (13.3%) with chromosomal abnormalities, including three cases of balanced chromosomal abnormalities, 15 cases with aneuploidies and 15 cases with pathogenic microdeletion or microduplication. In addition, variants of unknown significance (VOUS) were also detected in six cases (2.4%, 6/247). Whole-exome sequencing was carried out in each of ten cases with negative results from karyotyping/CMA with their parental DNAs and identified two *de novo* pathogenic SNVs. For instance, a pathogenic mutation was detected in autosomal dominant (AD) gene of *PTPN11* known to cause Noonan syndrome, while another pathogenic mutation was identified in gene *BRAF* (AD), which is related to cardio-facial-cutaneous (CFC) syndrome. **Conclusions:** Whole-exome sequencing can have 20% (2/10) additional diagnostic yield in the fetuses with high NT without a genetic diagnosis from karyotyping and CMA. It indicates that combination of whole-exome sequencing with current standard assays (karyotyping and CMA) can help to increase diagnostic yield in fetuses with high NT. A larger prospective study is warranted for further investigation due to the limited sample size in current study. This study was supported by the National Natural Science Foundation of China (Grant no. 81741004).

472W

Prenatal diagnosis of Cri-du-Chat (5p-) syndrome with persistent isolated fetal ascites: A case report. L. Mota-Vieira^{1,2}, I. Raposo³, M. Ormonde³, R. Monteiro⁴, A. Sampaio³, P. Cosme³, M. Cardoso³. 1) Molecular Genetics and Pathology Unit, Hospital of Divino Espírito Santo of Ponta Delgada, Azores, Portugal; 2) BiolSI - Biosystems & Integrative Sciences Institute, Faculty of Sciences, University of Lisboa, Portugal; 3) Department of Obstetrics and Gynecology, Hospital of Divino Espírito Santo of Ponta Delgada, Azores, Portugal; 4) Synlab - Genética Médica e Diagnóstico Pré-Natal, Porto, Portugal.

Cri-du-chat syndrome (CdCS) and isolated fetal ascites are two very rare conditions that have never been reported together. The former is a genetic disorder (OMIM #123450) caused by haploinsufficiency of contiguous genes in the critical region of 5p15.2 to the entire short arm of chromosome 5 (5p-). The second is defined as fluid accumulation in the abdominal cavity without fluid accumulation in any other serosal cavities or subcutaneous tissues, being thus a distinct entity from hydrops fetalis. Here, we describe a unique case of a prenatally diagnosed CdCS with persistent isolated fetal ascites. A 43-year-old Caucasian woman (gravida 4, para 3) was referred due to advanced maternal age. There was no history of consanguinity, teratogen exposure, miscarriages, mental retardation, or other inherited conditions. At the first ultrasound scan (12⁺WG) the only remarkable finding was a moderate to severe fetal ascites. Maternal blood group was O Rh-positive and indirect Coombs test was negative. Serologies were negative for parvovirus B19, CMV, HBV, HCV, VZV, HSV and *Toxoplasma gondii*. Maternal rubella IgG levels were consistent with previous infection. Due to advanced age, the pregnant woman underwent NIPT, which results revealed a low risk for trisomies 21, 18 and 13 (fetal fraction, 10.7%). At 16⁺ WG, the follow-up scan showed that isolated fetal ascites was still present with no signs of generalized hydrops. Taking these findings, an amniocentesis was performed and showed an abnormal female karyotype with terminal deletion of the short arm of chromosome 5 [46,XX,-del(5)(p15.2)], spanning the critical region for the CdCS, a result confirmed by MLPA. After counseling, the couple opted for the pregnancy termination at 19⁺ WG. Postmortem examination of the female infant (209g) revealed absence of dysmorphic features, although there was evidence of bilateral temporal lobe atrophy, coarctation of the aorta and an extra-hepatic cyst. Placental assessment showed villous dysmaturity and calcifications. Since parents refused genetic testing, we were unable to confirm whether this deletion was *de novo*, an inherited unbalanced translocation, or a rare aberration. Consequently, there is a potential genetic (familial) risk for CdCS, particularly in the parents' offspring. To conclude, we hypothesize that isolated fetal ascites has to be considered an antenatal ultrasonographic marker for cri-du-chat syndrome, a finding that should be confirmed in further cases.

473T

A systematic evaluation of 450 fetuses: Phenotypes, genetic causes and impact on genetic counseling. S.S. Nayak, P. Radhakrishnan, A. Shukla, A. Moirangthem, K.M. Girisha. Department of Medical Genetics, Kasturba Medical College, Manipal Academy of Higher Education, Manipal, India.

Introduction: Congenital anomalies occur in about 1 in 40 pregnancies and are common causes of perinatal loss. Perinatal losses manifest with diverse phenotypes, demonstrate etiological heterogeneity and challenge accurate genetic counselling. Perinatal autopsy is considered gold standard for evaluation of these conditions. New developments in genomic testing and experience of the perinatal team contribute to continual improvement in the performance of perinatal autopsy as a diagnostic tool. In this study we report the phenotypes and genetic cause of fetal losses in a large cohort and how it affected genetic counseling. **Materials and Methods:** A detailed perinatal autopsy was performed in all perinatal losses after taking informed consent. Antenatal history, sonographic findings, biometric measurements, photography, radiographs, external and internal examinations and histopathologic findings were systematically recorded. Cytogenetic tests, candidate gene sequencing and exome sequencing were carried out in fetuses as feasible. **Results:** Four fifty fetuses (251 male, 191 female and 8 undetermined) were evaluated over nine years with gestational age ranging 10-39 weeks. Radiographs were obtained in 433 (96%), cytogenetic tests in 94, candidate gene sequencing 14 and exome sequencing was performed in 19 fetuses. A definitive diagnosis was established in 375/450 (83%) fetuses on perinatal autopsy. Central nervous system (78/450, 17%) was most frequently affected followed by cardiovascular and urogenital system. We noted skeletal dysplasia in 30/450 (6.6%) cases. Hydrops (29/450), limb body wall complex /amniotic band disruption (19/450), arthrogryposis/fetal akinesia deformation sequence (17/450), neural tube defect (16/450) and Meckel syndrome (11/450) were the common diagnoses. Genomic diagnoses were established in 43 fetuses (22 fetuses with a cytogenetic abnormality and 21 fetuses with a monogenic syndrome). Autopsy redefined risk of recurrence in 55% fetuses tested. The work helped us to delineate new phenotypes and new genetic causes. **Conclusion:** Perinatal autopsy continues to evolve with new genomic testing facilities improving diagnosis and counseling for affected families.

474F

Imputed transcriptome analysis and phenome-wide association studies link the NAD metabolic pathway genes and congenital malformations in a biobank population.

T.W. Miller-Fleming^{1,2}, D.R. Velez-Edwards^{2,3}, G. Chenevix-Trench⁴, N.J. Cox^{1,2}. 1) Department of Medicine, Division of Genetic Medicine, Vanderbilt University Medical Center, Nashville, TN; 2) Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN; 3) Department of Obstetrics and Gynecology, Vanderbilt University Medical Center, Nashville, TN; 4) QIMR Berghofer Medical Research Institute, Brisbane, Queensland, 4029, Australia.

Congenital malformations (CM) occur in 3% of births and are a leading cause of infant mortality. Toxic environmental exposures, nutritional deficiency, and maternal age are known risk factors; however, many defects arise due to undetected genetic variation. Identifying the genetic risk factors, and the individuals at greatest risk for CM, could provide insight into the prevention of these disorders. Recent work uncovered rare genetic variation within the NAD metabolic pathway in four families presenting with CM (Shi et al. 2017). This study proposed a novel therapeutic for the prevention of birth defects in these families, as niacin supplementation can increase NAD levels in mice and humans; however, it was unclear whether these results could be replicated in larger populations. De-identified electronic health records (EHR) linked to genomic information allow researchers to evaluate genotype-phenotype relationships in extensive sample sizes among populations with strong phenotypic diversity. We evaluated the relationship between NAD pathway genes and congenital phenotypes using Vanderbilt's EHR-linked biobank, BioVU. A phenome-wide association study (PheWAS) across 23,000 individuals of European descent revealed an association between variants in *HAAO* (3-hydroxyanthranilate 3,4-dioxygenase) and *KYNU* (kynureninase) with CM phenotypes: vesicoureteral reflux, anatomical abnormalities of kidneys, and unspecified congenital anomalies ($p < 3 \times 10^{-5}$). Imputed expression levels of these genes were calculated using PrediXcan, a computational tool that derives genetically-predicted transcript levels from a reference transcriptome dataset (GTEx). We observed a significant association between decreased predicted expression of *HAAO* with respiratory conditions of fetus ($p < 3 \times 10^{-4}$) and *KYNU* with congenital anomalies of kidneys and limbs ($p < 0.005$ and $p < 0.004$). Additionally, we calculated a CM phenotype risk score (PheRS) for 23,000 individuals and uncovered an association between the burden of *HAAO/KYNU* pathogenic variants and the congenital PheRS within the BioVU cohort ($p < 0.05$). Our findings support a connection between the NAD pathway genes and CM phenotypes in our biobank population. Future studies will take advantage of the diverse medical phenome in BioVU to determine the relationship between lab values across the genome and phenome, uncover endophenotypes associated with CM, and leverage the congenital PheRS to identify novel genetic underpinnings of birth defects.

475W

Genomic autopsy to identify causes of abnormal fetal and neonatal development.

P. Arts¹, A. Byrne¹, L. Pais², D. Lawrence³, M. Babic¹, J. Feng², A. O'Donnell², T. Hardy^{1,4}, S. King-Smith^{1,5}, C. Barnett⁶, H.S. Scott^{1,3,5,7,8}. 1) Genetics and Molecular Pathology, Centre for Cancer Biology, SA Pathology and the University of South Australia, Adelaide, Australia; 2) Broad Institute Center for Mendelian Genomics and Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA; 3) Australian Cancer Research Foundation Cancer Genomics Facility, Centre for Cancer Biology, SA Pathology, Adelaide, Australia; 4) Repromed, Dulwich, Adelaide, Australia; 5) Australian Genomics Health Alliance, Melbourne, Australia; 6) Paediatric and Reproductive Genetics Unit, South Australia Clinical Genetics Service, Women's and Children's Hospital, North Adelaide, Australia; 7) School of Medicine, University of Adelaide, Adelaide, Australia; 8) School of Pharmacy and Medical Science, University of South Australia, Adelaide, Australia.

Loss of a pregnancy or neonate has a major impact on families in Australia and overseas. Despite pathological examination and genetic tests as part of standard care, ~25% of cases are currently unexplained. Understanding the (genetic) cause of perinatal death is essential to counsel these families for future pregnancies. The Genomic Autopsy Study was started to identify molecular defects underlying fetal and newborn abnormalities that result in termination of pregnancy, death *in utero* or death in the newborn period. To date, exome (WES) and genome (WGS) sequencing have been performed on 51 trios (parent-fetus). In this cohort, we have identified (likely) pathogenic mutations in known disease genes, leading to fetal (*in utero*) or neonatal presentations of developmental disorders for 13 families. Candidate (novel) variants affecting known disease genes were identified for 6 cases. Strong candidate variants in potentially novel disease genes were identified for 12 cases. We have gained sufficient evidence to confirm causality of the genetic variants for 3 of these families, resulting in 3 novel disease genes. For the remaining 9 cases, functional studies to characterise the effect of the genetic variant are planned or ongoing. Taken together, we have identified disease causing variants for 16 (32%) cases and strong candidate variants for another 15 (29%). Recently we have become the national referral centre for Genomic Autopsy in Australia, expecting to receive an increasing number of samples from centres nation-wide. This rapidly growing cohort will provide a great opportunity to gain knowledge about the genetic factors involved in perinatal death.

476T

Avoiding unnecessary trade-offs: Clinical experience for a noninvasive prenatal screen with both low no-call rate and high accuracy. S.E. Hancock, R. BenShachar, C. Adusei, C. Haverty, D. Muzzey. Counsyl, Inc, South San Francisco, CA.

Noninvasive Prenatal Screening (NIPS) via cell-free DNA analysis (cfDNA) has been rapidly incorporated into prenatal care since it became clinically available in 2011. The Counsyl Prelude Prenatal Screen utilizes a whole genome sequencing (WGS) approach to NIPS that has been widely reported upon by various laboratories. Here we describe the clinical experience of offering Prelude in our laboratory. We retrospectively analyzed results from 58,403 patients who underwent NIPS during a nine month period to assess clinical performance for Trisomy 21, Trisomy 18, and Trisomy 13 in singleton pregnancies. The patient cohort included both average-risk (48%, n=28,089) and high-risk patients (52%, n=30,315). Median maternal age of the patient population was 34 years (interquartile range (IQR): 29-37 years). Median gestational age was 12.4 weeks (IQR: 11.3-13.7 weeks). Unlike several NIPS offerings unable to provide a result for samples with fetal fraction below a specified threshold, Prelude does not have such limitations. The screen yielded a result for 99.9% of patients; no result was issued for the remaining 0.1% of cases (n=59) that failed for technical reasons. Median turnaround time was three days (IQR: 2-4 days). 572 cases (1%) screened positive ("aneuploidy detected" or "aneuploidy suspected") for common aneuploidies (362 positive for Trisomy 21, 142 positive for Trisomy 18, and 68 positive for Trisomy 13). "Aneuploidy suspected" reports were issued to 6.5% (n=37) of patients with positive results (0.06% of the total patient cohort). We sought pregnancy outcomes for all positive results and for 10% of negative results. Voluntarily reported false negatives were included in the analysis. Informative outcomes were received for 221 (38.6%) positive results and 700 negative results (12% of those requested). Based on cytogenetically confirmed cases and adjusting cohort size for ascertainment bias, observed sensitivities for Trisomy 21, Trisomy 18, and Trisomy 13 were 99.1%, 97.3%, and 92.3%, respectively. Outcome collection revealed 18 false positives overall (five for Trisomy 21, four for Trisomy 18, and nine for Trisomy 13). Importantly, this level of clinical performance is consistent with other characterizations of WGS-based NIPS, and was achieved in the absence of a fetal-fraction failure threshold. Our results spanning tens of thousands of patients demonstrate that Prelude achieves high accuracy and a low test-failure rate in a general obstetric population..

477F

Prenatal diagnosis of tuberous sclerosis complex using high-throughput DNA sequencing combined with fetal echocardiography. J. Chen¹, J. Wang², W.D. Kitcheners³, L. Han¹, X. Gu¹, X. Hao¹, Y. Fu¹, Z. Weng⁴, J. Shang⁵, Y. Huang⁶, Y. Xiong⁷, B. Liu⁸, H. Zhang⁹, Y. He¹. 1) Maternal-Fetal Consultation Center of Congenital Heart Disease, Department of echocardiography, Beijing An Zhen Hospital, Capital Medical University, Beijing, China; 2) College of life science, Tsinghua University, Beijing, China; 3) Department of Pathology, Stanford University, CA, USA; 4) Department of Ultrasonography, Fujian maternal and child hospital, Fujian, China; 5) Department of Pathology, An Zhen Hospital affiliated to Capital Medical University, Beijing, China; 6) Biodynamic Optical Imaging Center (BIOPIC), Peking University, Beijing, China; 7) Department of Ultrasonography, Shenzhen People's Hospital, Shenzhen, China; 8) Department of Ultrasonography, the Second Affiliated Hospital, Medical College of Xi'an Jiao Tong University, Xi'an, China; 9) Department of cardiac surgery, Beijing An Zhen Hospital affiliated to Capital Medical University, Beijing, China.

Objective To describe the role of high-throughput DNA sequencing and fetal echocardiography in the risk assessment and diagnosis of tuberous sclerosis complex (TSC) in high risk pregnancies. **Methods** This was a prospective cohort study, 45 pregnant women with family histories of TSC or fetal cardiac tumours were recruited. Amniotic fluid or umbilical cord blood samples from pregnant women and peripheral blood samples from all family members were collected. Fetuses and family members were further analyzed for potential TSC1 and TSC2 mutations using next-generation DNA sequencing. **Results** Of all 45 families, 42 families (31 in group □, 11 in group □) had members and/or probands with TSC gene abnormalities; 9 of which were family genetic mutations and 33 were caused by sporadic mutations. For prenatal diagnosis, TSC gene abnormalities were detected in 17 out of 45 fetuses (6 in group □, 11 in group □). Among these 17 fetuses, 16 had single or multiple cardiac rhabdomyoma (CR) by FE. **Conclusions** TSC target gene detection combined with FE assessment was obtained within a timeframe suitable for the prenatal setting in families with high-risk of TSC and that this information assisted in the genetic counselling and informed decision-making in a large proportion of pregnancies.

478W

Application of clinical next generation panel sequencing using exome enrichment for the genetic assessment of fetal malformations. U. Hehr¹, T. Roedl¹, R. Chaoui², U. Germer³, G. Schlueter⁴, A. Ovens-Raeder⁵, S. Hinreiner¹. 1) Center of Human Genetics, Regensburg, Bavaria, Germany; 2) Center for Prenatal Diagnosis and Human Genetics Berlin, Germany; 3) Center for Prenatal Medicine, Caritas Hospital St. Josef, Regensburg, Germany; 4) Prenatal medicine Nuernberg, Nuernberg, Germany; 5) Private Practice for Human Genetics, Muenchen, Germany.

Introduction: Advances in imaging technologies and clinical expertise result in an ever increasing rate of sonographically detected fetal malformations, while starting to shift the time of detection from midgestation to earlier weeks of gestation for some of the more severe forms. However, sonography still uncovers only a small part of the fetal phenotype, commonly associated with a broad differential diagnosis. Rapid identification of the underlying mutations allows more precise counseling of the families on the expected clinical spectrum and long-term course, directly relevant for therapeutic options regarding the ongoing pregnancy. We here report first results of our lab on prenatal genetic testing of 82 fetal samples with sonographically detected fetal malformations consecutively tested by next generation sequencing (NGS) after exome enrichment. **Methods:** After Nextera exome enrichment for all 82 fetal samples an individual virtual gene panel, considering provided findings in fetal imaging, was applied for each case (mean: 7.8 genes/case), followed by bioinformatic assessment of called variants and evaluation for copy number variations with our in house pipeline. Variants were categorized according to the ACMG guidelines and validated with conventional sanger sequencing and MLPA, respectively. **Results and discussion:** 82 prenatal cases were referred for diagnostic genetic testing in the context of either fetal brain malformations (N = 57) or fetal skeletal disorder (N = 25). Overall causal mutations (ACMG class 5) or likely pathogenic variants (ACMG class 4) were identified for 24 cases (29.3%). For 8 of the fetal samples variants of unknown clinical significance (ACMG class 3) were detected. For 23 fetal samples within the holoprosencephaly spectrum a causal mutation was observed for 6 families. 17 fetal samples with hydrocephalus or suspected Walker Warburg syndrome resulted in mutation detection for 2 families. Testing for other brain malformations including lissencephaly, Microcephaly or cerebellar hypoplasia (N = 17) allowed mutation detection for 8 families. 25 fetal samples were referred with suspected fetal skeletal disorder, resulting in a genetic diagnosis for 8 families. **Conclusion:** Clinical next generation panel sequencing is a fast and cost-effective tool in the differential diagnosis of fetal malformations. Detailed pretest counseling of the families is mandatory and should include the option, that the obtained results may remain inconclusive.

479T

Smith-Magenis syndrome in two fetuses diagnosed with congenital heart defect. Y. Li¹, X. Hao¹, H. Sun¹, X. Gu¹, X. Liu¹, Y. He¹, S. Ge². 1) Department of echocardiography, Beijing Anzhen hospital, Capital medical university, Beijing, Beijing, China; 2) Section of Cardiology, Heart Center for Children St. Christopher's Hospital for Children Erie Avenue at Front Street Philadelphia.

Background: Smith-Magenis Syndrome (SMS) is a rare genetic disorder with an estimated prevalence of 1:15,000 to 1:25,000 live births, which is characterized by congenital heart defect (CHD), mental retardation, distinct facial appearances, sleep abnormalities, neurological and behavioral features and obesity. Approximately 90% cases of SMS are caused by a microdeletion in chromosome 17p11.2. RAI1 gene within this region can also cause this disorder. **Method:** A multi-center network was established to pool data of fetal echocardiography and fetal tissue samples after the families opted for termination of pregnancy for whole genome sequencing (WGS) and whole exome sequencing (WES). **Result:** A total of 649 fetal cases of fetal CHD were collected from 2013 to 2018, in which there were 2 fetuses identified as SMS. The first fetus was noted to have ventriculomegaly and Tetralogy of Fallot. The second fetus had Tetralogy of Fallot, right aortic arch and ectopic ductus arteriosus. There was a 3.63Mb and a 4.86Mb micro-deletion in their chromosome 17p11.2 by WGS, respectively. **Conclusion:** Prenatal diagnosis of SMS requires high index of suspicion of fetuses with CHD, especially TOF, and associated anomalies, e.g. ventriculomegaly. WGS and WES of high risk cases may identify SMS and aid in improved prenatal diagnosis, prognosis and parental counseling.

480F

Clinical and genetic description of Mexican patients with prenatally identified cardiac tumors. L.F. Mariscal¹, R. Sevilla¹, A.J. Martínez¹, I.E. Monroy¹, K. Carrillo², C. Molina², L.L. Flores², C. Alaez², M. Aguinaga¹. 1) Genetics, Genomics and cardiology, Instituto Nacional de Perinatología, Mexico, Mexico city, Mexico; 2) Genetics, Instituto Nacional de Medicina Genómica, Mexico, Mexico city, Mexico.

Introduction: Rhabdomyomas represent the most common type of prenatal cardiac tumors. When isolated, 50-70% are related to tuberous sclerosis complex (TSC), this number increases to 80-90% when multiple. Commonly, they are asymptomatic and have a natural history of involution at birth, 10% persist and might alter heart hemodynamics, valvular function and heart rhythm. The aim of this study was to reinforce the importance of additional clinical data in patients with prenatal heart tumor(s). **Methodology:** From 2010 to 2017, 10 prenatally detected cardiac tumors were referred to the Genetics Department and a complete family history with parent's clinical evaluation was done. Postnatal echocardiographic and full clinical evaluation was completed. NGS sequencing of the *TSC1* and *TSC2* genes was performed. **Results:** Five female and 5 male patients were prenatally detected with cardiac tumor(s) at 29.2 to 33 weeks of gestation, all cases were confirmed postnatally as rhabdomyomas. Four had pericardial effusion, 2 supraventricular tachycardia and 1 hydrops fetalis with subsequent prenatal death. Extra cardiac manifestations included cortical tubers, Shagreen patch and hypomelanotic macules. Four *de novo* and 4 family cases were detected, only 1 case was previously aware of the TSC diagnosis. Two cases had familial manifestations of TSC but did not meet the criteria for a definite diagnosis. Molecular analysis by NGS was performed in 4 patients finding 3 *TSC2* mutations. One case reported as pathogenic in exon 34, rs45517340 (NM_000548.3: c.4375C>T, NP_000539.2: p.(Arg1459Ter), this patient had a giant tumor and received Everolimus with a good outcome. Another case had a pathogenic variant in exon 10 with a reported rs397515108 and described as NM_000548.3: c.910T>C, NP_000539.2: p.(Trp304Arg) with an *in silico* prediction as probably pathogenic. A mutation with no previous registry (June 2017) was found in *TSC2* with an *in silico* prediction as probably pathogenic. In 1 case, we were not able to identify a mutation by NGS in these 2 genes, family history describes mother and uncle with definitive diagnosis for TSC; we are still working on further molecular analysis. **Discussion:** We report a high sensitivity rate in prenatal diagnosis of cardiac tumors with a 60% association with TSC as reported in literature. An early diagnosis enables early treatment and permits a further analysis of family members using the fetus as a clue for familial disease diagnosis.

481W

Pre- and perinatal history in kabuki syndrome. C.E. Rosenberg¹, T. Daly², C. Hung², E. Gussoni², A.W. Lindsley¹, O. Bodamer². 1) Division of Allergy and Immunology, Department of Pediatrics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 2) Division of Genetics and Genomics, Department of Pediatrics, Boston Children's Hospital, Boston, MA.

Background: Kabuki Syndrome (KS), a disorder of epigenetic dysregulation, is due to heterozygous mutations in either *KMT2D* or *KDM6A*, genes which encode a lysine specific methyltransferase and a lysine specific demethylase, respectively. The clinical phenotype of KS is highly variable, including congenital cardiac and renal anomalies, developmental delay, hypotonia, failure to thrive, short stature and immune dysfunction. All affected individuals have characteristic facial features. As the natural history of KS has not been fully delineated, limited information exists on the pre- and perinatal history of patients with KS. **Methods:** Two tertiary centers (Boston Children's Hospital and Cincinnati Children's Hospital Medical Center) collected retrospective pre- and perinatal historical data from patients with KS (N=33) using a questionnaire followed by in-depth review of medical records. **Results:** Data from 33 patients with KS (age: 19m-33y; 14 male; 14 with mutations in *KMT2D*, 1 with mutation in *KDM6A*, 18 diagnosed clinically) were included in the analysis. Maternal age at conception was 22-39 years (average: 30.2 years) and paternal age was 26-43 years (average: 33.5 years). Eleven of 33 (33%) pregnancies were complicated by polyhydramnios, including one where the fetus was diagnosed with cystic hygroma and shortened femur length on ultrasound. One pregnancy was complicated by gestational diabetes and pre-eclampsia. Delivery mode was vaginal in 17 of 29 patients, with C-section due to various indications in the remainder. **Discussion:** These data comprise the first large dataset on pre- and perinatal history in patients with confirmed, either genetically or clinically, KS. A relatively large number of pregnancies were complicated by the identification of polyhydramnios on ultrasound, which may be secondary to abnormal craniofacial structures and functional impairment of swallowing. **Conclusion:** Kabuki syndrome needs to be considered in the differential diagnosis in pregnancies complicated by polyhydramnios.

482T

Rapid diagnostic testing for fetuses with multiple congenital anomalies detected through ultrasound scan using exome sequencing. B. Sikke-ma-Raddatz, N. Corsten-Janssen, J.C.D. Diphorn, J. el Mecky, K. Bouman, J.B.G.M. Verheij, R. Kinds, A.J. Scheper, R.F. Suijkerbuijk, Y.J. Vos, I.M. van Langen, R.J. Sinke, R.H. Sijmons, H. Westers, C.C. van Diemen. Genetics, University of Groningen, University Medical Center, Groningen, Groningen, Netherlands.

Introduction Identifying the cause of fetal anomalies seen on ultrasound provides important information to improve perinatal management. Conventional genetic testing (QF-PCR, microarray) leads to a diagnosis in approximately 40% of fetuses with multiple congenital anomalies. Implementing exome sequencing (ES) in the prenatal setting is challenging due to uncertainties about fetal phenotyping, variant interpretation, counseling of incidental findings, and the requirement of short turnaround times. ES was implemented in prenatal care to increase the number of genetic diagnoses to aid in medical decision-making for parents and physicians. **Methods** Phase 1: Retrospective anonymous ES analysis of six fetuses (including parents) with known postnatal genetic diagnosis to test if this diagnosis could be made based on the fetal phenotype only. Variants were filtered using human phenotype ontology (HPO) or using our custom virtual gene panel (approximately 3850 disease genes, excluding late-onset disease genes). Phase 2: Prospective study of rapid trio WES analysis in addition to conventional genetic tests for twenty-five fetuses with abnormalities detected through ultrasound. Inclusion criteria are at least the presence of two congenital malformations as seen on ultrasound or one congenital malformation in addition to a previous pregnancy with a fetus with a similar phenotype. Questionnaires and interviews with patients about the test will be conducted. **Results** Phase 1: Five out of six known diagnoses could be confirmed using our gene panel. One causal pathogenic *PTPN11* mutation was missed due to low coverage of the variant. Filtering with HPO terms was not helpful given the continuously changing fetal phenotype. Phase 2: the first six fetuses have been included (May 2018), resulting in a genetic diagnosis in three: (1) unclear sex, club feet, possible ASD and a *de novo* mutation in *SAMD9* (MIRAGE syndrome) (2) nuchal translucency, fluid surrounding thorax and abdomen, club feet and compound heterozygous mutations in *PEX1* (Zellweger syndrome) (3) severe hydrocephaly, also seen in a previous pregnancy and homozygous mutations in *POMGNT1* (Walker Warburg syndrome). Four incidental findings were reported. The turnaround time was 6-14 working days. **Conclusion** First results from our prospective study show that implementing ES as a routine test in the prenatal setting is feasible. The test has a promising diagnostic yield, helping parents in decision making.

483F

Damaging *de novo* variants in congenital diaphragmatic hernia patients are associated with neurodevelopmental outcomes. L. Qiao^{1,2}, J. Wynn¹, L. Yu¹, V. Duron³, G. Aspelund⁴, C. Farkouh-Karoleski⁵, A. Zygmunt⁶, R. Hernan⁷, F. Lim⁸, T. Crombleholme⁹, R. Cusick⁶, K. Azarow⁷, M. Danko⁸, D. Chung⁶, B. Warner⁶, G. Mychaliska¹⁰, D. Potoka¹¹, A. Wagner¹², M. ElFiky¹³, S. Soffer¹⁴, D. Schindel¹⁵, Y. Shen^{2,16,17}, W. Chung^{1,18,19}. 1) Department of Pediatrics, Columbia University Medical Center, New York, NY, USA; 2) Department of Systems Biology, Columbia University Medical Center, New York, NY, USA; 3) Department of Surgery, Columbia University Medical Center, New York, NY, USA; 4) Department of Neonatal Pediatrics, Columbia University Medical Center, New York, NY, USA; 5) Cincinnati Children's Hospital, Cincinnati, OH, USA; 6) Children's Hospital & Medical Center of Omaha, University of Nebraska College of Medicine, Omaha, NE, USA; 7) Department of Surgery, Oregon Health & Science University, Portland, OR, USA; 8) Monroe Carell Jr. Children's Hospital, Vanderbilt University Medical Center, Nashville, TN, USA; 9) Washington University, St. Louis Children's Hospital, St. Louis, MO, USA; 10) Division of Pediatric Surgery, Department of Surgery, University of Michigan School of Medicine, Ann Arbor, MI, USA; 11) Division of Pediatric Surgery, Department of Surgery, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA; 12) Medical College of Wisconsin, Milwaukee, WI, USA; 13) Department of Pediatric Surgery, Faculty of Medicine, Cairo University, Cairo, Egypt; 14) Department of Surgery, Northwell Health, Great Neck, NY, USA; 15) Children's Hospital Dallas, Dallas, TX, USA; 16) Department of Biomedical Informatics, Columbia University Medical Center, New York, NY, USA; 17) JP Sulzberger Columbia Genome Center, Columbia University, New York, NY, USA; 18) Department of Medicine, Columbia University Medical Center, New York, NY, USA; 19) Herbert Irving Comprehensive Cancer Center, Columbia University, New York, NY, USA.

Congenital diaphragmatic hernia (CDH) is a major congenital abnormality that is associated with significant mortality, long term morbidity, and developmental impairment in some but not all individuals. To identify genetic associations with differential developmental and clinical outcomes of CDH patients, we prospectively enrolled neonates with CDH and followed for up to five years. We performed whole genome or exome sequencing in 409 parent offspring trios to identify *de novo* variants. We classified CDH patients into three groups based on the type of *de novo* variant: (a) likely gene disrupting (LGD) and predicted deleterious missense variants in constrained genes or known risk genes of CDH or common comorbid disorders (Group A, n=72); (b) LGD in other genes or predicted mildly deleterious missense variants in any gene (Group B, n=133); (c) other variants (Group C, n=204). We assessed seven outcomes: pulmonary hypertension at one month (n=299) and three months of age (n=245), mortality prior to initial discharge (n=407), Bayley Scales of Infant Development (BSID-III, n=183) and Vineland Adaptive Behavior Scales (VABS-II, n=224) at two years of age, Wechsler Preschool and Primary Scale of Intelligence (WPPSI-IV, n=65) and VABS-II (n=80) at five years of age. We also collected baseline clinical factors and socioeconomic data. We tested the association separately in complex (CDH with additional anomalies, n=152, 37% of all cases) and isolated (n=256) cases. We observed an association of lower intelligence score assessed by WPPSI with group A in complex cases (P-value=0.02), especially significant when adjusted for neonatal factors (P-value=4.5×10⁻³). There is a marginal association of language delay (by the BSID-III) with group A variants (p=0.03) in isolated cases. The severity of pulmonary hypertension at three months were correlated with genetic risk variants in group A (P-value=0.03). No significant difference was observed in pulmonary hypertension at one month or mortality prior to initial discharge comparing the three genetic groups. In summary, we found *de novo* deleterious variants in constrained or CDH candidate risk genes to be associated with lower intelligence score of CDH patients at five years of age and marginally associated with pulmonary hypertension at three months of age and language scores at two years of age. These results have important implications for prognosis and potential intervention and long term follow up for children with CDH.

484W

Prenatal phenotypes with segmental overgrowth related to**PI3K-AKT-mTOR pathway mutations: Phenotypic and molecular char-**

acterization. N. Bourgon^{1,2}, P. Kuentz^{1,2,3}, V. Carmignac^{1,2}, A. Sorlin^{1,2,4}, Y. Dufourd^{1,2}, E. Tisserand^{1,2}, M. Chevarin^{1,2}, T. Jouan^{1,2}, N. Chassaing⁵, A. Guichet⁶, J. Martinovic⁷, C. Quelin⁸, B. Doray⁹, C. Abel¹⁰, M.-J. Perez¹¹, D. Amram¹², T. Attie-Bitach¹³, A. Vasiljevic¹⁴, C. Thauvin^{1,2,4}, J.-B. Riviere^{1,2}, P. Vabre^{1,2,15}, L. Faivre^{1,2,4}. 1) Génétique des Anomalies du Développement, INSERM UMR 1231, Université de Bourgogne, DIJON, France; 2) Fédération Hospitalo-Universitaire Médecine Translationnelle et Anomalies du Développement, CHU Dijon Bourgogne, DIJON, France; 3) Génétique Biologique Histologie, CHRU de Besançon, BESANCON, France; 4) Centre de Génétique, CHU Dijon Bourgogne, DIJON, France; 5) Service de Génétique Médicale, CHU Toulouse, TOULOUSE, France; 6) Unité de Génétique, CHU Angers, ANGERS, France; 7) Unité de Foetopathologie, Hôpital Antoine Bécclère, CLAMART, France; 8) Service de Génétique, CHU Rennes, RENNES, France; 9) Service de Génétique, CHU Felix Guyon-Saint-Denis, SAINT DENIS, La réunion, France; 10) Service de Génétique Médicale, Hospices Civils de Lyon, Hôpital de la Croix Rousse, LYON, France; 11) Département Génétique Médicale, Hôpital Arnaud de Villeneuve, CHRU de Montpellier, MONTPELLIER, France; 12) Service de Génétique pédiatrique, CHIC Créteil, CRETEIL, France; 13) Département de Génétique, Unité de Génétique Clinique, Hôpital Necker Enfants Malades, PARIS, France; 14) Centre de Pathologie Est, Hospices Civils de Lyon, BRON, France; 15) Service de Dermatologie, CHU Dijon, DIJON, France.

Introduction: Germline or post-zygotic activating mutations of *PI3K/AKT/mTOR* pathway lead to a spectrum of segmental overgrowth syndromes that can be detected during the prenatal period. We report the first cohort of prenatal phenotypes related to activating mutation of *PI3K/AKT/mTOR* pathway. **Material and methods:** We gathered fetal cases suspected of activating mutations of *PI3K/AKT/mTOR* pathway that led to fatal issues and tested in the reference molecular diagnostic laboratory of Dijon University Hospital. After precise phenotyping with prenatal and fetopathological data, we performed ultra-deep targeted sequencing of genes of the pathway on a MiSeq platform (*PIK3CA*, *PIK3R2*, *AKT3*, *CCND2*, *PTEN*) in several tissues depending on the fetal phenotype. **Results:** We ranked 14 fetuses from termination of pregnancies (12/14) and perinatal death (2/14) into two main groups of fetuses: fetuses without cerebral overgrowth and fetuses with cerebral overgrowth, symmetrical (megalencephaly) or asymmetrical (hemimegalencephaly). Eight fetuses had an asymmetrical cerebral overgrowth corresponding to a hemimegalencephaly detected prenatally, isolated in 6 cases, with the identification of a post zygotic *PIK3CA* or a germline *PTEN* mutation, and with significant extracerebral abnormalities discovered during autopsy in 2 fetuses, especially dermatological, corresponding to MCAP/CLOVES fetal phenotype secondary to a post-zygotic *PIK3CA* mutation. For the 6 fetuses with isolated hemimegalencephaly, the fetal phenotype was secondary to a. [A2] Three fetuses [A3] had a prenatal extended lymphangioma associated with wide feet with plantar hypertrophy, toe abnormalities with a wide sandal gap suggesting a CLOVES syndrome, with identification of a postzygotic *PIK3CA* variant for 2 of them. Three fetuses had a prenatal mild ventriculomegaly associated with macrocephaly corresponding to megalencephaly and diffuse polymicrogyria suggesting a MPPH syndrome, all carrying a *PIK3R2* variant, 2 germline *de novo* and one post-zygotic (mosaicism level=8.51%). **Conclusion:** This is the first prenatal cohort of extreme forms of phenotypes related to activating mutation of *PI3K/AKT/mTOR* pathway, which demonstrates a high clinical and molecular heterogeneity. Thanks to its results, we propose a decision tree according to the fetal phenotype to guide the targeted sequencing.

485T

Maternal factors associated with DNA methylation-based placental aging.

T. Workalemahu¹, D. Shrestha¹, S. Tajuddin², F. Tekola-Ayele¹. 1) Epidemiology Branch, Division of Intramural Population Health Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD; 2) Laboratory of Epidemiology and Population Sciences, National Institute on Aging, National Institutes of Health, Baltimore, MD, USA.

Disruption of the physiological aging of the placenta can deter its hormonal and transport function and may lead to several adverse pregnancy outcomes. Recent "epigenetic clock" studies have begun profiling lifestyle and socio-demographic factors that influence tissue age in adults. Understanding the maternal factors that predict DNA methylation-based placental aging provides intervention opportunities to improve pregnancy outcomes, but these factors are largely unknown. We investigated whether maternal socio-demographic factors and genetic ancestry are associated with placental aging. Participants with placental DNA methylation and genotype data (n=301) were obtained from a sub-cohort of low-risk pregnant women of the NICHD Fetal Growth Studies. Participants were composed of self-identified Non-Hispanic Black (n=72), Non-Hispanic White (n=77), Hispanic (n=102), and Asian/Pacific Islander (n=50) race/ethnicities. The Illumina Infinium 450 BeadChip array was used to quantify over 400k CpG sites. Placental DNA methylation age was predicted using 62 CpG sites that were known to predict placental aging. Placental age acceleration was defined to be the difference between gestational age at birth and placental DNA methylation age. Genome-wide proportions of African, European, Native American, and East Asian genetic ancestries of the mothers were estimated using ADMIXTURE. Placental DNA methylation age was positively correlated with gestational age at birth (r=0.20; p<0.01). Placental age acceleration was positively correlated with maternal age (r=0.13; p=0.03), proportion of European ancestry among non-Hispanic Whites (r=0.27; p=0.02), and proportion of Native American ancestry among Hispanics (r=0.26; p=0.01), and negatively correlated with proportion of African ancestry among Hispanics (r=-0.24; p=0.01). In multivariable models adjusted for maternal age, pre-pregnancy body-mass-index, infant sex, parity, medical insurance, marital status, education, mode of onset of labor, intra-uterine growth restriction status and preeclampsia status, the associations with African and Native American ancestry among Hispanics remained significant. In summary, findings suggest that mother's age and genetic ancestry may influence molecular markers of placental aging. Future studies are needed to understand whether adverse pregnancy outcomes known to be associated with advanced maternal age and population groups are mediated through accelerated aging of the placenta.

486F

Comparative Genomic Hybridization (aCGH) and Next Generation Sequencing (NGS) combine together give a comprehensive genetic diagnosis approach for prenatal anomalies. S.S. Chettiar^{1,2}, A.R. Patel^{1,2}, D.D. Jhala³, J.J. Dattani³, U. Radhakrishna⁴. 1) Human Genomics, Genexplore Diagnostics and Research Centre, Ahmedabad, Gujarat, India; 2) Department of Human Genetics, Gujarat University, Ahmedabad, India; 3) NVBDP, Ministry of Health and Family Welfare, Government of India; 4) Department of Obstetrics and Gynecology, Oakland University- William Beaumont School of Medicine, Royal Oak, MI, USA.

Fetal anomalies are identified in 2-5% of normal pregnancies and are reported as high as 12-18% in high risk pregnancies, which also attributes to 28% perinatal deaths. Most common developmental anomalies observed are in Central nervous system, Cardiovascular system, Gastrointestinal tract and Urinary tract. Diagnosis and prognosis approach vary in different anomalies observed, although most common and conventional approach adopted is karyotype followed by aCGH. Karyotype identifies 30% of aneuploidy and 5% of balanced translocation. While aCGH can detect additional 6.5% of CNVs, deletion and duplication. But together karyotype and aCGH only constitute to 37% of genetic changes contributing to the fetal anomalies observed, still leaving majority of cases undiagnosed. Thus there is need to add another dimension of approach to ascertain genetic etiology. Next generation sequencing give much more resolution upto single base pair level and at the same time encompassing a large number of genes or entire exome which will transcribe in to set of amino acid sequences in protein. In our study NGS has demonstrated 22% of diagnostic yield which was not been picked up by conventional methods. We studied 238 amniotic fluid samples with varied clinically reported radiological indications. Following conventional method cytogenetically 72 (30.25%) samples were found to have causative variation, CMA showed CNVs in 18 (7.54%) samples and NGS approach provided positive hits in 51 (21.42%) of cases. CMA data revealed CNVs in novel loci like Chr19p13.11-p12 duplication in fetus with echogenic intracardiac focus, 4q35.2 microdeletion in retrognathia, 15q11.1-11.2 duplication in intrauterine growth retardation, deletion in Chr8p11.22 associated with atrioventricular septal defect, fetus with bilateral Ventriculomegaly and facial dimorphism showed deletion in Chr3p26.2-26.1, ChrX22.33 and duplication in Chr10q26.3. NGS analysis provided novel mutation in genes DSC2, 2687GA with fetus reported presenting Arrhythmogenic right ventricular dysplasia; KCNQ1, G229D associated to fetal atrial fibrillation; FBN1, 5280G fetus with aortic aneurysm; DSG2, ARG48HIS associated to Arrhythmogenic right ventricular dysplasia; MYBPC3, ASP745GLY in aortic aneurysm and TNNT2,3-BP DEL, 285GGA in cardiac anomalies. Although data show no one method suffices to have significant mutation detection, we recommend a comprehensive parallel approach with all three methods.

487W

Aspirin enhances trophoblast invasion and represses sFlt-1 production: A putative mechanism for preventing preeclampsia. M. Su¹, C. Wang², H. Tsai¹. 1) Depart. of OBGYN, National Cheng Kung University Hospital, Tainan, Taiwan; 2) Department of Cell Biology and Anatomy, National Cheng Kung University, Tainan, Taiwan.

BACKGROUND Preeclampsia is a severe gestational hypertensive disorder that represents a major threat to mortality and morbidity in maternal and fetal health worldwide. Recent studies suggested that prophylactic aspirin prior to 16 weeks of gestation in high-risk patients may reduce the risk of developing preeclampsia, however, the exact mechanism of aspirin's effect on the pathophysiology of preeclampsia is not clear. **OBJECTIVE** This study was designed to investigate the effect of aspirin on trophoblast cell function and its effect on soluble Flt-1 (sFlt-1) production to elucidate the underlying molecular mechanisms in vitro. **STUDY DESIGN** We used two human trophoblastic cell lines (HTR-8/SVneo and JRA cells) to determine the effect of aspirin on trophoblast cell function. Trophoblasts were pretreated with aspirin, and then cell functions (cell proliferation, cell invasion) and sFlt-1 expression were assessed. In addition to RNA level and the activity of matrix metalloproteinase (MMP) 2 and MMP9 by real-time PCR and gelatin zymography, respectively, we also evaluated ERK and Akt signaling to decipher the underlying mechanism by which aspirin regulates trophoblast biology. The effect of sFlt-1 on trophoblast invasion and MMP2/9 was further validated by treating HTR-8/SVneo cells with human Flt-1 recombinant protein and performing siRNA knockdown studies. **RESULTS** We observed that aspirin promoted trophoblast invasion in HTR-8/SVneo and JAR cells. sFlt-1 protein production was repressed by aspirin in a dose-dependent manner in HTR-8/SVneo cells, but sFlt-1 expression was barely detected in JAR cells. By adding human Flt-1 recombinant protein, the invasion ability of HTR-8/SVneo cells was inhibited, which was reversed by FLT-1 siRNA knockdown. In addition, we found that MMP2/9 expression and activity in HTR-8/SVneo cells was activated by aspirin treatment but inhibited by sFlt-1. Aspirin also downregulated Akt phosphorylation at Thr308 and Ser473 and trophoblast invasiveness was facilitated under Akt inhibitor treatment. **CONCLUSION** Aspirin enhances cell invasion ability and inhibits sFlt-1 production in trophoblasts. Moreover, sFlt-1 itself also inhibits trophoblast invasion. Our data suggest that the preeclampsia prevention effect of aspirin may be exerted through two mechanisms: the effect of aspirin itself on promoting trophoblast invasion and the effect of aspirin in repressing sFlt-1 production.

488T

Improved non-invasive prenatal testing by mitigating false predictions caused by maternal mosaic aneuploidy using size-based DNA. H.J.

Kwon¹, S.Y. Yun¹, A. Goyal¹, J.H. Kim¹, H.S. Im¹, J.S. Joo¹, J.S. Bae¹, M.S. Lee^{1,2,3}, S.H. Lee¹. 1) Eone-Diagnomics Genome Center, Inc., 291 Harmony-ro, Yeonsu-gu, Incheon, 22014, Republic of Korea; 2) Diagnomics, 5775 Kearny Villa Rd., San Diego, CA 92123, USA; 3) Incheon Nat'l University, 119 Academy-ro Yeonsu-gu, Incheon, 22012, Republic of Korea.

Background A noninvasive prenatal test (NIPT) is a method of identifying fetal chromosomal abnormalities using cell-free DNA in the maternal plasma. Although many new algorithms have been developed and improved in accuracy, the false positives caused by the low fetal fraction and the maternal mosaic aneuploidy are still difficult to avoid. Circulating fetal DNA molecules are generally shorter in size than the corresponding maternal DNA molecules. In view of fact that the DNA fragment size can be used as diagnostic parameter, we calculated z-score of enriched fetal and maternal DNA respectively and found a solution to reduce the false NIPT results which might be caused by low fetal fraction, partial duplication and deletion of target chromosome or maternal mosaic aneuploidy. **Method** We used paired-end sequencing reads to identify the size distribution of fetal and maternal DNA in cell-free maternal plasma. First, paired-end reads from short cfDNA fragments of less than 150bp, "fetal reads", were selected to enrich cfDNAs derived from fetus cells while those from larger fragments of over 185 bp, "maternal reads" were used to enrich cfDNAs from maternal cells. Second, we calculated z-scores using GC corrected fetal and maternal reads across all chromosomes and analyzed the presence of aneuploidies of both fetal and maternal reads. Finally, we assessed how the maternal cfDNA affected the overall results or how the NIPT algorithm can be more adequate to detect the fetus aneuploidy. Finally, we made final decision by comparing the two results from fetal and maternal reads. **Results** Compared to conventional NIPT algorithm, our method could detect aneuploidy more clearly by comparing the results from fetal and maternal reads when analyzing trisomy 21 samples of very low fetal fractions of 3~5%. In addition, our algorithm could distinguish correctly fetal X monosomy from maternal mosaic aneuploidy that can cause false prediction result. **Conclusion** We amplified the fetal signal by distinguishing the maternal DNA specific fragments by size selection and consequently reduced the false negative or false positive detection due to the low fetal fraction or maternal mosaic aneuploidy and partial duplication/deletion, respectively.

489F

GWAS for gestational hypertension using time-series home blood pressure data in maternity log study. Y. Tsunemoto¹, T. Yamauchi^{1,2}, D. Ochi^{1,2}, K.

Kojima², T. Mimori², S. Kumatani¹, R. Yamashita², M. Wagata², Y. Harada², F. Katsuoka², N. Minegishi², O. Tanabe², S. Hiyama¹, M. Nagasaki², J. Sugawara^{2,3}, Maternity Log Study Group. 1) Research Laboratories, NTT DOCOMO, INC., Hikarino-oka, Yokosuka, Kanagawa, Japan; 2) Tohoku Medical Megabank Organization, Tohoku University, Seiryomachi, Aobaku, Sendai, Miyagi, Japan; 3) Department of Gynecology and Obstetrics, Tohoku University Graduate school of Medicine, Seiryomachi, Aobaku, Sendai, Miyagi, Japan.

Objective: Hypertensive disorders of pregnancy (HDP) including gestational hypertension (GH) and preeclampsia (PE) is a major cause of perinatal mortality in both mothers and children, but the pathogenic mechanisms of HDP have not been clarified. Although there exist several genome-wide association studies (GWASs) for HDP, most of the studies considered only the binary disease status, and significant loci identified in these studies are limited. In order to overcome this limitation, we conducted a GWAS for a continuous phenotype on daily home blood pressure of pregnant women. Since most of HDP patients in this study were diagnosed as GH and since there were fewer reports of GWASs for GH compared to PE, we focused on GH in this study. **Methods:** We used daily home blood pressure data and whole genome sequencing data of 190 pregnant women including GH patients (n=16) and controls (n=174), who participated in Maternity log study (MLOG). MLOG is an add-on cohort study on a part of participants of a prospective cohort study by Tohoku Medical Megabank Organization in Japan. In MLOG, pregnant women were recruited in the first or second trimester of pregnancy at Tohoku University Hospital. By using a smart phone and monitoring devices, participants uploaded daily life-logs such as home blood pressure. Furthermore, specimens such as blood samples for genome sequencing and multi-omics analyses were collected. Blood pressure of GH patients showed greater increments during pregnancy than controls. Therefore, we conducted a GWAS for the increment of daily home blood pressure as a continuous phenotype, not only GWASs for binary phenotypes. **Results:** We identified several significant ($p < 5 \times 10^{-8}$) SNPs by conducting a GWAS for the continuous phenotype instead of binary phenotypes. Moreover, these significant SNPs were located near the genes related to hypertension or placental function. **Discussions:** In this study, we conducted an unprecedented GWAS for GH using time-series data of home blood pressure. Despite the small number of samples, we could identify significant SNPs that were not found by GWASs with binary phenotypes. This result suggests that using detailed and continuous biometric information is effective for GWAS with a limited sample size.

490W

Deciphering the causes of early human pregnancy loss by exome analysis of abortuses. S.K. Garushyants¹, E. Nabieva¹, M.D. Logacheva¹, T.V. Neretina², A. Zhdanova¹, N. Libman³, D. Pyankov⁴, G.A. Bazykin¹. 1) Skolkovo Institute of Science and technology, Moscow, Russia; 2) Lomonosov Moscow State University, Moscow, Russia; 3) Genetico LLC, Moscow, Russia; 4) Genomed, Moscow, Russia.

Only a third of human conceptions result in live births, which comprises a substantial burden to human reproduction. A large fraction of pregnancy losses are associated with chromosomal abnormalities. The genetic basis of embryonic lethality in the remaining euploid embryos has been the subject of a few recent analyses employing whole exome sequencing. However, those studies have focused on embryos of advanced gestational age. The causes of early pregnancy loss in karyotypically normal embryos remain largely unknown. Here, we present the early results of a whole exome sequencing study of spontaneous euploid abortuses below 16 weeks gestational age, aiming to determine the inherited or *de novo* genetic variation (DNV) that could have led to pregnancy loss. We plan to sequence 100 mother-father-abortus trios; to date, we have sequenced eight trios. The mean gestational age was 10.5 weeks. As candidate causative variants, we considered inherited or *de novo* variants in genes associated with embryonic lethality in mice, and in genes with high intolerance to protein truncating variants (PTVs). We considered nonsynonymous variants with Polyphen score >0.996. For inherited variants, we required allele frequency below 1% and absence in homozygous state in the general population. For *de novo* variants, we required absence in the general population. In 3/8 cases, we detected no candidate variants. In one case, we detected two possible causative DNVs: one DNV in *SOX4* (p.Gly420Asp), a gene whose knockouts in mice were shown to cause embryonic lethality; and one DNV in *BCL9L* (p.Lys275Asn), which is known to be PTV intolerant, while mutations in this gene are predicted to be autosomal dominant. In another case, we detected a possible causative DNV at splice site of *DLD* (c.40-2A>G); a knockout of this gene causes embryonic lethality in mice. In three additional cases, we detected compound heterozygotes in essential genes *PNPT1*, *LAMA5*, and *PTPN6*. This study will add to the emerging picture of the diversity of causes of embryonic lethality.

491T

Cell-free DNA in embryo culture medium of in-vitro fertilization (IVF) for preimplantation genetic testing (PGT) for aneuploidy. K.W. Choy, Q. Yeung, Y. Zhang, T.C. Li, G.W.S. Kong, J.P.W. Chung. Department of Obstetrics and Gynaecology, The Chinese University of Hong Kong, Hong Kong, China.

To circumvent the risk of causing irreversible damage to a preimplantation embryo, the development of Non-Invasive Prenatal Testing (NIPT) on prenatal diagnosis provides a new insight to the methodology of genetic screening. Literatures have proven that embryonic DNA is released when an embryo begins to divide and grow during its preimplantation development. Since an embryo is confined to an in-vitro environment during IVF, the culture media in which the embryo has been handled and cultured provide a valuable source for sampling for PGT. We aim to study could the same genome profile be identified if spent culture media (SCM) instead of trophectoderm (TE) were used for aneuploidy screening (PGT-A). This is a prospective study comparing PGT-A results of SCM samples to the routine results based on TE of the same embryos. Couples undergoing PGT-A cycles were recruited from a university hospital infertility centre. The study is still ongoing, and this interim review is consisted of 70 blastocysts from 24 PGT-A cycles between March 2017 and September 2017. All recruited cycles underwent intracytoplasmic sperm injection (ICSI). All TE samples were subjected to routine PGT-A by aCGH (GenetiSure Pre-Screen 8x60K microarray, Agilent), from which 69 (98.6%) produced an interpretable result. For SCM from the 70 corresponding embryos, 60 samples (85.7%) had cell free DNA that were successfully amplified, then subjected to NGS (Veriseq, Illumina). Thirty-nine of them (65.0%) giving an interpretable result. There was an 88.2% concordance rate in sex determination. When SCM results were classified into euploid and aneuploid, the concordance rate with TE results was 71.8%. Overall, 12 euploids were identified by SCM, in which 6 were also euploids based on TE biopsy, 3 contained low level of mosaicism, 1 involved a segmental error on one chromosome, and 2 false negatives probably due to presence of cumulus cells in the SCM. Interestingly, there were 3 aneuploids identified by SCM which were euploids by TE biopsy. These three embryos were already transferred based on their euploid status as established from TE, resulting in 2 miscarriages and 1 ongoing pregnancy. In summary, SCM from embryos have the potential to become the first non-invasive method in euploid identification and gender determination. Much research is needed to elucidate the origin of cell free DNA found in SCM.

492F

Correlation of hemolysis and fetal fraction in cell-free DNA blood collection tubes for noninvasive prenatal testing. R. Stokowski, C. Hacker, K. White, R. Jones, M. Schmid. Ariosa Diagnostics Inc., Roche Sequencing Solutions Inc., San Jose, CA.

Objective: Lysis of maternal white blood cells in prenatal cell-free DNA (cfDNA) test samples will increase the level of maternal genomic DNA and consequently decrease fetal fraction. The objective of this study was to determine whether hemolysis, traditionally used as a marker for cell lysis, is correlated with a decrease in fetal fraction in maternal blood samples collected in specialized cfDNA tubes for noninvasive prenatal testing. **Methods:** Blood from pregnant women was collected into multiple cfDNA blood tubes. First, paired samples from the same subject were evaluated for a visual difference in hemoglobin level as a measure of hemolysis and fetal fraction measured using polymorphic DANSR assays. In a second arm of the study, clinical Harmony® test samples with hemoglobin levels of 500mg/dL and higher were tracked through the laboratory and their fetal fraction distribution compared to concurrently processed samples with lower hemoglobin levels. **Results:** There was no significant difference in fetal fraction in 170 paired samples with a difference in hemoglobin levels ranging from 30 to 980 mg/dL. The fetal fraction distribution in 203 tracked clinical samples with hemoglobin levels greater than or equal to 500mg/dL was statistically equivalent to the distribution in a concurrent series of 12,705 samples. **Conclusion:** Hemolysis as a measure of cell lysis in maternal blood samples collected in specialized cfDNA tubes does not correlate with a decrease in fetal fraction; therefore, it should not be a cause for rejection of samples submitted for prenatal cfDNA testing.

493W

Pan-ethnic expanded carrier testing for fatty acid oxidation disorders. N. Echeverri¹, J. Bardos¹, C. Terhaar², K. Hill-Harfe², R.E. Longman¹. 1) University of Miami Miller School of Medicine, Miami, FL; 2) Progenity, Inc., Ann Arbor, MI.

Introduction: Fatty Acid Oxidation Disorders (FAOD) are a group of autosomal recessive conditions characterized by impaired catabolism of mitochondrial fatty acid oxidation. The most common clinical presentation is hypoketotic hypoglycemia and sudden death in the neonatal period after an episode of fasting or infection. Individuals with milder defects may present later in life or even be asymptomatic. Historically, carrier testing guidelines were ethnicity-based and did not include testing for FAODs. New technologies and increasing ethnic diversity have enabled expanded carrier testing as an acceptable clinical practice. This study assesses how many individuals screened positive for five different types of FAOD when a pan-ethnic expanded carrier testing panel was implemented. **Methods:** A retrospective database analysis was performed on samples received for expanded carrier testing via a commercially-available genotyping platform. Expected number of positive carriers for five different types of FAOD were calculated and compared to observed rates: Carnitine palmitoyltransferase I deficiency (CPT I), Carnitine palmitoyltransferase II deficiency (CPT II), Long-chain 3-hydroxyacyl-CoA dehydrogenase deficiency (LCHAD), Medium-chain acyl-CoA dehydrogenase deficiency (MCAD), and Very long-chain acyl-CoA dehydrogenase deficiency (VLCHAD). Chi-squared analysis was performed to assess for statistical significance ($p < 0.05$). **Results:** A higher-than-expected number of individuals tested positive for CPT II deficiency ($p < 0.001$) and MCAD ($p < 0.001$). A lower-than-expected number of individuals tested positive for LCHAD ($p < 0.001$) and VLCAD ($p = 0.009$). The number of individuals that screened positive for CPT I deficiency was consistent with expected values ($p = 0.405$).

Disease	Total tested	Observed Positives	Expected Positives	p-value
Carnitine palmitoyltransferase I deficiency	24158	3	4.84	0.405
Carnitine palmitoyltransferase II deficiency	24150	121	37.19	<0.001
Long-chain 3-hydroxyacyl-CoA dehydrogenase deficiency	24158	28	84.07	<0.001
Medium-chain acyl-CoA dehydrogenase deficiency	24155	308	151.5	<0.001
Very long-chain acyl-CoA dehydrogenase deficiency	24157	14	27.77	0.009

Conclusion: Carriers of CPT II deficiency and MCAD were more common than expected in this cohort. Pan-ethnic expanded carrier testing will identify carriers of FAODs, allowing for more complete genetic counseling, family planning options, and prompt treatment of affected neonates.

494T

Conflicting cfDNA and microarray prenatal results: A unique repair process in a chromosomally normal fetus. V.R. Potluri¹, T. Boomer², S. Caldwell², M.K. Rudd², R. Mccullough³, P. Papenhausen², S. Schwartz². 1) Laboratory Corporation of America Holdings, Houston, TX; 2) Laboratory Corporation of America Holdings, Research Triangle Park, NC; 3) Sequenom Inc., San Diego, CA.

The field of prenatal diagnosis has evolved considerably over the past decade to include new screening tests (e.g cfDNA) as well as improved diagnostic testing (e.g. microarray analysis and exome sequencing). These tests are often complementary but on occasions provide conflicting results. This presents an interesting patient with seemingly conflicting results but in reality provide a unique underlying mechanism to illustrate how the results are inter-related. Initial cfDNA analysis suggested the presence of a deletion of part of the long arm of chromosome 13. Follow-up microarray analysis at 16.4 weeks gestation did not confirm the deletion, but indicated a long stretch of homozygosity (~57.4 Mb) in the long arm of chromosome 13 (13q21.1->13qter). Parental analysis indicated that while the proximal region of chromosome 13 was inherited in a biparental manner, it confirmed maternal segmental uniparental disomy for this terminal region. The pregnancy continued to term at which time placental microarray studies demonstrated a 57.5 Mb mosaic terminal deletion and allelic mosaicism of 13q21.1->13qter. This study is important clinically and mechanistically for a number of reasons including: (1) The cfDNA result clearly detected an anomaly in the placenta; (2) The follow-up microarray studies indicated that the fetus did not appear to have a deletion, but indicated the possibility of a repair process; (3) Standard cytogenetic analysis would not have provided an explanation for the cfDNA findings, just indicating that both chromosome 13s were structurally normal; (4) The placenta studies confirmed that there was a deletion in the placenta, but that a repaired process occurred leading to the normal fetal findings; (5) Although no fetal deletion was detected, there always remains a possibility of a phenotype due to developmental effects of the deficit before complete repair or a recessive disorder associated with the homozygosity and the patient needs to be counseled to these possibilities.

495F

Genome-wide trans-ethnic meta-analysis identifies a novel locus influencing longitudinal fetal growth at sensitive periods in pregnancy. F. Tekola-Ayele, J. Wu, M. Ouidir, M. Rahman, K.L. Grantz, C. Zhang, J. Grewal, T. Workalemahu. Division of Intramural Population Health Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD, USA.

Fetal growth is an important predictor of infant morbidity and mortality, and is associated with cardiometabolic diseases in children and adults. Genome-wide association studies (GWASs) involving European ancestry populations have identified genetic loci associated with birthweight. However, genetic influences on fetal growth vary at different windows of pregnancy and birthweight may not reflect *in utero* growth trajectory. We conducted GWAS meta-analyses of fetal weight at end of first trimester (13 gestational weeks), mid-gestation (20 weeks), late second trimester (27 weeks), and third trimester (40 weeks) of pregnancy in 1,849 self-identified African American, Asian, White, and Hispanic pregnant women recruited through the NICHD Fetal Growth Studies. SNPs were genotyped with the Infinium Multiethnic Global BeadChip and imputed using the 1000 Genomes reference panel. We carried out random-effects GWAS meta-analyses followed by trans-ethnic GWAS meta-analysis using a Bayesian approach to account for variations in allelic effects and linkage disequilibrium among ethnicities. We identified a genome-wide significant association of a novel SNP in the inositol 1,4,5-triphosphate receptor type 1 (*ITPR1*) gene with 30.8 g reduced fetal weight at second trimester of pregnancy (95% CI -41.9, -19.8, $\log_{10}BF=6.297$; $P=4.09 \times 10^{-6}$). The locus was also associated with reduced fetal weight at first trimester ($P=3.91 \times 10^{-6}$), mid-gestation ($P=2.55 \times 10^{-7}$), and third trimester ($P=4.41 \times 10^{-4}$) of pregnancy. Fine-mapping showed that the fetal weight reducing allele was common among African-Americans, Hispanics and Whites (12.5%-20.8%), and had low frequency among Asians (1.6%) in our cohort. The significantly associated *ITPR1* locus is within regulatory histone marks in several tissues including the placenta, and knockout of *itpr1* in mice has been shown to lead to extra-embryonic vascular defects, early growth retardation and embryonic lethality. Together, the present study sheds new light on the role of common genetic variants in the inositol receptor signaling pathway on fetal growth, potentially by regulating vascular development at the embryonic-maternal interface. The study also demonstrates that longitudinal fetal growth measures, beyond birth size, are valuable for discovering novel genetic loci, particularly through trans-ethnic genomic approaches.

496W

German Neonatal Network: Resource for genetic analyses in very low birth weight infants. *T.K. Rausch^{1,2}, I.R. König¹, W. Göpel²*. 1) Institut für Medizinische Biometrie und Statistik, Universität zu Lübeck, Universitätsklinikum Schleswig-Holstein, Campus Lübeck, Lübeck, Germany; 2) Department of Pediatrics, Universität zu Lübeck, Universitätsklinikum Schleswig-Holstein, Campus Lübeck, Lübeck, Germany.

Founded in 2009, the German Neonatal Network examines the development of very low birth weight infants until the age of five in Germany with 63 study centers. During their stay in hospital after birth the assessment includes the infants' basic data, medication, surgeries and therapies. After discharge from hospital, the parents report on their infants' development via an annual questionnaire. A follow-up team of study nurses and a medical doctor from Lübeck visits all study sites to examine the infants at the age of five years. Besides recording basic data they conduct several tests to examine eyes, ears, lung, movement and development. Umbilical cord tissue frozen after birth is used to genotype the DNA of the infants. Affymetrix Axiom™ Genome-Wide CEU 1 Array Plate 2.0 and Illumina Infinium® Global Screening Array-24 v1.0 were used for chip genotyping. Our continuously growing database already contains clinical data of 17,900 very low birth weight infants, genetic data of 7,400, and five-year follow-up data of 1,850 infants. In the main data set 49.1 % of 15,462 very low birth weight infants are female, the mean birth weight is 1,059.82 ± 306.76 g and the mean gestational age is 28.7 ± 2.76 weeks. Ultimately, genetic information will be available from all infants who were examined at the age of five years. A resource is already available providing answers to a wide variety of research questions.

497T

Whole exome sequencing of genomes of Korean patients with non-obstructive azoospermia. *M. Go¹, S. Sung², J. Park², K. Kang², Y. Shin², S. Song³, S. Shim^{1,2}*. 1) Biomedical Science Dept., College of Life Science, CHA University, Seoul, South Korea; 2) Genetic laboratory, Fertility Center of CHA Gangnam Medical Center, Seoul, South Korea; 3) Department of Urology, CHA Gangnam Medical Center, Seoul, South Korea.

Azoospermia affects approximately 1% or more of adult men. The most of genetic causes of non-obstructive azoospermia (NOA) cases are idiopathic except AZF (Azoospermia factor) microdeletion and Klinefelter syndrome. To date, whole exome sequencing (WES) technologies allow fast identification of variants in causative genes of target disease such as NOA (He et al. 2018; Nakamura et al, 2015). In this study, we used WES to identify possible pathogenic variants in genomes of Korean patients with idiopathic non-obstructive azoospermia. Sixteen Korean patients who diagnosed as idiopathic non-obstructive azoospermia were subjected to whole exome sequencing (WES). The Illumina HighSeq 2500 platform (Illumina, San Francisco, USA) was used and the libraries are prepared using SureSelect Exome V5 library kit. Cytological examination showed maturation arrest in all patients. As a result, we identified total 77 variants in 40 genes (data not shown) that might cause azoospermia. Two variants (c.551T>C of *TEX11* (NM_001003811) and c.724G>T in *DMRT1* (NM_021951)) were filtered with databases (1000G or ExAC) and prediction software (Mutation taster, SIFT, Polyphen-2, M-CAP, and CADD). *TEX11* and *DMRT1* are causative genes of non-obstructive azoospermia which were identified in previous studies (Nakamura *et al.*, 2015; Don *et al.*, 2011.). *TEX11* variant (c.551T>C) was expected to cause p.Leu184Pro and might affect the activity of domain of TEX11. Likewise, c.724G>T in *DMRT1* could cause p.Gly242Trp and predicted to alter the structure of DMRT1. As a further study, 74 patients with azoospermia and 96 normal males were screened. Because no one showed same variants on each *TEX11* and *DMRT1*, the variants that we identified have a possibility to be pathogenic. To summarize, we identified two novel variants that were predicted to be pathogenic in *TEX11* and *DMRT1*, respectively. Our result could not fully elucidate functional impacts of variants on spermatogenesis and therefore further studies are required. However, this study showed WES is a useful tool to identify variants in causative genes of non-obstructive azoospermia.

498F

CETN1 variants are associated with idiopathic male infertility: Genetic and functional perspectives. D.V.S. Sudhakar¹, R. Phanindranath¹, S. Jayashankar², A. Pawar³, D. Verma³, P. Nagarjuna², M. Deenadayal², K.V. Chary³, R. Singh⁴, K. Singh⁵, R. Dada⁶, A. Anand⁷, Y. Sharma⁸, J. Gopalakrishnan¹⁰, N.J. Gupta⁹, K. Thangaraj¹. 1) Centre for Cellular and Molecular Biology, Hyderabad, India; 2) JNCASR, Bangalore, India; 3) TIFR, Mumbai, India; 4) CSIR-CDRI, Lucknow, India; 5) Mangalore University, Mangalore, India; 6) IIRM, Kolkata, India; 7) IIRC, Secunderabad, India; 8) Banaras Hindu University, Varanasi, India; 9) All India Institute of Medical Sciences, New Delhi, India; 10) Center of Molecular Medicine, Cologne, Germany.

Introduction: Approximately, one out of every seven couples are infertile worldwide and the male factor infertility accounts for 30-50% of the infertility cases. The etiology of large proportion of infertile men still remains unknown. Therefore, to identify novel genetic factors associated with idiopathic male infertility, we selected *CETN1* based on our gene expression and exome sequencing studies and sequenced it in large cohort of case-controls and performed various functional studies. **Materials and methods:** We have performed exome sequencing of 44 idiopathic infertile men (azoospermia and severe oligozoospermia) using Illumina HiSeq-2000 platform with 100X coverage. Various bioinformatics tools and custom pipelines were used for data analysis and prioritizing the variants. We have identified 61 novel and rare variants (from 58 genes) and genotyped (using Sequenom) them in 740 cases and 700 ethnically matched controls. Of the novel genes identified, *CETN1* has been found to be a potential candidate; hence we have sequenced *CETN1* in additional 1073 infertile men and 874 ethnically matched fertile control men. Functional characterization of two significantly associated variants was performed using biophysical (circular dichroism, 2D-HSQC and ITC etc.), cell biology approaches and luciferase reporter assays. **Results:** rs61734344 (p.Met72Thr) and rs367716858 (5'UTR) showed significant association with infertility (p-values: 8.46e-06 and 0.001, respectively). Biophysical studies showed that mutation p.Met72Thr alters calcium binding affinity, surface hydrophobicity, calcium induced changes in Centrin-1. Somatic cells (HEK293T and RPE) and differentiated iPSC over-expressing mutant Centrin-1 (p.Met72Thr) showed higher multipolar cells, fragmented nuclei and cell death. Studies using RPE cell lines and germ cell lines (derived from iPSC) showed that the mutation p.Met72Thr affects ciliary assembly dynamics and cell proliferation, respectively. The 5'UTR mutant expressing cells showed 3-fold higher luciferase activity compared to wild type (p-value < 0.05), due to loss of a potential methylation site. Thus, rs61734344 causes defective cell division and rs367716858 leads to dysregulation of *CETN1* expression, affecting spermatogenesis and male fertility. This is the first study to demonstrate the role of *CETN1* gene variants in human male infertility.

499W

Genotoxicity of interferon beta in pregnant women with multiple sclerosis. M. Bouhlel¹, M. Guirat¹, N. Gharbi¹, I. Benayed¹, S. Hammami², A. Dammak³, R. Frikha¹, F. Abdelhedi¹, S. Guidara¹, N. Belguith¹, H. Kamoun¹. 1) Genetics, Hospital Hedi Chaker, Sfax, Tunisia; 2) Pharmacology department, Faculty of medicine, Sfax, Tunisia; 3) Gynecology department, Hospital Hedi Chaker, Sfax, Tunisia.

Multiple sclerosis (MS) has a female preponderance and affects people during their reproductive years. Thus, evaluating the reproductive effects of interferon β , a commonly used therapy for MS, is clinically important. The literature review indicates that there is an increased risk of spontaneous abortion related to the use of INF β . However, the effects of this drug on fertility have not been well studied. So, we propose to evaluate the genotoxicity of INF β in a tunisian case. Our patient is a 36-years-old female who has been treated by Rebif for multiple sclerosis since 2010. She was referred for genetic counseling. A family history of multiple sclerosis in a sibling was reported. Married in 2016, she had 3 spontaneous abortions in the first trimester with a therapeutic window of one or two weeks prior to conception without B9 vitamin therapy. A standard karyotype on G-band blood lymphocytes was done. This test showed a structural chromosome abnormality that involves a translocation affecting chromosomes 10 and 14 in one mitose from 31 mitoses analysed (0.032%). Rebif, interferon β -1a, is a pregnancy category C drug. There is no evidence of infertility or teratogenicity with the INF β s; however, exposure during the first trimester was reported to cause a dose-dependent abortive effect in monkeys. Our patient had 3 spontaneous abortions while she was under treatment. Her karyotype showed a structural abnormality that may be due to the effect of Rebif. Her brother who has also MS treated with INF β married, his wife had two spontaneous abortions. To support the hypothesis of the genotoxicity of this drug, we propose to complete with a Sister chromatids exchange test.

500T

Toll-like receptor 1 (TLR1) is key gene in TLR6-TLR1-TLR10 gene cluster associated with preterm birth in Wisconsin cohort. *D.M. Pillers¹, C. Hoffman², T. Ranade², J. Eickhoff², B.R. Pattnaik², M. Baker², S.A. Tokarz², W. Luo¹.* 1) University of Illinois, Chicago, IL; 2) University of Wisconsin, Madison, WI.

Preterm birth (PTB) occurs in 11% of US births, causing significant morbidity and mortality. A subset of PTB is complicated by infection and potential aberrations in the microbiome. The innate immune system via the Toll-like receptors (TLR) is associated with susceptibility to infectious diseases. We speculated that the SNPs in fetal TLR genes may be associated with PTB. We have previously reported results for SNPs in three TLRs (1, 2, and 4) in Wisconsin infants, with only TLR1 SNP (N248S) and SNP (H305L) associated with preterm birth. TLR1 is part of the TLR6-TLR1-TLR10 gene cluster reported to be associated with inflammation and infection. We hypothesized that TLR1 SNP rs5743618 (S602I) would similarly mark an increased risk for PTB. In addition, we sought to determine whether SNP rs5743810 (P249S) in TLR6 or SNP rs11096957 (N241H) in TLR10 in the TLR6-TLR1-TLR10 cluster are also associated with PTB. Residual newborn screening DNA was tested with IRB approval. 3000 samples were used for TLR1 SNP (N248S) analysis, with 1616, 1813, 866, and 581 samples adequate for subsequent TLR1 (H305L), TLR1 (S602I), TLR6 (P249S) and TLR10 (N241H) testing. SNP genotype was by TaqMan SNP Genotyping Assay (Lifetechnologies, CA). Ancestral background was by maternal declaration. The TLR1 (S602I) minor allele T was associated with PTB: GA < 34 wks, Odds Ratio (OR) 1.72 (P < 0.0001); GA < 37 wks, OR 1.34 (P = 0.0004). The minor allele combined effects of TLR1 305 (T) and TLR1 248 (C) showed a significant association with PTB (GA < 34 and < 28 wks) (P < 0.00004), as well as low birth weight (< 1250g) (P < 0.0000015) and were most prominent in the Wisconsin Black cohort (GA < 34 wks; OR for Whites is 1.81 vs. 3.27 for Blacks). The TLR6 SNP (P249S) minor allele T was associated with preterm GA < 34 wks, OR 1.41 (P < 0.00087) and with GA < 37 wks, OR 1.30 (P < 0.032). However, in GA < 28 wks, the TLR6 SNP (P249S) minor allele T was not consistent with those in GA < 37 wks and GA < 34 wks. TLR 10 SNP (N241H) was not associated with PTB. In the TLR6-TLR1-TLR10 gene cluster, TLR1 SNPs in discrete locations were consistently associated with PTB in a Wisconsin cohort. TLR1 is part of a tightly linked cluster of TLR genes that act as co-receptors with TLR2, a key step in the inflammation cascade. Aberrant TLR1 may lead to PTB via an altered downstream pathway causing disturbed commensal homeostasis leading to an altered microbiome and finally, to pathogen invasion.

501F

Real-time variant interpretation for full-exon sequenced expanded carrier screening panels. *C. Lo, K. Kaseniit, M. Rasmussen, D. Muzzey.* Counsyl, South San Francisco, CA.

INTRODUCTION: Next-generation sequencing (NGS) of entire exons in disease-relevant genes is now common in clinical genomic tests. When such sequencing identifies novel variants, they must be classified (e.g., as benign, pathogenic, or unknown) in a timely manner in order for the test results to be clinically meaningful to providers and their patients. Because variant classification is a time-sensitive and resource-intensive process, it is important to understand the feasibility, dynamics, and operator-dependence of performing reliable interpretation in response to changes in test volume and panel size. **METHODS:** We analyzed the empirical variant-interpretation volume, duration, operator, and method (i.e., manual or automatic) in a clinical laboratory running a 176-condition NGS-based expanded carrier screening panel. We evaluated hypotheses to explain the observed trends using simulations leveraging data in public repositories (e.g., allele frequency in gnomAD). **RESULTS:** The rate at which novel alleles are detected depended on the number of screened patients and the size of the genomic region sequenced. We compiled the empirical distribution of novel-allele detection per condition as a function of sequenced patients and found that, for most conditions, novel-allele detection reached steady state before 20,000 patients had been tested. The trajectory is comparable to simulated cohorts of variable size using condition-specific allele frequency data from gnomAD. For 94.7% of variants, interpretation according to ACMG guidelines can be performed automatically based on allele frequency and other in-silico methods (e.g., conservation). Manual consideration is needed for the remaining variants. Based on the self-reported estimates of variant-interpretation operators, the median time required to manually curate a variant was 42 minutes (average was 67.8 minutes). Permutation testing revealed low classification bias among operators. **CONCLUSION:** We demonstrate how real-time variant interpretation scales as a function of ECS panel content and the number of patients tested. Performing this characterization is important to maintain prompt reporting of results and proper downstream patient care, especially as panel content and test volume grows. .

502W

Clinical validity of expanded carrier screening: High concordance of inter-lab variant classifications. D. Muzzey, E. Collins, K.E. Kaseniit, C. Lo, K. Moyer, H.P. Kang. Counsyl, South San Francisco, CA.

Introduction Consensus on the classifications of individual variants among submitters to databases such as ClinVar is often used to assess the clinical validity of a genetic test. The ACMG/AMP variant classification guidelines recommend requiring more evidence for pathogenic assertions in an unaffected population, e.g., in expanded carrier screening (ECS), where most patients are asymptomatic. We evaluated the clinical validity of ECS by analyzing variant-classification concordance between Counsyl's Foresight ECS and ClinVar submissions. **Methods** Concordance was assessed between ClinVar entries (April 2018; ignoring old entries, submitters with few entries, and common variants) and Counsyl interpretations, and discordances were categorized by their underlying cause. To estimate the worst-case scenario for Foresight performance, disease-level positive predictive value (PPV) and sensitivity were calculated from carrier rates of different variant categories assuming ClinVar assertions are correct, e.g., "false positives" arose from variants considered pathogenic by Counsyl but not in ClinVar. **Results** Out of 12,834 variants evaluated, 98.2% were concordant. Most discordances (76.8%) were due to pathogenic assertions in ClinVar that we considered not to be reportable (VUS or benign). Many discordances had a clear explanation: 25.7% were due to unclear submitter classifications, 14.8% were variants with no published cases, 3.4% were variants with homozygotes observed in the population, and 4.6% were due to categories where reporting of variants in a carrier-screening setting might not be appropriate compared to a diagnostic setting. The remaining 51.5% of discordances could not be clearly categorized. Overall, 31.7% of Foresight patients are carriers for concordant alleles, and only 0.9% carry uncategorized discordant alleles (0.7% carry variants that do not meet our ECS threshold for pathogenicity, while 0.2% carry variants that do). Importantly, our concordance data suggest a level of clinical validity for ECS diseases that yields high estimated clinical sensitivity and PPV. For example, the estimated clinical sensitivity for cystic fibrosis was 99.7% (PPV: 99.8%), comparable to levels reported for hereditary cancer screening. **Conclusions** With a proper variant-interpretation workflow, ECS has high clinical validity. Residual discordances relative to ClinVar may be unavoidable because submitters sample from different patient populations (e.g., affected vs. unaffected).

503T

Genome-wide association study of a refined sub-phenotype of pregnancy duration: Pre-labor rupture of membranes. P. Sole-Navais¹, J. Juodakis¹, H. de Haan², T. Laisk³, R. Mägi³, G. Zhang⁴, L.J. Muglia⁴, J. Bacelis¹, B. Jacobsson^{1,2}. 1) Department of Obstetrics and Gynecology, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden; 2) Division of Health Data and Digitalization, Department of Genetics and Bioinformatics, Norwegian Institute of Public Health, Oslo, Norway; 3) Estonian Genome Center, University of Tartu, Tartu, Estonia; 4) Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA.

While pregnancy duration is a heterogeneous phenotype, where different onsets of delivery may reflect different etiologies, most genetic studies have treated gestational age (GA) as a single entity. Furthermore, GA is a time-to-event phenotype, and should be analyzed with statistical techniques developed for this purpose (survival models). Because of these analytical simplifications, the only successful attempts of identifying genetic components of GA so far required very large sample sizes. Here, we aimed to discover genetic variants associated with pre-labor rupture of membranes (PROM), a sub-phenotype of GA affecting 10-15% of all pregnancies, by using survival models. HARVEST (discovery sample) is a Norwegian genotyping (Human-CoreExome chip) effort on 11000 family trios from the MoBa Cohort. Only live-born singleton pregnancies were genotyped and imputation of non-genotyped variants was performed against the reference haplotypes of the HRC (release 1.1). Samples with major ethnicity other than CEU or cryptic relatedness to other samples were excluded. Association analysis was restricted to pregnancies without risk factors of PROM and to genetic variants with $MAF > 0.01$. We conducted a genome-wide analysis to test for the association of ~10M genetic variants with time-to-PROM on 9196 unrelated women (1117 PROM events). Time-to-event (GA) was estimated by ultrasound or last menstrual period. The effect of each SNP on time to event was estimated by Cox regression (implemented in ProbABEL). We used the Estonian Biobank ($n = 21956$, PROM cases = 936) to test for replication of candidate from the discovery cohort loci with genome-wide or suggestive ($p < 1e-5$) evidence of association with PROM by using logistic regression. We identified two variants (rs11978196 and rs6463018) in perfect LD showing genome-wide significant association with PROM near *INHBA* gene (hazard ratio: 1.65, 95% CI: 1.38-1.96, $p = 2.4 \times 1e-9$, $MAF = 0.10$). Previous studies suggest a role for inhibin in both preterm delivery and preterm PROM. In the Estonian Biobank, none of the discovery-stage associations were considered significant (most extreme $p = 0.04$, 70 variants tested), but top variants effect direction was identical to that in the discovery-stage. Our genome-wide survival analysis of PROM as a sub-phenotype of GA identified a suggestive locus previously implicated in preterm PROM pathophysiology, but warrants follow-up in large independent replication cohorts with a well-characterized phenotype.

504F

Revealing aging-related transcriptome and methylome dynamics in mouse oocytes development by single-cell parallel sequencing. *TL. Lee¹, Y. Qian¹, DYL. Chan², MP. Zhao³, KYL. Yip³, Q. Chao³, KL. Chow⁴, TY. Leung², NLS. Tang⁴, JY. Liao¹, ACS. Luk¹, WT. Lee¹, TC. Li².* 1) School of Biomedical Sciences, The Chinese University of Hong Kong, Shatin NT, Hong Kong, China; 2) Department of Obstetrics and Gynaecology, The Chinese University of Hong Kong, Shatin NT, Hong Kong, China; 3) Department of Computer Science and Engineering, The Chinese University of Hong Kong, Shatin NT, Hong Kong, China; 4) Department of Chemical Pathology, The Chinese University of Hong Kong, Shatin NT, Hong Kong, China; 5) Division of Life Science, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, China.

The decline in oocyte quality at advanced maternal age (AMA) is a major factor of female infertility. As such, identifying AMA oocytes with good developmental competence is critical for *in vitro* fertilization (IVF) success. Currently, oocyte quality measurement largely relies on morphological assessment. However, the method is subject to personal bias and the diverse approaches adopted in different clinics. As a result, the assessment failed to accurately define the respective developmental competence. To establish a foundation for precise gene-based assessment, we aimed to dissect the genomic landscapes of oocyte maturation and aging by applying an unbiased single-cell genomics approach. Mouse oocytes at both germinal vesicle (GV) stage and metaphases II (MII) stage from young female mice (6 weeks) and aged female mice (12 months) were collected. Single-cell parallel methylome and transcriptome sequencing (scM&T-seq) was then performed on individual oocytes. Overall, we identified 1377 and 1432 aging-related differentially expressed genes (DEGs) in GV and MII stage respectively. Gene ontology and KEGG pathways analysis showed that DEGs in GV stage were enriched in mitochondrial function, cell cycle process, and RNA splicing. In contrast, DEGs in MII stage were associated with microtubule and spindle formation. Repression of novel gene targets like maturation-specific oocyte-secreted factors and interleukins in both GV and MII stages were identified. Interestingly, expression of repetitive elements (RE) such as long terminal repeat (LTR) family was not affected by age, which is to our surprise as retrotransposition of RE was suggested to be activated during normal aging. Similar to transcriptome, the methylome exhibited dynamic changes in CpG sites and non-CpG sites. Global methylation level of non-CpG sites increased in GV stage but decreased in MII stage during oocyte aging. It was consistent with the expression dynamics of Dnmt3s, especially in GV stage. Methylation level of CpG sites declined during GV oocyte aging but increased in MII stage. Besides, global methylation level of non-CpG sites increased during maturation. In summary, we successfully revealed the transcriptome and epigenetic landscapes in mouse oocyte development and identified novel targets that may reflect oocyte quality. We suggest the use of high-resolution genomic data, along with morphological evidence, would improve IVF outcomes.

505W

Unlocking the mysteries of preterm birth: Meta-analysis of maternal and fetal transcriptomic data elucidates the role of adaptive and innate immunity. *B. Vora, A. Wang, I. Kost, H. Huang, I. Paranjpe, T. Woodruff, T. MacKenzie, M. Sirota.* University of California, San Francisco, San Francisco, CA.

Preterm birth (PTB) is the leading cause of newborn deaths around the world. Spontaneous preterm birth (sPTB) accounts for two-thirds of all PTBs; however, there remains an unmet need of detecting and preventing sPTB. Although the dysregulation of the immune system has been implicated in various studies, small sizes and irreproducibility of results have limited identification of its role. Here, we present a cross-study meta-analysis to evaluate genome-wide differential gene expression signals in sPTB. A comprehensive search of the NIH genomic database for studies related to sPTB with maternal whole blood samples resulted in data from three separate studies consisting of 339 samples. After aggregating and normalizing these transcriptomic datasets and performing a meta-analysis, we identified 210 genes that were differentially expressed in sPTB relative to term birth. These genes were enriched in immune-related pathways, showing upregulation of innate immunity and downregulation of adaptive immunity in women who delivered preterm. An additional analysis found several of these differentially expressed at mid-gestation, suggesting their potential to be clinically relevant biomarkers. Furthermore, a complementary analysis identified 473 genes differentially expressed in preterm cord blood samples. However, these genes demonstrated downregulation of the innate immune system, a stark contrast to findings using maternal blood samples. These immune-related findings were further confirmed by cell deconvolution as well as upstream transcription and cytokine regulation analyses. Overall, this study identified a strong immune signature related to sPTB as well as several potential biomarkers that could be translated to clinical use.

506T

Parent-of-origin determined allelic expression of imprinting candidate genes in the human placenta. *M. Laan, D. Pilvar, M. Reiman, A. Pilvar.* Institute of Biomedicine and Translational Medicine, University of Tartu, Tartu, Estonia.

Imprinting adjusts the gene dosage via parent-of-origin specific silencing of a gene copy. The emergence of imprinting coincides with the evolution of placenta and seminal discoveries of imprinted loci reported genes critical for placental and fetal development. Recently, the analysis of human imprinted genes has been facilitated by RNA-Seq and/or allele-specific WGBS applications. However, the published data has reached inconsistent claims and family-based studies are limited. We aimed to analyse parent-of-origin of placental transcripts of nearly 400 imprinting candidate genes retrieved from the GeneImprint database and recent literature. In total, 54 families (38 parental-placental trios, 16 maternal-placental duos) representing normal 1st, 2nd and term pregnancies as well as pregnancy complications were included. The study utilized our published genome-wide genotyping data of trios/duos (Kasak et al 2015, 2017 Sci Rep) and the placental RNA-Seq dataset of the same cases (Sõber et al 2015, 2016 Sci Rep). The data shows that nearly half of the imprinting candidate genes have no or minimal transcription in the human placenta. Among the genes with adequate read counts for the assessment of parental allelic expression, the vast majority of the analysed genes showed convincingly biallelic origin of the transcripts. Only a small fraction of genes exhibited clear signals of imprinting, i.e. nearly monoallelic parent-of-origin specific gene expression. These loci represent previously well-established placenta-enhanced imprinted genes (e.g. *IGF2*, *H19*, *DLK1*, *MEG3* etc). Imprinting status of these genes was conserved across gestation, and was unaffected by the occurrence of pregnancy complications such as preeclampsia. In addition, some loci demonstrated parent-of-origin biased expression, but not complete silencing of one parental allele (>65%, but <90% of reads from one parental copy). This group of genes is expressed in a broad range of human tissues and showed rather variable proportions of maternal/paternal reads across gestation and among individual placentas. As these loci were mostly biased towards paternal expression, it could be explained by incomplete demethylation of maternal and/or remethylation of paternal alleles during epigenetic reprogramming in the trophoblast lineage. This hypothesis is also consistent with polymorphic imprinted methylation that has been shown for the human placental genome. Grant support: Estonian Research Council IUT34-12.

507F

Serotonin transporter (5-HTTLPR) genotypes and trinucleotide repeats of androgen receptor exert a combinatorial effect on hormonal milieu in patients with lifelong premature ejaculation. *S. Bhatti^{1,2}, H. Latif Khan¹, S. Abbas¹, Y. Latif Khan¹, R. Gonzalez³, M. Aslamkhan³, G. Gonzalez⁴, H. Hakan Aydin⁵.* 1) Department of Human genetics, Lahore Institute of Fertility and Endocrinology, Lahore, Punjab, Pakistan; 2) Department of Human Genetics and Molecular biology, University of Health Sciences Lahore-54600, Pakistan; 3) Centro de investigacion Biomedica de Occidente, IMSS, Universidad de Guadalajara, Jalisco Mexico; 4) Universidad De Guadalajara CIBO, IMSS Sierra Mojada 800 Independencia, 44340 Guadalajara, Jalisco Mexico; 5) Department of Medical Biochemistry, Ege University School of Medicine, Bornova Izmir, Turkey.

Premature ejaculation (PE) is one of the most common sexual disorders in men due to uncontrolled modulation of spinal reflexes which are controlled by cortico-limbic centers in the brain. In the present study, we investigate the combinatorial effects of tri-nucleotide repeats of androgen receptor and allelic variants of the 5-HTTLPR gene on sex steroids, hypophyseal hormones, sexual performance, and premature ejaculation assessment parameters among evidence-based lifelong PE subjects. A total of 571 outpatients (age: 26.6 ± 1.9) consulting for evidence-based lifelong premature ejaculatory dysfunction (LPE) were selected in this study. The control group consists of 255 men with normal IELT (>4 minutes). The study revealed that the subjects who have the highest (≥ 26) CAG stretches depicted a significantly high serum oxytocin levels (102.1 pg/ml; $n = 126$, $p < 0.001$) compared with the patients with the medium (22-25) and short (≤ 21) CAG stretches (76.63 ng/ml; $n = 64$, $p < 0.001$ versus 77.4 ng/ml; $n = 81$, $p < 0.001$) and the control group (71.2 pg/ml; $n = 75$, $p = < 0.001$). Almost 33 (26.1%) LPE patients had AR variant of longer (≥ 26) CAG repeats were homozygous for S alleles (SS), 45 (35.7%) was homozygous for L allele (LL), and 48 (38%) had the L/S or S/L genotype of 5-HTTLPR gene. Homozygous (SS) alleles have a significant positive correlation ($r = 0.44$, $p < 0.0001$) with the high score of BDI-II (39.1, $n = 126$, $p < 0.001$). However, LL alleles have shown a significant positive correlation with PEDT ($r = 0.46$, $p < 0.001$) and negative correlation with self-estimated IELT and intercourse satisfaction ($r = -0.35$, $p < 0.001$). The innovative study design elaborates that Androgen receptor trinucleotide repeats and 5-HTTLPR genotypes have a combinatorial impact on hormonal milieu and sexual function regarding evidence-based lifelong premature ejaculatory dysfunction patients.

508W

FANCM biallelic mutations cause non-obstructive azoospermia. L.

Kasak¹, M. Punab², L. Nagiraja³, M. Grigorova⁴, A. Minajeva¹, A.M. Lopes^{4,5}, A.M. Punab⁶, K.I. Aston⁶, F. Carvalho^{6,7}, E. Laasik¹, L.B. Smith^{8,9}, D.F. Conrad^{8,10}, M. Laan¹, GEMINI consortium. 1) Institute of Biomedicine and Translational Medicine, University of Tartu, Tartu, Estonia; 2) Andrology Center, Tartu University Hospital, Tartu, Estonia; 3) Department of Genetics, Washington University School of Medicine, St. Louis, Missouri, USA; 4) IPATIMUP - Institute of Molecular Pathology and Immunology of the University of Porto, Porto, Portugal; 5) i3S - Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal; 6) Division of Urology, Department of Surgery, University of Utah School of Medicine, Salt Lake City, Utah, USA; 7) Serviço de Genética, Departamento de Patologia, Faculdade de Medicina da Universidade do Porto, Porto, Portugal; 8) MRC Centre for Reproductive Health, University of Edinburgh, Edinburgh, UK; 9) School of Environmental and Life Sciences, University of Newcastle, Callaghan, Australia; 10) Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, Missouri, USA.

Infertility is a highly complex and multifactorial condition affecting around 7% of men worldwide. Unexplained non-obstructive azoospermia (NOA) is defined as the absence of spermatozoa in the ejaculate due to failure of spermatogenesis. Around a quarter of azoospermia cases can be explained by genetic causes, but a large number still remains idiopathic. Whole-exome sequencing (WES) in two brothers diagnosed with NOA and Sertoli cell-only syndrome (SCOS) led to the discovery of compound heterozygous loss-of-function (LoF) variants in the *FANCM* (Fanconi anemia complementation group M) gene as the most likely cause for their condition. A previously undescribed paternally inherited splicing variant and a rare frameshift variant derived from the mother introduce a premature STOP codon leading to a truncated protein missing critical functional domains. *FANCM* shows enhanced testicular expression and is involved in maintaining genomic stability in meiosis and mitosis. In controls with normal spermatogenesis, immunohistochemical staining localized *FANCM* to the Sertoli and spermatogenic cells of seminiferous tubules with gradually enhanced intensity along germ cell maturation stages. In the SCOS case carrying biallelic *FANCM* LoF variants, none or only faint expression was detected in the Sertoli cells. Identification of two further unrelated NOA cases with independent *FANCM* homozygous nonsense variants omitting the protein domain for DNA repair activity provided additional support to the link between *FANCM* LoF variants and male infertility. Our study demonstrates that biallelic recessive *FANCM* LoF variants cause azoospermia. *FANCM* pathogenic variants show also implications for somatic health, being linked with doubled risk of familial breast and ovarian cancer. This provides a possible mechanism for the known epidemiological association between infertility and cancer risk, supported by published data on *Fancm*-mutant mouse models. Grant support: Estonian Research Council IUT34-12.

509T

Identifying genetic factors that contribute to human female infertility.

K.M. Tyc¹, W. El Yakoubi¹, X. Tao², J. Landis³, Y. Zhan², N. Treff³, R.T.Jr. Scott³, K. Schindler¹, J. Xing¹. 1) Department of Genetics, Human Genetics Institute of New Jersey, Rutgers, the State University of New Jersey, 145 Bevier Rd. Piscataway, NJ 08854, USA; 2) Foundation for Embryonic Competence, 140 Allen Rd, Basking Ridge, NJ 07920, USA; 3) Reproductive Medicine Associates of New Jersey, 140 Allen Rd, Basking Ridge, NJ 07920, USA.

Infertility is a common health issue that affects about 10% of women. Maternal age is an established contributing factor that correlates with an increased production of aneuploid eggs—a major cause of early miscarriage and infertility. Nevertheless, clinical data report cases of reproductively young (<35y) women producing unusually high percent of aneuploid concepti, and women of advanced maternal age (>35y) producing lower than average percent of aneuploid concepti for their respective age categories. This observation suggests genetic factors contribute to fertility decline. Understanding the contribution of genetics in female infertility is the first step towards a genetic-based approach to reach a clinical efficacy in treating infertility. The scope of this work is to identify mutations that could be used as predictors of egg quality which would help in advancing assisted reproduction. Moreover, identification of key mutations and genes contributing to the risk of aneuploidy will allow prescreening patients that have better chances to benefit from preimplantation genetic screening. We obtained whole-exome sequencing data from over 500 reproductively challenged women, and performed a search for functional mutations that cause gametogenesis defects. Thus far we identified several candidate genes that could be linked to processes important for proper chromosome packaging and/or cell division. For instance, we identified nuclear pore components mutated in women with high rate of aneuploid concepti. Studies in lower eukaryotes show that nucleoporins are essential for timely nuclear import of cytosolic proteins involved in centromere formation before the completion of meiosis I. Moreover, we identified a condensin gene whose product is an important player in compacting the genetic material before meiotic divisions. Thus far, we established a list of mutations likely contributing to the production of poor quality eggs. Using a mouse model for assessing meiotic functions, we are currently validating the impact of the identified mutations on regulating chromosome segregation in oocytes.

510F

Clinical impact and cost effectiveness of a 176 condition expanded carrier screen. K.A. Beauchamp, K.K. Wong, K.J. Taber, G. Lazarin, E.A. Evans, J.D. Goldberg, D. Muzzey. Counsyl, South San Francisco, CA.

IntroductionCarrier screening identifies couples at high risk for conceiving offspring affected with severe heritable conditions. Minimal screening guidelines mandate testing for two diseases (cystic fibrosis and spinal muscular atrophy), but expanded carrier screening (ECS) assesses reproductive risk for hundreds of conditions simultaneously. Although key medical societies consider ECS an acceptable practice, and hundreds of thousands of patients have undergone ECS testing, the health economics of ECS remain incompletely understood. **Methods**The clinical impact and cost effectiveness of a 176-condition ECS panel were investigated using a decision-tree model comparing minimal screening and ECS in a preconception setting. Carrier rates from >50,000 ECS patients informed disease-incidence estimates, while cost and life-years-lost data were aggregated from the literature and a cost-of-care database. Model robustness was evaluated using one-way and probabilistic sensitivity analyses. **Results**For every 100,000 pregnancies, approximately 300 are predicted to be affected by conditions on the ECS panel, which individually incur approximately \$1,000,000 in lifetime costs and increase mortality by approximately 30 undiscounted life years on average. Relative to minimal screening, ECS reduces the affected-birth rate by 66% and is cost-effective (i.e., <\$50,000 incremental cost per life year), findings that are robust to reasonable model-parameter perturbation. Increased cost-effectiveness is observed when the net impact of downstream costs is included (e.g., savings from averted disease and spending on in vitro fertilization). **Conclusions**ECS is predicted to substantially reduce the population burden of Mendelian disease in a manner that is cost-effective compared to many other common medical interventions.

511W

Embracing the challenges of isochromosomes in cell-free DNA prenatal screening. T. Boomer, E. Soster, J. Chibuk, S. Caldwell, E. Almasri, J. Wardrop, S. Boshes, M. Hackbardt, P. Cacheris, R. McCullough. Sequenom / Integrated Genetics, 3595 John Hopkins Court, San Diego, CA 92121.

Background: Cell-free DNA (cfDNA) prenatal screening continues to increase insight into early gestational placental findings not previously recognized or well understood. Genome-wide sequencing combined with advanced bioinformatics enables detailed interrogation of a variety of complex chromosomal changes, including isochromosome formation. Herein we review our laboratory experience with a variety of isochromosome rearrangements.

Methods: Maternal blood samples submitted for cfDNA screening were subjected to DNA extraction, library preparation, and whole-genome massively parallel sequencing as described by Jensen et al. Sequencing data were analyzed using a novel algorithm as described by Lefkowitz et al. **Results:** Over 60 cases involving isochromosomes were detailed in our clinical database, including: i(9) – 5 cases i(12) – 7 cases i(13) – 6 cases i(15) – 3 cases i(18) – 9 cases i(20) – 1 case i(21) – 9 cases i(22) – 2 cases i(X) – 8 cases i(Y) – 16 cases. Over half of these cases (57%) were correctly identified via cfDNA screening and confirmed as ‘true positives’, a subset (18%) pending outcome, and a minority (6%) were not confirmed via diagnostic testing and thus considered ‘false positives’. Interestingly and unique to this particular chromosomal subset are the 19% reported to have eluded cfDNA screening and reported as ‘false negatives’, consistent with four European / Australian laboratories’ cfDNA i(21) experience (ISPD 2017) and the chorionic villus sampling (CVS) literature. Mosaicism was commonly suggested and reported in cfDNA data, as well as diagnostic testing. Discrepancy for fetal sex was common among i(Y) cases. [Totals will be updated prior to final publication.] **Conclusion:** Isochromosomes arise from an error in centromere division during meiosis or mitosis. Most cases are *de novo* and mosaic in origin, consequently having variable impact on developing embryologic tissues. Since cfDNA screening is thought to primarily reflect early developing placental tissue (cytotrophoblast), results may be concordant or discordant with downstream diagnostic studies. It is essential for clinicians to remain mindful of the unique biologic limitation that isochromosomes pose for cfDNA screening. The contribution of cfDNA data can lend new and important insight into the timing and origin of such complex events, as well as help to explain recurrent and unavoidable inconsistencies between a variety of fetoplacental testing modalities.

512T

Trisomy 8 mosaicism detected by NIPT: A +30 week diagnostic odyssey for fetus, parents, and clinicians. J.L. Giordano¹, A. Thomas², C. Buchovecky², A. Tanaka², N. Hoffman², S. Caldwell³, E. Bonifero⁴, D. Lewis⁴, R. Wapner¹, B. Levy². 1) Dept of OBGYN-MFM, Columbia University, New York, NY; 2) Dept of Pathology and Cell Biology, Columbia University, New York, NY; 3) Sequenom Laboratories, 3595 John Hopkins Ct, San Diego, CA 92121; 4) Mount Sinai West 1000 10th Avenue, Suite 11A, NY NY 10019.

NIPT for common trisomies has become standard practice for AMA pts. Screening all aneuploidies is the next logical step in NIPT. However, the clinical utility remains unclear. Our 35 yr old pt consented to a whole genome NIPT at 10 weeks, which revealed an increased risk for Trisomy 8 (T8) mosaicism (20% z score 11.33 FF 10%). Her 1st trimester screening revealed normal NT (1.26 MoM) and analytes (hCG 1.15 MoM PAPP-A 2.40 MoM). The NIPT laboratory recommended an amnio follow up, which the patient pursued after genetic counseling. Amniocytes showed a normal chr 8 FISH (100 interphase cells), karyotype, and CMA. The fetal anatomy was nl. As T8 can elude detection in cultured amniocytes and the pt was highly motivated to investigate the etiology of the abn NIPT, the pt opted for a placental biopsy, repeat NIPT, and mat blood karyotype with FISH at 17w4d. The placental biopsy and mat blood karyotype were normal (620 & 270 cells counted respectively, direct FISH + 20 G-banded metaphase) and the repeat NIPT reaffirmed the finding of T8 mosaicism (18.8% z score 7.93 FF 8%) at 10 wks. This led to concern for occult maternal malignancy or subclinical T8 in an otherwise healthy individual. The pt had a 2nd blood karyotype (20 G-banded metaphases), FISH on buccal sample and skin biopsy, which were all nl (250 direct, 600 direct & 500 cultured cells by FISH, respectively). Given the association of T8 and hematological cancers, CBC w/ differential and flow cytometry were ordered and nl. The patient pursued cordocentesis and FISH analysis with a dual locus probe for chr 8, which revealed 5 of 900 cells with T8. The pt continued the pregnancy and delivered a full term healthy appearing female. The placenta was sectioned to 12 parts and screened using dual locus probes. 2,450 cells were scored, 32 cells positive for T8 (1.3% Average; Range: 0 – 5%). Cord blood with a dual locus probe found 1/100 (1%) had T8. Repeat NIPT was drawn 6 weeks postpartum and no evidence of T8 was seen, suggesting that the T8 cells were indeed placental in origin. Since adverse outcomes have been observed in the many rare autosomal trisomy cases followed up after NIPT, T8 mosaicism by NIPT could not be ignored. The extensive work up in this case proved the NIPT was an accurate reflection of placental mosaic aneuploidy, despite its absence in placental biopsy at 17wks. However, the findings did not affect pregnancy outcome and the NIPT proved to have low clinical utility.

513F

Real-time quantification of cell free fetal DNA (cffDNA) in maternal plasma and its applications for non-invasive prenatal diagnosis. D. Misra¹, R. Singh², S.R. Phadke³, N.S. Verma⁴, M. Banerjee¹. 1) Molecular & Human Genetics Laboratory, Department of Zoology, University of Lucknow, Lucknow-226007, Uttar Pradesh, India; 2) Department of Obstetrics and Gynecology, King George's Medical University, Lucknow-226003, Uttar Pradesh, India; 3) Department of Medical Genetics, Sanjay Gandhi Post Graduate Institute of Medical Sciences, Lucknow-226023, Uttar Pradesh, India; 4) Department of Physiology, King George's Medical University, Lucknow-226003, Uttar Pradesh, India.

Background: A step towards better management of healthy births lies in prenatal diagnosis of genetic disorders. The quantitative measurement technique (QMT) was used to analyze different quantities of cell free fetal DNA (cffDNA) in total plasma DNA (pDNA) using non-pregnant, healthy pregnant and high risk pregnant cases. **Purpose of the Study:** Presence of certain pregnancy related complications and fetal abnormalities manifest as changes in pDNA or cffDNA amounts in maternal plasma when compared to normal pregnancies. **Objective:** The aim was to quantitatively analyse cffDNA present in pDNA and its application as an early non-invasive prenatal diagnostic tool.

Method Used: The Circulating DNA from Plasma Kit was used to obtain pDNA. The primers used to quantitate pDNA were β -actin and GAPDH and Y-specific viz. DYS and SRY for cffDNA using SYBR Green I based assay in real-time quantitative PCR (qPCR). **Key Results:** During gestation, the average copy number of pDNA was 1.05×10^{10} and 1.04×10^9 using β -actin and GAPDH primers and the percentage of cffDNA was 2.87 and 28.94 respectively. A significant rise in concentration of pDNA in maternal circulation was observed in pregnancies suspected of Down syndrome (DS), preeclampsia (PE) and gestational diabetes mellitus (GDM). The amount of cffDNA showed highly significant (<0.0001) and significant ($P \leq 0.05$) rise in pregnancies suspected with PE and Duchenne muscular dystrophy (DMD) respectively. **Conclusion:** The pDNA and cffDNA concentrations could be a prenatal indicator of genetic abnormalities. The study indicates that quantitative measurement of pDNA/cffDNA in qPCR may be used as a screening marker for ascertaining pregnancy associated complications.

514W

Maternal fragile sites resulting in maternal mosaic deletions: An additional explanation for false positive cell-free DNA. V. Nitibhon, B. Dyer, T. Boomer, J. Chibuk. Sequenom Inc., Laboratory Corporation of America Holdings, San Diego, CA.

Objectives Genome wide cell-free DNA (cfDNA) allows for the potential to detect copy number variants due to maternal low level mosaicism leading to false positive screening results. As previously reported, deletions detected at 10q25-qter are associated with maternal FRA10B expansions. FRA10B is a known rare fragile site that is not associated with a clinical phenotype. Expanded FRA10B alleles have been noted to delay replication distal from the expansion and can predispose to deletions. This case series explores our experience with seven positive 10q25.2 deletion results and is consistent with recent literature as a possible cause of some false positive cfDNA results in that genomic region. **Methods** Maternal blood samples submitted to Sequenom Laboratories® for MaterniT® GENOME testing were subjected to DNA extraction, library preparation, and genome-wide massively parallel sequencing as described by Jensen et al. Sequencing data were analyzed using a novel algorithm to detect trisomies, select microdeletions, and genome wide events 7 Mb and larger, as described by Lefkowitz et al. Follow-up information and pregnancy outcomes were elicited from the clinicians as part of routine, ongoing laboratory protocol for positive cases. **Results** Seven MaterniT® GENOME samples returned a positive result for a 10q25.2 deletion, where the fragile site FRA10B is located. Among these, one was confirmed as a maternal mosaic deletion, three were false positive following prenatal diagnostic testing with no maternal testing performed, two were lost to follow up and one had a second cfDNA screening drawn confirming the same deletion but no diagnostic testing. For six of the cases no phenotypic abnormalities or adverse outcomes were reported. **Conclusions** These positive cfDNA results may be attributed to maternal mosaic deletions associated with expansion of the fragile site FRA10B. Although the sample size is limited, preliminary conclusions can be drawn from this cohort as it is reflective of literature previously published correlating positive genome wide cfDNA for 10q25.2 deletions with maternal low-level mosaic deletions and FRA10B expansions. When a maternal mosaic deletion is identified there is concern for repeat false positive cfDNA results in future pregnancies. The available data provides further evidence that maternal fragile sites might explain false positive cfDNA results.

515T

Assessing causality in associations between maternal adiposity and obstetric and perinatal outcomes: A Mendelian randomization study. M. Taylor, M.C. Borges¹, R.C. Richmond¹, R. McEachan², J. West², J. Wright², EGG Consortium. 1) MRC Integrative Epidemiology Unit, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom; 2) Bradford Institute for Health Research, University of Bradford, UK.

Background: Observational studies consistently suggest that maternal obesity is related to a number of adverse perinatal outcomes. However, which of these are causal or confounding is unclear. We used Mendelian randomization (MR), which uses genetic variants as instrumental variables (IV), to determine the causal effect of maternal BMI on obstetric and perinatal outcomes. **Methods:** Data from eight cohorts (ALSPAC, BiB, HAPO, EFSOCH, Generation R, GOYA, Project Viva and Gen3G) were used (N=3,554-19,244 depending on outcome). We performed multivariable regression between maternal BMI (pre-pregnancy or at the start of pregnancy) and 23 outcomes. We then combined genetic variants shown to be robustly associated with BMI in a weighted allele score and used this in IV analyses to obtain causal estimates (Wald ratios) of maternal BMI on each outcome. We used MR-Egger and other sensitivity methods to explore potential bias due to horizontal pleiotropy. When available, MR analyses for outcomes that could be influenced by offspring genetic predisposition to higher BMI were adjusted for offspring genotype. **Results:** In multivariable analysis, a 1 standard deviation (SD, ~4.95 kg/m²) higher maternal BMI was associated with a 0.12 SD (95%CI 0.11-0.14) higher offspring birthweight, and 1.68 (95%CI 1.60-1.77) odds ratio (OR) of hypertensive disorders of pregnancy (HDP). In MR analyses, a 1 SD higher maternal BMI was associated with a 0.19 SD (95%CI 0.01-0.37) higher offspring birthweight and 3.06 (95%CI 1.20-7.85) OR of HDP. We also found positive associations of maternal BMI with birth ponderal index and early membrane rupture in multivariable regression analyses. MR point estimates for effects of BMI on these outcomes were in the same direction and of a similar magnitude to multivariable results, but the 95% CIs were wide and included the null. Lastly, in multivariable regression analyses we found a positive association of BMI with gestational diabetes and caesarean section, which was not supported by MR results. Results from MR Egger regression and other sensitivity methods were generally consistent and did not suggest strong bias by pleiotropy. **Conclusions:** Our findings support a causal effect of increased maternal BMI on hypertensive disorders of pregnancy and offspring birthweight, suggest the same for birth ponderal index and early membrane rupture, but not for caesarean section and gestational diabetes.

516F

Genetically determined modulation of NLRP7-associated inflammatory response in the context of reproductive failure. U. Wysocka¹, A. Gach¹, A. Sakowicz², I. Pinkier¹, M. Rybak-Krzyszowska³, W. Alaszewski¹, L. Dudarewicz¹, M. Kozłowska¹, L. Jakubowski¹. 1) Department of Genetics, Polish Mother's Memorial Hospital Research Institute, Lodz, Poland; 2) Department of Medical Biotechnology, Medical University of Lodz, Poland; 3) Department of Obstetrics and Perinatology, Jagiellonian Medical University of Cracow, Poland.

Introduction Recurrent reproductive wastage is a global health issue affecting a significant number of women. Miscarriages occur in up to 30% of pregnancies. It is a severe problem for diagnostics and healthcare system. Repeatability of recurrent reproductive wastage in a certain number of couples shows that the phenomenon is not random and urges to define a cause. Genetic variation may have a high impact on reproductive failure and thus delineating of specific genetic factors is of great importance for genetic counselling. The c.1441G>A polymorphism (p.Ala481Thr) in the *NLRP7* (MIM 609661) gene has a proven functional character and its presence is directly linked to a decreased levels of cytokines IL1 β and TNF in the plasma. It is possible that proinflammatory cytokine secretion and as a consequence altered inflammatory response in patients with c.1441G>A variant of *NLRP7* gene can promote the occurrence of miscarriages. In the current study the relationship of recurrent reproductive wastage and non-synonymous *NLRP7* variant in Polish population have been investigated. **Materials and Methods** The study was carried out on the blood samples from women with a history of recurrent miscarriages (study group) and from women with unremarkable obstetric history (control group). Additionally, material from CVS (chorionic villus sampling) with confirmed fetal chromosomal aberration (study group) and normal fetal karyotype (control group) was investigated. Study of the c.1441G>A polymorphism in the *NLRP7* gene was performed by Sanger sequencing. The frequencies of genotypes were compared with using the Chi-square test and the Statistica 10 software. **Results** The study was conducted in a group of 369 patients with reproductive failure, 178 patients from the control group and in 73 trophoblast samples (41 from the study group and 32 from the control group). The polymorphic variant (GA and AA genotypes) was demonstrated in 118 tested blood samples (82 from the study group and 36 from the control group) and in 16 trophoblast samples (8 in the study group and 8 in the control group). **Conclusions** Obtained results suggest that the link between reproductive failure and non-synonymous variant of *NLRP7* gene in Polish population is not statistically significant because the c.1441G>A polymorphism in the *NLRP7* gene appears with the similar frequency in the general population. To explore consequences of genetic variation in recurrent reproductive wastage further analysis should be carried out.

517W

Embryo sample tracking and identification during PGT using the mitochondrial genome. M.J. Jasper¹, M. Barry², S.A. Myers¹. 1) RHS Ltd, Thebarton, South Australia, Australia; 2) FertilitySA, Adelaide, South Australia, Australia.

The mitochondrial genome is maternally inherited and contains single nucleotide variants (SNVs) that can be used to differentiate individuals. This combination of characteristics provides a novel opportunity for DNA-based confirmation of maternal origin of embryo biopsies, and sibling embryo identification. Mitochondrial DNA (mtDNA) is sequenced during Preimplantation Genetic Testing for Aneuploidy (PGT-A) by Next Generation Sequencing (NGS), and the depth and breadth of coverage of the mitochondrial genome obtained from PG-Seq™ readily allows SNV analysis, even from a 48 sample NGS run. Our aim was to demonstrate the use of the RHS Embryo ID panel for accurate and economical embryo identification as part of routine PGT-A using the PG-Seq™ kit. A large putative panel of mtDNA SNVs was collated from published literature and PG-Seq™ data. SNVs associated with disease-related markers or in regions with low depth of coverage were excluded. A panel comprising 48 SNVs was compiled for evaluation. Three PG-Seq™ kit run scenarios were modelled, wherein 1, 2 or 4 sibling embryos were analysed per individual. The performance of the panel in each scenario was tested across 10,000 in-silico PG-Seq™ runs using randomly selected mtDNA genomes from (1) a globally-diverse database of 377 individuals and (2) embryo mtDNA genomes from 52 individuals. For analysis of the embryo dataset, a population-specific panel comprising another 23 SNVs was added. Using the global mtDNA genome database and modelling embryos from 12, 24 or 48 individuals, on average the SNV panel differentiated 91.5%, 84.6% and 75.0% of embryos, respectively. For the embryo PG-Seq™ NGS files analysed using the RHS Embryo ID SNV panel and an additional population-specific SNV panel, with embryos from 12, 24 or 48 individuals, on average 99.3%, 98.3% and 96.3% of embryos were differentiated, respectively. If the panel failed to differentiate individuals, other regions of mtDNA could be analysed to achieve differentiation. It should be noted that maternally-related individuals will share mitochondrial genomes. In this case, it will not be possible to distinguish between embryos from these individuals using mtDNA. PG-Seq™ and the RHS Embryo ID panel readily generates a unique embryo signature, potentially improving PGT-A practice by providing a DNA-based confirmation of maternal origin and sibling embryo identification.

518T

A novel single tube approach for combined PGT-A and PGT-M using Ion Torrent™ NGS. S.A. Myers, M.J. Jasper. RHS Ltd, Thebarton, South Australia, Australia.

Preparing embryo biopsies for Next Generation Sequencing (NGS) for Preimplantation Genetic Testing for Aneuploidy (PGT-A) typically requires a laborious, multi-step process wherein Whole Genome Amplification (WGA) and library preparation are performed separately. The DOPlify® kit (RHS Ltd) provides a flexible technology to not only amplify whole genomes but also target sequences for PGT for Monogenic disorders (PGT-M) using RHS' Target Sequence Enrichment (TSE) protocol. Furthermore, the attributes of the DOPlify® kit also present a unique mechanism to incorporate the specific sequences needed for NGS, bypassing the need for a separate library preparation step. PCR barcoding during WGA provides laboratory efficiencies compared to typical workflows, including a reduction of hands on and total protocol time, and decreased reagent requirements for sample preparation. Our aim was to develop a novel approach that allows combined amplification and PCR barcoding of low template samples in a single tube for Ion Torrent™ NGS. Five-cell samples were manually sorted from aneuploid cell lines (Coriell Institute) and prepared for sequencing using a two-stage, single tube protocol. DNA was amplified using standard DOPlify® kit reagents with a modified primer (RHS Ltd). Subsequently, publicly available Ion Torrent™ (Thermo Fisher) NGS adapter sequences and barcodes were incorporated utilising a second PCR step within the same PCR tube. The barcoded samples were pooled and sequenced on an Ion Torrent™ instrument. The sequencing data was aligned to hg19, sequencing metrics collated and the data analysed to determine sample ploidy status. Whole chromosome aneuploidy and sub-chromosomal 7-32 Mb duplication and deletion results were concordant with the expected karyotypes of the cell lines, and TSE using haemoglobin subunit beta gene-specific PCR primers in PGT-M for β -Thalassemia was achieved. The protocol produced sequencing ready samples within 3 hours, including clean-up time. Leveraging the unique characteristics of the DOPlify® kit this novel method provides a single tube amplification and barcoding protocol for Ion Torrent™ NGS to allow rapid, scalable and economical sequencing for PGT-A, and the incorporation of RHS' TSE protocol for combined PGT-M and PGT-A from a single embryo biopsy.

519F

Generation and validation of reference standards for non-invasive prenatal testing (NIPT). A. Mele¹, E. Surmann¹, R. Santos¹, D. Anderson¹, J. Wickenden¹, D. Pasupathy², B. Thilaganathan³, A. Khalil³. 1) Horizon Discovery Ltd, 8100 Beach Drive, Cambridge Research Park, Waterbeach, Cambridge CB25 9TL, United Kingdom; 2) St Thomas Hospital, Department of Women and Children's Health, School of Life Course Sciences, 10th Floor North Wing, London SE1 7EH, United Kingdom; 3) St George's Hospital, Fetal Medicine Unit, Department of Obstetrics and Gynaecology, St George's University of London, 4th floor Lanesborough Wing, London SW17 0QT, United Kingdom.

Non-invasive prenatal testing (NIPT) has been widely adopted in clinical practice as a screening tool for fetal chromosomal abnormalities. However, there is little regulatory oversight on test performance underlining the need for appropriate reference standards that accurately mimic the complexity of NIPT samples to establish assay sensitivity, specificity and reproducibility. Here we describe our unique approach on generating NIPT reference materials from clinical fetal tissues with or without chromosomal aneuploidies, as well as from corresponding maternal and paternal samples. We have generated immortalized cell lines that have been extensively characterized by karyotyping and quantitative polymerase chain reaction (qPCR) to confirm genomic stability and retainment of the additional chromosome over several cell culture passages. We show that fragmentation of maternal and fetal DNA to 160 base pairs and subsequent blending into buffer or synthetic plasma at different fetal fractions can closely mimic the size distribution profile of clinical cell-free DNA (cfDNA). This reference material demonstrated superior performance across various NIPT platforms in our beta testing study. We are the first to generate commutable NIPT reference standards containing defined fractions of fetal DNA in a matched maternal DNA background that can be used by platform developers and laboratories to validate and routinely monitor assay performance. The addition of the corresponding paternal cell lines will allow us to further extend our material in the future by introducing mutations that mimic single gene disorders inherited by the father. .

520W

Whole genome amplification for STR profiling on single cells in cell-based non-invasive prenatal testing. A. S. Vander Plaetsen¹, J. Weymaere¹, S. Cornelis^{1,2}, D. Deforce¹, F. Van Nieuwerburgh¹. 1) Faculty of Pharmaceutical Sciences, Ghent University, Ghent, Belgium; 2) Department of Life Science Technologies, imec, Leuven, Belgium.

Novel emerging isolation techniques and the rapidly evolving field of next generation sequencing have invigorated the interest in single cell analysis within cell-based non-invasive prenatal testing (cbNIPT). Although the presence of fetal cells in peripheral maternal blood has been recognized for a few decades, cbNIPT is still very challenging due to the extreme rarity and fragility of these fetal cells in maternal blood. Furthermore, the absence of a unique fetal biomarker necessitates fetal identity confirmation of each isolated cell before further downstream genetic analyses can occur. To allow both fetal identification, by means of short tandem repeat (STR) profiling, and copy number variant (CNV) analysis on a single fetal cell, unbiased whole genome amplification (WGA) is essential. This study investigates the performance of three WGA methods for STR profiling on single cells, isolated from a blood sample. To imitate a beginning-to-end workflow within cbNIPT, peripheral blood samples are first collected in blood tubes that contain a cell preservative. Next, cell-fixation and cell-permeabilization are performed to allow intracellular immunostaining of target cells, using fluorescently labeled antibodies. Then, single white blood cells are isolated and amplified, using three WGA methods in parallel: an advanced degenerate oligonucleotide primed PCR (DOP-PCR WGA) method, a ligation-adaptor PCR-based (LA-PCR WGA) method and a semi-linear WGA method that combines multiple displacement amplification and PCR (MDA-PCR WGA). Finally, STR analysis is performed on all WGA products by simultaneously amplifying the Amelogenin locus *AMEL* and 15 tetrameric STR loci across the human genome: *D8S1179*, *D21S11*, *D7S820*, *CSF1PO*, *D3S1358*, *TH01*, *D13S317*, *D16S539*, *D2S1338*, *D19S433*, *VWA*, *TPOX*, *D18S51*, *D5S818*, and *FGA*. Results demonstrate that the LA-PCR WGA method and the MDA-PCR WGA method yield STR profiles that show several dropouts, but are nevertheless adequate to allow fetal cell identification. Unfortunately, DOP-PCR WGA yields unusable STR profiles in the tested cbNIPT workflow.

521T

Full-gene sequencing based expanded carrier screening of over 150 disorders. C. Gijavanekar¹, B. Yuan^{1,2}, J. Scull^{1,2}, Z. Chen^{1,2}, L. Meng^{1,2}, S. Wen^{1,2}, R. Xiao^{1,2}, J. Zhang^{1,2}, E. Schmitt^{1,2}, S. Peacock^{1,2}, J. Dong¹, H. Cheng^{1,2}, D. Muzny³, R. Gibbs^{2,3}, A. Beaudet², F. Xia^{1,2}, C. Eng^{1,2}. 1) Baylor Genetics, Houston, TX; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX; 3) Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX.

Background: While the American College of Obstetricians and Gynecologists recommends carrier-screening of only 22 conditions, next-generation high-throughput full-gene sequencing has the potential to identify all known and predicted pathogenic variants associated with far more conditions. Based on the American College of Medical Genetics and Genomics (ACMG) policy statement (2013), GeneAware™ was designed to provide comprehensive pan-ethnic carrier screening of 159 genes for single nucleotide variants (SNV), indels, copy number variants (CNV), and CGG repeat expansion of the *FMR1* gene. **Methods:** We screened 6,291 females and 2,965 male samples by GeneAware. The test includes full-gene sequencing of 152 genes; full-gene CNV scan for *CFTR* and *DMD*; targeted CNV detection of 13 genes, including *MECP2*, *HBA1/2* etc., *FMR1* analyzed by CGG repeat expansion PCR, and *SMN1* carrier and silent carrier analysis were performed. Detection rates were >95% and varied by ethnicity for some conditions. Variant classification and interpretation used modified ACMG guidelines for automated or exhaustive manual curation to augment clinical sensitivity. Phenotype correlation was provided when possible. Reporting was limited to pathogenic or likely pathogenic variants. **Results:** Of the 9,256 individuals screened, three major ethnicities were Caucasian, Hispanic American and, African American. Notably, 53.3% were carriers of at least one disorder, with 16.9% being carriers of two or more disorders. Depending on ethnicities, 42% to 64% of individuals were identified as carriers for at least one disorder. The top five genes with positive findings were *HBA1/2* (10.4%), *SERPINA1* (7.5%), *CFTR* (4.1%), *HBB* (4%), and *GJB2* (3.4%). *FMR1* repeat expansion analysis in females detected 120 (1.3%) intermediate allele carriers and 25 (0.2%) premutation carriers. Caucasians had the highest incidence of *FMR1* expansions (0.65%). Of 1,011 couples, 30 (1.4%) were found to be carrier-couples. Half of the carrier couples had indicated a family history of an affected child. Six high-risk couples underwent prenatal testing for the identified variants in our laboratory. **Conclusion:** GeneAware is an expanded, full-gene sequencing based, pan-ethnic carrier screening panel which uses a variety of high-throughput technologies to increase the detection of carriers and carrier-couples. This expanded screening facilitates informed reproductive counseling.

522F

Fine mapping the MHC region identified two independent loci associated with non-obstructive azoospermia in Han Chinese males. M. Huang¹, M. Zhu¹, T. Jiang¹, Y. Jiang¹, G. Jin¹, Z. Hu^{1,2}. 1) Department of Epidemiology and Biostatistics, School of Public Health, Nanjing Medical University, Nanjing, China; 2) State Key Laboratory of Reproductive Medicine, Nanjing Medical University, Nanjing, China.

Objectives: Previous genome-wide association studies (GWAS) studies have identified the association between genetic variants in the major histocompatibility complex (MHC) region and non-obstructive azoospermia (NOA). However, the mechanism underlying the association remain unclear because of the extensive linkage disequilibrium (LD) existing in the MHC region. **Materials and Methods:** In the present study, using the specific Chinese Han-MHC database as a reference panel, we imputed all missing SNPs and HLA variants in the MHC region of GWAS data from 981 Chinese NOA patients and 1,657 controls. Combined with in silico bioinformatics analysis, we explored MHC variations in relation to NOA susceptibility comprehensively.

Results: We identified two independent risk loci: rs7194 (OR = 1.37, 95% CI = 1.21-1.56, $P = 1.43 \times 10^{-6}$) at MHC class II molecules, and rs4997052 (OR = 1.30, 95% CI = 1.15-1.46, $P = 2.26 \times 10^{-5}$) at MHC class I molecules. Functional annotation showed rs7194 was probably tag the effect of multiple amino acid residues and the expression of *HLA-DQB1* and *HLA-DRB1*; while rs4997052 showed the effect of amino acid changes of *HLA-B* at position 116 as well as the expression of *HLA-B* and *CCHCR1*, which co-expressed with genes enriched in pathways of spermatogenesis and male gamete generation. Combined novel findings with previously identified variants, approximately 6.71% of the phenotypic variances can be explained according to our study.

Conclusion: We fine mapped the MHC region and identified two loci that independently drove the NOA susceptibility. The results provided a deeper understanding of the association mechanisms of MHC and NOA risk.

523W

Association analysis of HLA-G gene 3'UTR variant (rs1063320) with recurrent miscarriage. V. Kalotra^{1,2}, G. Singh¹, V. Sharma³, I. Sharma³, M. La¹, I.C. Verma³, S. Sharma³, A. Kaur¹. 1) Department of Human Genetics, Guru Nanak Dev University, Amritsar, India; 2) Cytogenetics laboratory, Institute of Medical Genetics and Genomics, Sir Ganga Ram Hospital, New Delhi, India; 3) Human Genetics Research Group, School of Biotechnology, Shri Mata Vaishno Devi University, Katra, Jammu and Kashmir, India.

Recurrent miscarriage (RM) is defined as the loss of two or more clinical pregnancies and affects 15-20% of couples. About 50% cases of RM are idiopathic. A successful pregnancy is the outcome of coordinated interaction of several components between mother and fetus. Human leukocyte antigen (HLA)-G gene is a non-classical HLA class Ib gene and favours maternal acceptance of the semi-allogenic fetus by modulating the maternal immune system during pregnancy. It undergoes alternative splicing which produces seven different isoforms (four are membrane bound proteins three are soluble forms). *HLA-G* +3142G/C variant is located in 3' Untranslated region (UTR) of the gene and is involved in micro (mi) RNA mediated post transcriptional regulation. We analyzed the association of *HLA-G* rs1063320 (+3142G/C) variant in 91 RM women who had experienced at least two or more consecutive miscarriages, their 91 male partners and 100 healthy control females who had at least one healthy live born and no history of miscarriage. The genotyping was performed by using AGENA MassARRAY system. The distribution of variants followed the Hardy-Weinberg equilibrium ($p > 0.05$). The study depicted no significant association for rs1063320 variant and recurrent miscarriage. We also analyzed the association of allele sharing between the partners of 91 RM couples and number of miscarriages in RM couples. However, the study revealed no significant association for this comparison. Thus, it is suggested that further investigations involving other variants of the *HLA-G* 3'UTR and their haplotype analysis on a larger sample size may help to understand their role in RM. **Keywords:** Recurrent miscarriage, Human leukocyte antigen, 3'UTR The financial assistance to Vishali Kalotra (DST/ INSPIRE Fellowship/ 2013/ 889) from Govt. of India is duly acknowledged. No conflicts of interest are declared by the authors.

524T

A comprehensive roadmap of murine spermatogenesis defined by single-cell RNA-seq. Q. Ma¹, C.D. Green¹, G.L. Manske², A.N. Shami², X. Zheng¹, C. Sultan¹, S.J. Gurczynski³, B.B. Moore^{3,5}, J.Z. Li^{1,4}, S.S. Hammoud^{1,2,6,7}. 1) Department of Human Genetics, University of Michigan, Ann Arbor, MI; 2) Cellular and Molecular Biology Program, University of Michigan, Ann Arbor, MI; 3) Department of Internal Medicine, University of Michigan, Ann Arbor, MI; 4) Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI; 5) Department of Microbiology and Immunology, University of Michigan, Ann Arbor, MI; 6) Department of Obstetrics and Gynecology, University of Michigan, Ann Arbor, MI; 7) Department of Urology, University of Michigan, Ann Arbor, MI.

The process of spermatogenesis requires intricate interactions between the germline and the somatic cells. Within a given cross-section of a seminiferous tubule in mammalian testis, multiple germ and somatic cell types co-occur and these are different from the cell ensemble in the next cross-section. Thus, cellular heterogeneity is coupled with spatial complexity, making it difficult to understand the interaction among distinct cell types at different stages of germ cell development. To address this challenge, we collected single-cell RNA sequencing data from ~35K cells from the adult mouse testis, and identified all known germ and somatic cells, as well as two novel somatic cell types. Unlike previous datasets from similar studies we increased the representation of rare cell types by targeted experimental enrichment using a series of transgenic models. Our analysis revealed a continuous developmental trajectory of germ cell maturation from spermatogonia to the elongated spermatids, and identified candidate transcriptional regulators at several critical transition points during differentiation. Focused analyses delineated the presence of four subtypes of spermatogonial cells and nine subtypes of Sertoli cells, the latter can be linked to histologically defined developmental stages over the seminiferous epithelial cycle. Overall, this high-resolution cellular atlas represents a community resource to study germ cell development and *in vivo* gametogenesis. (Supported by NIH-R21HD090371 and the Michigan Center for Single-Cell Genomic Data Analytics.).

525F

Optimization of advanced DOP-PCR WGA for STR profiling in cell-based non-invasive prenatal testing. J. Weymaere, A.S. Vander Plaetsen, P. Caenen, L. Tilleman, D. Deforce, F. Van Nieuwerburgh. Faculty of Pharmaceutical Sciences, Ghent University, Ghent, Belgium.

In the context of cell-based non-invasive prenatal testing (cbNIPT), fetal cells are isolated from the maternal peripheral bloodstream and used to perform prenatal testing such as copy number variant (CNV) analysis. Unfortunately, fetal cells in the maternal bloodstream are extremely rare and cannot be identified with a unique fetal marker. Therefore, short tandem repeat (STR) profiling is performed to verify the fetal identity of the isolated cell before prenatal diagnosis. As a single cell contains only 6-7 pg DNA, whole genome amplification (WGA) is essential to increase the amount of DNA up to ng- or µg- levels to allow both STR profiling and CNV analysis on a single cell. Since each WGA method introduces bias and thereby influences the outcome of downstream assays, the WGA method is selected based on the required downstream applications. Until now, there is no WGA method well suited for both STR- and CNV analysis. For instance, WGA based on Multiple Displacement Amplification (MDA) give rise to faultless and complete STR profiles but CNV profiles show extreme representation bias. In return, advanced Degenerate Oligonucleotide Primed PCR (DOP-PCR) is well fitted for CNV analysis although STR profiles are incomplete. The aim of this work was to optimize advanced DOP-PCR and increase the quality of the STR profiles. Therefore, specific STR primers were added to the advanced DOP-PCR WGA, which is called advanced DOP-PCR with Targeted Sequence Enrichment (TSE). In that way, STR amplicons are already selectively amplified during WGA. The STR primers that were used, originate from the Identifier Plus AmpFISTR PCR kit. Four different concentrations of STR primers were tested: no primers, 2.67 µM, 4 µM, and 8 µM of the STR primer pool. The performance of this advanced DOP-PCR with TSE was assessed using one and three micromanipulated cells from the Loucy cell line. STR profiles were generated using the same Identifier Plus AmpFISTR kit, which is a multiplex PCR kit that amplifies 15 STR loci (*D8S1179*, *D21S11*, *D7S820*, *CSF1PO*, *D3S1358*, *TH01*, *D13S317*, *D16S539*, *D2S1338*, *D19S433*, *VWA*, *TPOX*, *D18S51*, *D5S818*, *FGA*) and the Amelogenin locus *AMEL*. Results demonstrate that implementation of TSE has a positive effect on the drop-out rate, which increases the quality of the STR profiles. Addition of STR primers in an 8 µM concentration results in the lowest drop-out rate and most complete profiles.

526W

Transcriptional profiling of circulating exosomes defines molecular phenotypes of preeclampsia. J. Schuster^{1,2}, V. Zarate¹, C. Felber¹, A. Uzun^{1,2,3}, J. Padbury^{1,2,3}. 1) Pediatrics, Women and Infants Hospital of Rhode Island, Providence, RI; 2) Pediatrics, Brown University, Providence, RI; 3) Center for Computational and Molecular Biology, Brown University, Providence, RI.

Preeclampsia is a multi-system hypertensive disorder of pregnancy, characterized by variable degrees of maternal symptoms including elevated blood pressure, proteinuria and fetal growth retardation. There is currently no effective intervention for preeclampsia short of delivery of the fetus. Although the heritability of preeclampsia has been well documented, specific genetic contributions have not been identified. We have shown that genes associated with preeclampsia are phenotype specific. To better understand the underlying mechanism, we are investigating the transcriptional profile of exosomes from the plasma of women with the extreme phenotype of early onset severe preeclampsia. Nearly all tissues and cell types release vesicles from their plasma membrane. The contents of extracellular vesicles, including both mRNA and miRNA, are tissue specific can be functionally transferred to recipient cells and lead to alterations in downstream signaling. We hypothesize that the mechanism for severe preeclampsia can be revealed by comparing the transcript profile of the extracellular vesicles from women with severe preeclampsia to gestational age matched normotensive women. To test this hypothesis, we are enrolling a cohort of patients, isolating plasma derived exosomes and sequencing the exosome RNA. We have successfully isolated exosomes and have confirmed their size, shape and concentration with western blots Nanosight Analysis and electron microscopy. In a preliminary analysis of 8 normotensive term women, we have sequenced the exosomal messenger RNA cargo and used STAR, HT_Seq and DESeq2 to identify differentially expressed genes in the pregnant mothers compared to 24 hours post-delivery. 79 genes were differentially expressed after FDR correction. The Citrulline Mediated Endocytosis pathway, involved in vesicle-target cell communication, was enriched for upregulated genes. We are using innovative bioinformatics techniques to further analyze this data and the severe pre-eclamptic/gestational age matched control samples. This includes qSVA to correct for bias in RNA expression due to delayed processing and Nonnegative Matrix Factorization to deconstruct the transcription profiles of the exosomal cargo into the contribution from their tissues of origin. Identification of the transcript profiles in these extracellular vesicles will provide insights into the mechanisms severe preeclampsia and lead to new strategies for treatment, prediction and prevention.

527T

Cause for clinical concern: Genome-wide cfDNA for cases screening positive for trisomy 15. E. Soster, T. Boomer. Sequenom/Integrated Genet-ics, San Diego, CA.

Cell-free DNA (cfDNA) testing has been integrated into routine prenatal care. Expansion of cfDNA technology includes select microdeletions, large copy number variants, and esoteric aneuploidies. Initial data regarding outcomes from a commercial genome-wide cfDNA test has been described.^{1,2,3} Based on this commercial laboratory experience, the outcomes observed for cases positive for trisomy 15 (T15) suggest significant risk for complications in these pregnancies. A retrospective analysis was performed on over 40,000 maternal blood samples submitted for genome-wide cfDNA analysis. Samples were subjected to DNA extraction, library preparation, and genome-wide massively parallel sequencing.⁴ Sequencing data were analyzed using a novel algorithm to detect trisomies and subchromosomal, genome-wide copy number variants 7Mb and larger.¹ The cases that screened positive for T15 were reviewed. Clinical outcomes were requested from ordering providers as part of routine follow-up of positive cases. There were 18 cases that screened positive for T15. Of those, 3 cases were confirmed by diagnostic testing. An additional 8 cases ended in miscarriage or fetal demise. There were 2 discordant cases with a known co-twin demise and 3 discordant cases with no obvious biological explanation, but raw sequencing data in those cases suggests mosaicism. Two cases were lost to follow-up and outcomes are unknown. Fourteen of 18 cases included maternal age as a reason for referral. Ultimately, 50% (9/18) of cases ended in pregnancy loss. One of the confirmed cases resulted in a fetus with Prader-Willi syndrome (PWS), suggesting a trisomic rescue. T15 is a fairly common cause of pregnancy loss, accounting for 1.68% of all spontaneous abortions and 7.4% of trisomies seen in spontaneous abortions.⁵ As T15 is usually an error in maternal meiosis, it follows that many of these cases were referred due to maternal age.⁶ The high rate of adverse outcomes can be explained by the frequent meiotic origin of T15, which has been previously described.^{6,7} These outcomes are similar to those reported in another cohort of cfDNA cases.⁸ Although the sample size is relatively small and unfortunately, most of the losses did not have diagnostic testing, the cause for concern is apparent. Whether caused by full or mosaic T15 in the fetus or placenta, or UPD leading to PWS/Angelman syndrome, a positive cfDNA result for T15 carries significant concern for an adverse outcome, particularly pregnancy loss.

528F

Improved DNA amplification used for human karyomapping to phase single gene defects. S. Rulli¹, M. Horňák², A. Janosch³, D. Unt³, N. Hochstein³, R. Kläver³. 1) QIAGEN Sciences Inc., 6951 Executive Way, Frederick, MD 21703, US; 2) REPROMEDA BIOLOGY PARK, Studentska 812/6, 62500 Brno, Czech Republic; 3) QIAGEN GmbH, QIAGEN Strasse 1, 40724 Hilden, Germany.

Karyomapping allows to phase single gene defects in embryos and is performed when a couple has a family history of a genetic disorder. In most cases, the method is applied on a single cell or small number of cells from an embryo during preimplantation genetic diagnosis. One obstacle when performing karyomapping on single or few cells is the small amount of sample DNA. Whole genome amplification (WGA) can be used to overcome this limitation. The performance of WGA is essential as the quality of amplified DNA affects karyomapping results. To improve reliability and reproducibility of WGA and subsequent karyomapping experiments we developed an advanced version of the Phi-29 amplification system and added a single cell storage buffer optimized for collecting and storing eukaryotic single cells. In this study, we have validated the performance and robustness of the new WGA components and protocol (REPLI-g Advanced DNA Single Cell Kit, QIAGEN) and its compatibility with subsequent karyomapping experiments. For karyomapping we performed SNP genotyping of 300.000 SNP markers across human genome using the Human KaryoMap-12 SNP array (Illumina) and used the corresponding QC metrics for validation of WGA performance. These QC metrics comprised overall call rate, AB rate, Allelic drop out rate and Miscall rate. We analyzed numerous cases (on average 2 - 9 embryos per case) with family history of different genetic disorders e.g. fragile X syndrome, cystic fibrosis or nephronophthisis. Overall, the QC metrics of all cases displayed good quality of the amplified sample DNA from parental origin as well as from embryo biopsies. In addition, the karyomapping could be used to identify embryos free of chromosomal aneuploidies, which frequently occur in human embryos. Our study confirms that the improved DNA amplification is highly suitable for human karyomapping and gives reliable and reproducible results for phasing of single gene defects.

529W

Parents' attitude of chromosomally affected child towards carrier screening. N.N. Kadam¹, B.B. Ganguly². 1) MGM Medical College & Hospital, Navi Mumbai, Maharashtra, India; 2) MGM New Bombay Hospital, Navi Mumbai, Maharashtra, India.

The information generated from genotype-phenotype correlations has importance in clinical setting. Genetic counseling and screening of blood-linked relatives not only prevents transmission of irreparable genetic alterations to future generations but also help in understanding its impact on future health. We present a 6 days old female child born with choanal atresia and subsequently detected with an abnormal karyotype with an extra recombinant marker chromosome. She also had anteposed anal opening, ASD and developed hyperbilirubinemia. The child cried immediately after birth; however, shortly developed acute respiratory distress. An emergency nasogastric connection had detected bilateral choanal atresia, which had opened the left nostril partly. Conventional karyotyping in peripheral blood detected a constitutional abnormality with 47,XX,+mar pattern. The marker chromosome was confirmed as an acrocentric one from its involvement in acrocentric association, which might have been produced from a balanced translocation present in one of the parents. The father and the elder sibling also complained of some amount of nasal blockade (? unilateral atresia) causing chronic cold and cough. However, the family was reluctant to undergo karyotyping and confirm the source of the recombinant chromosome in their female child. This is the first case of choanal atresia caused by unbalanced chromosomal rearrangement. Carrier screening of the parents was extremely important for identification of the chromosomes involved in the proband, for their elder son and future life of their offspring. Presence of a balanced translocation, if present in one of the parents, had influenced meiotic non-disjunction resulting in aneuploidy in the fertilized gamete.

530T

Clinical utility of expanded carrier screening: Results-guided actionability and outcomes. G.A. Lizarin, K. Johansen Taber, K.A. Beauchamp, A. Arjunan, D. Muzzey, K. Wong, J.D. Goldberg. Counsyl, South San Francisco, CA.

Introduction: Expanded carrier screening (ECS) facilitates reproductive and pregnancy management, such as in vitro fertilization (IVF) with preimplantation genetic diagnosis (PGD) and prenatal diagnostic testing. However, studies exploring the actions of at-risk couples (ARCs) after receiving ECS results have been limited by small sample sizes. Here, we describe the impact of ECS results on planned and actual pregnancy management in the largest sample of ARCs studied to date. **Methods:** Couples who elected ECS and were found to be at high risk of having a pregnancy (ARCs) affected by at least one of 176 genetic conditions were invited to complete a survey about the impact of ECS results on planned and actual pregnancy management. **Results:** 391 ARCs completed the survey. 77% of ARCs who received results before they were pregnant reported that they altered or were planning to alter pregnancy management to reduce the risk of an affected pregnancy, including IVF with PGD, use of donated gametes, adoption, avoidance of pregnancy altogether, and prenatal diagnosis if/when they became pregnant. 37% of ARCs who received results while they were pregnant elected to undergo prenatal diagnosis. Approximately one-third of pregnancies were found to be affected, and 42% of affected pregnancies were terminated. Among those who received ECS results before they were pregnant, 126 pregnancies occurred. In those who were pregnant when they received ECS results, 40 subsequent pregnancies occurred. This yielded 166 pregnancies in which ARCs knew their ECS results prior to conception. 35% of these pregnancies were achieved through IVF with PGD. Prenatal diagnosis was pursued in approximately 30% of pregnancies, with 29% of these pregnancies found to be affected; 75% of affected pregnancies were terminated. In couples who chose not to pursue prenatal diagnosis, common reasons cited were to avoid the risk of miscarriage, a perception that it was not necessary because the couple had undergone IVF with PGD, a perception that the risk of an affected pregnancy was low, and a perception that a positive result would not have led to pregnancy termination or informed management. **Conclusions:** ECS results impacted couples' reproductive decision-making and led to altered pregnancy management that reduced the risk of an affected pregnancy.

531F

Mendelian mutations may explain a significant number of unexplained recurrent spontaneous abortions. H.M. Byers¹, M.E. Mehaffey², A. LaCroix³, R.J. Silver³, M.A. Bamshad², I.A. Glass², H.C. Mefford². 1) Pediatrics, Stanford University, Stanford, CA; 2) Pediatrics, University of Washington Medical Center, Seattle, WA; 3) Obstetrics & Gynecology, University of Utah, Salt Lake City, UT.

BACKGROUND: Infertility is a global problem with high social, economic, medical and personal costs. Approximately 50%, the evaluation for an etiology is unsuccessful and couples don't receive an explanation, leaving them with few options. We predicted that some couples with recurrent spontaneous abortions may have a Mendelian etiology for the embryonic or fetal loss. If detected, the couple could be offered individualized treatment to support the pregnancy, more precise recurrence recurrence and/or diagnostic testing.

METHODS: Couples with idiopathic recurrent pregnancy losses, defined as ≥ 2 or more unexplained spontaneous abortions, were enrolled. Medical records were reviewed to ensure recurrent pregnancy losses had no attributable etiology. DNA was extracted from abortus, male and female partner. Chromosomal microarray was performed if chromosomal assessment had not been previously performed on the abortus. Trio exome sequencing with variant analysis was performed. Candidate results were Sanger sequenced for confirmation.

RESULTS: Ninety-eight couples (77 trios) met eligibility criteria after medical record review, provided a sample for DNA extraction and consented for enrollment. Chromosomal microarray was performed on the 24 without chromosomal analysis prior to enrollment, 3 were abnormal and excluded. After DNA quantity, DNA quality and fingerprinting were assessed, 84 samples from 26 trios were prepared for exome sequencing. 11 samples failed amplification. 73 samples from 20 trios were successfully sequenced and analyzed. Variants were assessed considering first the fetus, male and female partner as the proband. Of the 20 trios, a candidate diagnostic, Mendelian change was identified in 10 of the 20 trios, for a diagnostic rate of 50%. Of these, 4 were *de novo* variants in the fetus, 1 was inherited from a mosaic father, 2 were changes in mother that were thought to contribute to inability to support the pregnancy and 1 was X-linked. **CONCLUSION:** Trio-exome sequencing is a powerful tool for hypothesis-generation research for conditions predicted to have a significant Mendelian genetic component with limited phenotypic information. Despite limited and heterogeneous phenotypic information, a diagnostic rate of 50% (10/20) was achieved for couples with recurrent pregnancy losses, providing the possibility for targeted treatment in female partner and more precise information about likely recurrence risks. Further study is warranted. .

532W

Calculating power to detect maternal and fetal effects in genetic association studies. *G-H. Moen^{1,2}, N.M. Warrington³, D.M. Evans^{3,4}.* 1) Department of Endocrinology, Morbid Obesity and Preventive Medicine, Oslo University Hospital, Oslo, Norway; 2) Faculty of Medicine, University of Oslo, Institute of Clinical Medicine, Oslo, Norway; 3) University of Queensland Diamantina Institute, Translational Research Institute, Brisbane, Australia; 4) Medical Research Council Integrative Epidemiology Unit, University of Bristol, Bristol, UK.

Background: Perinatal offspring outcomes, such as birthweight, are a function of maternal genetics operating on the intrauterine environment, fetal genetics and environmental factors. Partitioning genetic effects into maternal and fetal components is not straightforward because maternal and child genotypes are correlated. Such partitioning requires genotyped mother-offspring duos or genotyped individuals with phenotypic information on both themselves and their offspring. **Methods:** We derived closed form solutions, performed asymptotic power calculations and data simulations to estimate power to detect maternal and fetal genetic effects after accounting for the correlation between the genotypes. We performed calculations for both a one degree of freedom test where the maternal or fetal effect was constrained to zero and a two degree of freedom test where both parameters were set to zero. We tested the effect of different study designs, missing data and strength of residual correlation between maternal and fetal phenotypes on power. To make our power calculations accessible, we developed an online power calculator, the "Maternal and Fetal Effects Power Calculator", allowing others to conveniently estimate the power to detect these quantities in their own studies. **Results:** 53,000 genotyped mother-offspring duos were required to achieve 80% power to detect a variant exerting a maternal (or fetal) effect that explained 0.1% of the trait variance at genome-wide levels of significance (one degree of freedom test). Genotyped mother-offspring duos were in general the most powerful study design for detecting maternal or fetal effects; however, studies of genotyped individuals with their own and their offspring's phenotype had similar power when the residual correlation between maternal and offspring phenotypes approached 0.5. A two degrees of freedom test to estimate whether the genetic variant had any effect on the outcome (maternal and/or fetal) was considerably more powerful for locus detection purposes compared to a one degree of freedom test estimating either the maternal or fetal component. Studies where genotyped individuals had either their own or their offspring's phenotype (but not both) were the least powerful of the research designs considered. **Conclusion:** Samples sizes required to detect maternal or fetal effects that explain realistic proportions of the trait variance with appreciable power are achievable and within the range of current research consortia.

533T

When is a disease too rare? A statistical framework for estimating clinical sensitivity for expanded carrier screening panels. *R. Ben-Shachar, A. Svenson, D. Muzzey.* Counsyl, South San Francisco, CA.

Current guidelines issued by the American College of Obstetricians and Gynecologists (ACOG) provide criteria for determining if a disease should be included in an expanded carrier screening (ECS) panel. The criteria endeavor to confine screening to diseases that are severe, phenotypically well understood, testable via prenatal diagnosis, early-onset, and common (a carrier rate above 1 in 100). We analyzed de-identified data from 44,483 patients who underwent ECS on a 176-condition panel to evaluate how exclusion of diseases that do not meet certain criteria impacts the rates of detection of both individual carriers and at-risk couples. The criterion that a disease must have a carrier frequency above 1 in 100 can be interpreted in different ways, as carrier frequencies vary widely by ethnicity. Depending on the definition used, this criterion is satisfied by between 12 (7%) and 75 (43%) conditions on the 176-condition panel; restricting the panel only to conditions meeting this criteria would reduce identification of at-risk couples by between 5% and 51% and carriers by between 20% and 83%, respectively. To determine a less ambiguous and more clinically relevant carrier-frequency cutoff, we introduce a statistical framework that models clinical sensitivity, an analysis based on the probability that sufficient reported cases exist in the literature to interpret the pathogenicity of observed variants. We show that the allele-frequency distribution of pathogenic variants for most diseases is well approximated by an exponential distribution, and that estimated clinical sensitivity depends on the value of the exponential decay parameter: Diseases with lower decay parameters have lower estimated clinical sensitivities than diseases with higher decay parameters. Our modeling suggests that a disease with a carrier frequency as low as 1 in 4,000 could still achieve a clinical sensitivity of 50%. In summary, we show that, while disease-inclusion criteria are needed for an ECS panel to be clinically meaningful, the 1-in-100 carrier-frequency threshold is unclear and arbitrary, limiting at-risk couple detection. We suggest that this threshold should be revisited and propose an alternative model to determine when a disease is too rare to include in an ECS panel.

534F

GenetiSure pre-screen assay for pre-implantation genetic screening by array CGH. N. Novorodovskaya¹, S. Basehore², S. Happe². 1) R&D, Agilent Technologies, La Jolla, CA; 2) R&D, Agilent Technologies, Cedar Creek, TX.

Preimplantation genetic testing for aneuploidy (PGT-A) is a technique that is used to screen for chromosomal abnormalities in embryos prior to transfer in an *in vitro* fertilization (IVF) cycle. Next generation sequencing (NGS) and array-based comparative genomic hybridization (aCGH) are two major platforms that have become widely adapted for aneuploidy screening in human embryos with a current detection limit of 5-10 Mb and a turnaround time of at least 12 hours. The aim of this study was to develop a high throughput, cost-effective PGT-A workflow with shorter turnaround time, to avoid the need for embryo cryopreservation and enable embryo transfer in the same IVF cycle, and higher resolution, based on the Agilent aCGH platform. The GenetiSure Pre-Screen assay, that works with two common whole-genome amplification (WGA) methods (multiple displacement amplification (MDA) and PCR-based) was developed. The workflow includes the following steps: WGA, enzymatic labeling, purification of amplified labeled DNA and hybridization to the GenetiSure Pre-Screen microarrays specifically designed for an aneuploidy screening. A single-color array design allows processing of up to 14 test and 2 reference samples per slide, providing high-resolution and cost-effective results in less than 8 hours when using the MDA method. Three single cell data analysis methods implemented in the CytoGenomics software allow detection of: (1) aberrations > 5 Mb; (2) small aberrations > 1 Mb and (3) aberrations > 20 Mb in mosaic samples. Cell lines with known aberrations were used as a model system and biopsied cells from day 3 or day 5 embryos were used to validate the workflow. Using single cell (1-2 cells) and few cell (3-5 cells) samples isolated from cell lines with known aberrations (Coriell), we demonstrated successful detection of copy number (CN) gains and losses > 5 Mb using the Single Cell Recommended analysis method and detection of expected aberrations in 1 Mb range using the Single Cell Small aberration analysis method. Using mixtures of two cell lines to mimic mosaic samples, we demonstrated the detection of a 30 Mb aberration at 20% mosaicism with the Single Cell Long Low analysis method. High correlation of the GenetiSure PreScreen results with other confirmatory methods (NGS, FISH or aCGH) used during assay validation demonstrated high accuracy of the GenetiSure PreScreen assay.

535W

Pre-implantation genetic screening: One year experience at a single center in Northwest fertility clinic in Mexico. L. Patron^{1,2}, H.E. Almanza-Reyes¹, J.J. Hernández-Ledezma², J.F. Walters-Arballo². 1) Universidad Autonoma de Baja California, Tijuana, CA; 2) Hospital Ángeles Tijuana.

Objective: Describe the indications for pre-implantation genetic testing for aneuploidy (PGT-A) cycles in our clinic population. **Desing:** Retrospective cohort study. **Setting:** Private medical reproduction center **Methods:** This is a retrospective cohort study from June 2017 to June 2018 at the Fertilite Clinic in Tijuana, Baja California, Mexico. The samples were obtained during (75 PGS cycles) undertaken for 57 couples. Indications for PGT-A are such as maternal advanced age, recurrent implantation failure, ectopic pregnancy, male factor infertility, recurrent spontaneous abortion, and others. Biopsy was performed at day 5-7 post-fertilization, approximately 6-10 cells from the trophectoderm layer of blastocyst was analyzed. Next Generation Sequencing (NSG) technology was performed to screen all chromosomes for aneuploidy. This test is limited to screening for aneuploidies greater than 20MB in size.

Results: The mean female age (years,±) 38.3±5.3 vs. mean male age 37.2 ± 5.7. 17/57 couples received egg from donors, the mean age in this group was 26.5 ± 1.6. The main indication for the genetic screening was advancing maternal age/low ovary reserve in 38 (66%) couples. Infertility duration was 4.8 ± 2.6. AMH(ng/mL) 3.03±2.9, BMI(kg/m²) 30.8±4.7. 231 embryos were analyzed, blastocyst with genetic result 225 (97.4%), blastocyst no result 6(2.5%). Euploid blastocysts: 121 (52.38%) vs. 91 (47.63) aneuploid blastocysts. In the euploid group, 55(45.5%) were female vs. male 66 (54.5%), in the aneuploidy group 48 (52.7%) female embryos vs. 43 (47.5%) male embryos. In this group trisomy was present in 47 (53%) samples, 1 mosaic of trisomy 8, and 46 (51%) cases of monosomy. 9 cases report trisomy and monosomy simultaneously. 12 (5.2%) blastocyst report complex chromosomal abnormalities, 2 (0.8%) were inconclusive and 4 (1.7%) reported low DNA. **Conclusion:** In our study the main indication for IVF/PGT-A was maternal advanced age/low ovary reserve, this differs from other studies, were recurrent spontaneous abortion is more common. It was expected that, the proportion of aneuploid embryos was higher in women above 35 years. In order to measure the impact of the effectiveness of PGT-A we will follow the implantation rates, miscarriage rates and the full term pregnancies. The limitation of our study is that is only 12 months; we will extend this period to analyze this data.

536T

Fetal fraction evaluation in non-invasive prenatal testing. *M.S. Hestand^{1,2,3}, M. Bessem³, P. van Rijn³, R.X. de Menezes⁴, D. Sie³, I. Bakker³, E.M.J. Boon³, E.A. Sistermans³, M.M. Weiss³.* 1) Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA; 2) Department of Pediatrics, University of Cincinnati, Cincinnati, OH, USA; 3) Department of Clinical Genetics, VU University Medical Center Amsterdam, Amsterdam, NL; 4) Department of Epidemiology and Biostatistics, VU University Medical Center Amsterdam, Amsterdam, NL.

A sufficient percentage of fetal DNA is required for non-invasive prenatal testing (NIPT) to avoid false negative results. During a six month period, we successfully performed shallow whole-genome sequencing on 14,379 NIPT samples. Technical and biological factors were evaluated for their influence on the percentage of fetal DNA. We found shipping time did not influence the percentage of fetal DNA. However, bioinformatics algorithms had a profound influence on fetal DNA fractions. DEFRAG (Y-chromosome based analysis) performed best on male fetuses while SeqFF (count-based analysis) performed best on female fetuses. Biologically, a negative correlation was found between fetal fraction and an increased weight/BMI. Gestational age of up to 21 weeks and maternal age had no influence on fetal fraction. Finally, though not statistically significant, we did observe a trend for increased fetal fractions in trisomy 21 samples and decreased fetal fractions in trisomy 13 and 18 samples. In the future, individual labs should further optimize detection methods and associations with larger cohorts performed to successfully develop appropriate fetal fraction cut-offs for reliable detection of fetal aneuploidies and smaller copy number variants with NIPT.

537F

Performance verification of the OnePGT preimplantation genetic testing system: Concurrent detection of aneuploidies, structural rearrangements, and monogenic disorders in single cells from *in vitro* fertilized (IVF) embryos. *S. Happe¹, K.O. François¹, J. Allemeersch¹, R. Richards¹, S. Moeyers¹, J. Theuns¹, M.Z. Esteki^{2,4}, C. Melotte⁴, E. Coonen^{2,3}, J. Dreesen^{2,3}, A. Paulussen², J. Dumoulin³, C. Van Uum², M. Drusedau², J. Engelen², K. Derks², R. Van Golde³, E. Dimitriadou⁴, H. Masset⁴, C. De Die-Smulders², J.R. Vermeesch⁴.* 1) Agilent Technologies, Diagnostics and Genomics Group; 2) Department of Clinical Genetics, GROW School for Oncology and Developmental Biology Maastricht University Medical Center+, Maastricht, The Netherlands; 3) Center for Reproductive Medicine, GROW School for Oncology and Developmental Biology, Maastricht University Medical Centre+; 4) Center for Human Genetics, University hospital of Leuven, KU Leuven, Leuven, Belgium.

OnePGT (Agilent Technologies) is a genome-wide next-generation sequencing (NGS) solution for concurrent preimplantation testing for monogenic disorders (PGT-M) structural rearrangements (PGT-SR), and aneuploidies (PGT-A), based on the haplarithmis concept which combines haplotyping and copy-number profiling of DNA from single cells (Zamani Esteki et al, 2015). In collaboration with Maastricht University Medical Center (MUMC) and Catholic University of Leuven (KU Leuven), a total of 213 embryo samples (177 PGT-M; 36 PGT-SR) were tested for concordance with the respective gold standard PGT method (SNP array with haplarithmis (siCHILD) analysis at KU Leuven; STR-PCR for PGT-M, or FISH/array CGH for PGT-SR at MUMC). For the KU Leuven study, left-over whole genome-amplified DNA from the embryo biopsies was processed, while at the MUMC site, a second biopsy was taken and amplified with REPLI-g sc (Qiagen). In addition, a total of 86 cell line samples (36 single cells and 48 few (5-10) cells) originating from 12 Coriell cell lines, of which 10 were aneuploid and 2 were euploid, were tested internally to demonstrate the sensitivity and specificity of the OnePGT solution for PGT-A. NGS libraries were generated using OnePGT library preparation reagents, sequenced on NextSeq500 (Illumina) and data analyzed within the Agilent Alissa OnePGT software suite. For PGT-M analysis, OnePGT demonstrated a 91% call rate (n=161). The 16 embryos which did not receive a conclusion were shown to possess LOH-like regions in the family members (n=10) that precluded haplotyping, or single gene disorder loci that either fell on a recombination breakpoint (n=5) or within uninformative regions of low SNP density (n=1). After excluding samples with chaotic genome-wide copy number profiles or those for which gold standard results were inconclusive, 100% of the 154 remaining samples agreed with the gold standard PGT result. For PGT-SR analysis (n=36), 2 embryos were shown to have additional mitotic aberrations due to an unstable genome; of the remaining 34 embryos, 100% concordance with the reference PGT methods was demonstrated. Of the 84 Coriell cell line samples processed for PGT-A concordance analysis, an observed sensitivity of 98.5% and an observed specificity of 94.7% was demonstrated. In addition, a limit of detection of 5 Mb and 20 Mb for single and few cell sample types respectively was demonstrated. For Research Use Only. Not for use in diagnostic procedures.

538W

Digital karyomapping following genomic autopsy: Implications for preimplantation genetic diagnosis in the genomic era. T. Hardy^{1,2,3}, D. Lawrence⁴, A. Byrne⁴, P. Arts⁴, S. King-Smith⁴, M. Babic⁴, C. Barnett⁴, H. Scott^{1,4}. 1) SA Pathology, Adelaide, Australia; 2) Repromed, Adelaide, Australia; 3) Women's and Children's Hospital, Adelaide, Australia; 4) Centre for Cancer Biology, University of South Australia, Australia.

Background: Karyomapping identifies informative single nucleotide polymorphism (SNP) profiles in embryos using an array sampling ~300,000 genomic loci. Karyomapping may not be feasible if there are insufficient loci around the gene of interest due to poor SNP coverage or consanguinity. Whole genome sequencing could define all informative variants individualised to a couple; however, the expected resolution of this strategy has not been assessed. **Aim:** To develop an automated method for digital karyomapping feasibility studies using genomic data and compare to standard karyomapping. **Method:** A user-friendly graphical tool for digital karyomapping was built on top of VariantGrid (variantgrid.com), a variant database developed at SA Pathology. WGS data was analysed for two families with an abnormality in a previous pregnancy. Informative variants within a standard karyomapping window of 2Mb of the gene of interest were identified and compared to a virtual SNP array. **Results:** Compound heterozygous mutations were identified in *MDFIC* (Family 1) and *MKS1* (Family 2). Genome-wide analysis demonstrated 97,067 informative variants on SNP array compared to 3,107,644 variants on WGS for Family 1, and 103,682 informative variants on SNP array compared to 3,389,623 informative variants on WGS for Family 2. This represents a 30-fold increase in available informative variants using WGS. In the 2Mb window around the gene of interest, there were 2,968 variants around *MDFIC* and 3,915 variants around *MKS1*. SNP array would have identified only 2% of the available information (61 variants for Family 1 and 87 variants for Family 2). **Conclusion:** Digital karyomapping provides a 30-fold increase in the availability of informative variants for analysis in embryos and may be utilised in cases in which higher resolution is required. Immediate transition to IVF and PGD without further feasibility studies will be possible for couples who have undergone WGS of a rare disorder in a previous pregnancy.

539T

Rapid whole exome sequencing for diagnosis of prenatal phenotypes.

A. Haworth¹, S. Drury¹, Y. Patel¹, R. Dubis¹, J. Rizvi¹, N. Lench¹, J. Short², S. Mansour³, T. Homfray³. 1) Congenica Ltd, Hinxton, Cambridgeshire, CB10 1DR, UK; 2) South West Thames Regional Genetics Laboratory, St George's Hospital, London, SW17 0QT, UK; 3) Clinical Genetics, St Georges Hospital, London, SW17 0QT, UK.

Objectives: i) Development, refinement and implementation of a rapid whole exome sequencing workflow from DNA to report for the diagnosis of fetal abnormalities detected on ultrasound; ii) Expansion of fetal phenotype in a single centre. **Method:** DNA (CVS, amniotic fluid, fetal blood or tissue, parental gDNA) was enriched using the Agilent Sureselect CREv2 and sequenced using Illumina NextSeq 500. Maternal cell contamination exclusion was performed where relevant. HPO terms and pedigree information were loaded into the NextSeq run folder at initiation of the sequencing run. Data from completed runs was automatically collected from the instrument daily, processed using DRAGEN BIO-IT processor and Sapientia™. Gene panels and variant filter settings were preconfigured in Sapientia™ to further expedite analysis. Analysis was stratified by candidate genes, phenotype relevant panels and DDG2P with Exomiser prioritisation as relevant for each case. **Results:** To date 18 cases with a prenatal phenotype have been tested; 13 trios and 5 singletons. 1 case with tested in the NICU. A molecular diagnosis was made in 5/18 (28%), including mutations in *CHD7*, *PIEZO1*, *PTPN11*, *BICD2* and *FGFR2*. DNA to primary result using our current workflow can be reduced to 7 working days. In three cases, the molecular diagnosis was not the suspected clinical diagnosis (*CHD7*, *BICD2*, *FGFR2*). The causative mutation was identified using DDG2P and Exomiser variant prioritisation in 2 cases (*CHD7*, *FGFR2*) and phenotype relevant gene panel (Genomics England PanelApp) in one (*BICD2*). **Conclusion:** Whole exome sequencing for prenatal diagnosis can be challenging due to limited phenotype and requirement for rapid turnaround. The utility of stratifying analysis to candidate genes, followed by extension to phenotype relevant panels is highlighted in this setting. Using streamlined laboratory and data processing workflows we continue to identify areas for improvement to facilitate translation to routine diagnostic use. We are validating a parallel SNP barcoding method to help obviate the need for Sanger confirmation which adds significant time to that reported here. Whole exome sequencing is a valuable tool for furthering gene-prenatal presentation association.

540F

Parental age effects on the incidence of human de novo pathogenic variants. Y. Cao¹, P. Stankiewicz¹, J. Salvo², F. Xia^{1,2}, Y. Guan^{1,2}, C. Eng^{1,2}, P. Liu^{1,2}, J. Zhang^{1,2}. 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, USA; 2) Baylor Genetics, Houston, Texas, USA.

De novo DNA change(DNC) is the major mechanism causing human monogenic disorders demonstrated by clinical whole-exome studies that nearly 60% of solved cases were caused by dominant disorders and >70% of them were attributed to de novo pathogenic variants. Approximately 80% of de novo germline point variants arise from paternal origin likely due to the DNA replication and repair errors during mitotic spermatogenesis. Advanced paternal age at conception (PAC) is associated with the increased incidence in several dominant diseases including FGFR2- and FGFR3- related disorders. However, for the majority of dominant diseases it is still unclear whether there is a paternal and/or maternal age effect on the occurrence of de novo changes. We retrospectively studied more than 2,000 cases which were referred for whole exome sequencing, diagnostic panel, and non-invasive single gene disorder panel at the Baylor Genetics from November 2012 to April 2018. 1328 patients from 1325 unrelated families were found to carry at least one apparently de novo variant, with a total of 2394 de novo variants on 1984 different genes. A total of 296 genes were found with at least two de novo changes in unrelated patients, including 14 variants arising in more than two times. Paternal age at conception ranged from 17-78 yrs(average age 32.4 yrs with the standard deviation (SD) of 6.9 yrs) while maternal age at conception from 15-46yrs (avg age 29.5 yrs with SD of 5.7 yrs). The average number of DNC in the exome region per case gradually increased from 20-40 years in both maternal and paternal group with similar rate. The increasing trend was persistent for fathers from 40yr to 60yr. Interestingly, the growth rate of DNC accelerated when PAC was older than 40 yrs compared with that in the 20-40 yrs group. Furthermore, after functionally annotating genes, we found that pathogenic de novo variants were clustered in pathways involving in cell growth suggesting that there might be an advantage for gain-of-function mutations during spermatogenesis or fertilization. The average PAC was increased to 34.6 yrs in father ($P = 0.01$) and 31.1 yrs in mother ($p=0.03$) of probands with de novo variants associated with RASopathies. In summary, our study provides insights into the effects of advanced parental age on the occurrence of human de novo pathogenic variants as well as how the risks for monogenic diseases associated advanced paternal age can be better calculated for genetic counseling.

2787W

The SureMASTR BRCA Screen assay combined with MASTR Reporter analysis is an accurate and precise workflow for SNV, indel and CNV detection in blood- and FFPE-derived DNA. K. Bettens¹, B. Slabbinck², J. Vermeire¹, M. Vyverman², B. Tegenbos¹, L. Claes¹, L. Heyrman¹, H. Swennen¹, C. Marques¹, A. Rotthier¹, J. De Schrijver¹, D. Goossens¹, J. Crappé², J. Del Favero¹. 1) R&D, Agilent Technologies, Niel, Antwerp, Belgium; 2) R&D, Agilent Technologies, Ghent, Belgium.

Introduction: Targeted next-generation sequencing has tremendous potential to steer targeted therapy. Inhibitors of poly-(ADP-ribose) polymerase (PARP) have emerged as a new class of targeted anti-cancer drugs, specifically for tumors showing homologous recombination repair deficiency in ovarian and breast cancers. **Methods:** The SureMASTR BRCA Screen workflow comprises a sample to result solution. The library preparation is a single plex assay enabling mutation analysis of all coding regions of *BRCA1*, *BRCA2*, *PALB2* and *CHEK2*. The assay comprises 355 amplicons ranging in size from 80 to 224 bp. Data quality control as well as variant calling of SNPs, indels and CNVs are performed using MASTR Reporter. The SureMASTR BRCA Screen workflow was extensively examined in house and by 3 external clinical labs using both blood- and FFPE-derived DNA. **Results:** In total, 105 clinical samples were analyzed by three external clinical sites for germline mutation detection during a beta testing study. These samples harbored 114 annotated, mostly pathogenic variants, of which 8 in *CHEK2*, 14 in *PALB2*, 49 in *BRCA1* and 43 in *BRCA2*. The 114 variants included 8 CNV events and 49 indels up to 13 basepairs. All indels and SNVs were correctly called with the BRCA Screen workflow. In addition, all CNV events were correctly observed. The assay is optimized for cost-efficient use of the sequencing capacity obtained by (i) > 99% of the reads mapping to the target region and (ii) > 99.4% of the amplicons having a coverage within 20% of the mean coverage. Testing was performed on Illumina's MiSeq sequencer, and compatibility with Illumina's NextSeq and MiniSeq was demonstrated. FFPE-derived reference samples of Horizon were assessed for the quality of the DNA with the QC Plex assay. Uniformity of amplification remained high, with on average 99.1% of amplicons reaching a coverage of 20% of the mean coverage. A study with clinical FFPE derived samples is initiated to confirm these data. **Conclusions:** The SureMASTR BRCA Screen assay is a single plex assay which significantly reduces the hands-on time, DNA input, required sequencing capacity and in combination with the MASTR Reporter results in an accurate and precise workflow. Using the SureMASTR BRCA Screen workflow, we demonstrated that it can be routinely applied with a LOD down to 5% to detect somatic or germline *BRCA1*, *BRCA2*, *PALB2* and *CHEK2* mutations.

2788T

Molecular genetics study of hereditary colorectal cancer in Algerian patients. F. Cherbal¹, F. Khider¹, L.A Boumehdi¹, I. Tabet¹, S. Hamine¹, K. Layaïda², H. Mahfouf³, F. Zebboudj⁴, M. Maaoui⁵.

1) Unit of Genetics, LMCB, FBS Université des Sciences et de la Technologie Houari Boumediene, Algiers, Algeria; 2) Gastroenterology service, University Hospital Mustapha Bacha, Algiers, Algeria; 3) Mohamed El Kolli Public Hospital, Academic Medical Oncology Services, School of Medicine, University of Algiers, Rouiba, Algeria; 4) Mohamed El Kolli Public Hospital, Academic General Surgery Services, School of Medicine, University of Algiers, Rouiba, Algeria; 5) Bachir Mentouri Public hospital, Academic General Surgery Services, School of medicine, University of Algiers, Kouba, Algiers, Algeria.

Background Colorectal cancer is the second most common cancer in men and women in Algeria. The present study focuses on the two major forms of hereditary colorectal cancer, Familial Adenomatous Polyposis (FAP) and Lynch Syndrome (LS) in Algerian patients. Here, we analyzed by PCR-direct sequencing *APC* gene of 20 FAP patients and *MLH1*, *MSH2*, *MSH6* genes of 20 patients with LS, respectively. **Methods** The patients and their families were referred from 2012 to 2017 through the main medical oncology services and gastroenterology and general surgery services of Public Hospitals of Algiers. We screened by PCR-direct sequencing exon 15 of *APC* gene in 20 patients with a strong family history of FAP. 20 patients with Lynch syndrome who fulfilled Amsterdam criteria were screened by PCR-direct sequencing for germline mutations in the *MMR* genes: *MLH1* (exons 9,10,13,16), *MSH2* (exons 5,6,7,12) and *MSH6* (exon 4). **Results** The analysis of DNA samples of 20 FAP patients revealed that five patients carried pathogenic variants and one patient carried a VUS in *APC* gene. Four distinct pathogenic variants: c.3784delT, c.3905deT, c.3927_3931delAAAGA, c.4459dupA and the VUS p. Glu1317Gln were identified in *APC* gene. The *APC* c.3784delT and c.4459dupA pathogenic variants have never been reported in LOVD, UMD and ClinVar databases. These new *APC* germline variants could be specific of Algerian FAP patients. All the carriers of pathogenic variants in *APC* gene have developed colorectal carcinoma. Three distinct germline mutations: c.1852_1854delAAG in exon 16 of *MLH1*, c.1165C>T in exon 7 of *MSH2* and the Newfoundland mutation c.942 +3A>T in exon 5 of *MSH2* were detected in four Lynch syndrome patients. Interestingly, the c.942+3A>T *MSH2* mutation was found in two unrelated Lynch families. Moreover, the *APC* European founder mutation c.3927_3931delAAAGA has been also detected in two unrelated FAP families. Haplotype analysis of the *APC* mutation c.3927_3931delAAAGA and the c.942+3A>T *MSH2* mutation will establish if the origin of these two mutations in Algeria is linked to the common history and migration flow of Mediterranean countries. **Conclusion** We report here the first screening of *APC* gene in FAP patients and *MMR* genes in Lynch syndrome patients of Algerian national cohort. The accumulating knowledge about the prevalence and nature of *APC* variants and *MMR* genes variants will contribute in the future to the implementation of genetic testing and counseling for patients and families at risk.

2789F

Comparison of enzymatic fragmentation vs sonication for the production of cfDNA reference standards for liquid biopsy assay validation. *H. Child, A. Mele, K. Wilczynska, J. Wickenden.* Horizon Discovery, Cambridge, UK, Cambridgeshire, United Kingdom.

Introduction: Liquid biopsies hold great promise to revolutionise the field of clinical oncology testing. cfDNA can be extracted from a routine patient blood sample and used to determine the genetic profile of a solid tumour located elsewhere within the body. This facilitates more informative disease management for the clinician, without the need for invasive surgery for the patient. With new cfDNA NGS assays being able to detect variants from as little as 2-10ng DNA, assay validation to ensure sufficient accuracy has never been so critical. Reference materials that closely mimic real cfDNA samples are essential to support this effort. Here we present results from a comparative study into the production of cfDNA reference standards, assessing enzymatic fragmentation compared to fragmentation by sonication, and the benefit of including a size selection step. **Methods:** DNA was extracted from engineered cancer cell lines and fragmented to 160bp using either sonication or enzymatic shearing, as assessed by Tapestation. In addition, a size selection step was included to investigate the ability to purify the fragment peak to a size distribution profile most similar to real cfDNA samples. The presence of mutations was confirmed by ddPCR, allowing for accurate variant allele frequency to be determined. The sample set was tested by NGS on Thermo breast cfDNA assay, and Illumina TST-15 assay, following the manufacturer's instructions. **Results:** Variant detection by ddPCR confirmed the presence of 8 mutations across 4 genes (*EGFR*, *KRAS*, *NRAS* and *PIK3CA*) at either 0.1% or 5% variant allele frequency. Tapestation analysis confirmed that both sonication and enzymatic shearing produced cfDNA with an average fragment size of 160-170bp. A size selection step proved useful to concentrate the amount of DNA within the desired fragment size range. Analysis by NGS confirmed that all variants were detectable at the correct allele frequency as expected. **Conclusion:** Results demonstrate that both sonication and enzymatic shearing can be used to create cfDNA reference material with an average fragment size of 160bp to closely mimic cfDNA from real patient samples. This study has compared the utility of two alternative production methods, highlighting the potential of enzymatic shearing and a size selection step to further improve the commutability of high quality cell-line derived reference material for the validation of liquid biopsy assays.

2790W

The utilization of public population databases for somatic variant classification. *J.P. De La O, H. Cox, M. Jones, B. Coffee, L. Esterling, S. Manley.* Myriad Genetic Laboratories, Inc., Salt Lake City, UT.

Background: The presence of pathogenic variants in genes associated with homologous recombination (HR) may inform treatment decisions for some cancers. For example, the PARP inhibitor olaparib is FDA approved for women with ovarian cancer only if they have a pathogenic variant in *BRCA1* or *BRCA2*. The allele frequency (AF) of a variant in research grade population databases, such as gnomAD, is often used in variant classification by clinical testing laboratories, based on recommendations from ACMG/AMP. Specifically, variants that are present in these databases at an AF higher than the expected gene-associated disease prevalence may be classified as benign. In order to ensure accurate variant classification and appropriate treatment decisions, it is imperative that laboratories apply additional scrutiny to this research grade data. Here, the data quality from gnomAD was analyzed for HR related genes to assess the feasibility of using such data in variant classification.

Methods: Data quality from gnomAD was assessed for 6 genes associated with HR to identify high frequency variants that could be classified as benign. Quality metrics, including NGS read frequencies, mapping quality, pseudogene homology, and importantly somatic contamination, were analyzed. Variants of uncertain significance identified in the course of clinical testing were downgraded to benign if their AFs were above the ACMG/AMP recommended population specific AF thresholds ($\geq 5\%$ or an overall AF of $\geq 1\%$) and the available data passed all quality filters. **Results:** A total of 22 variants were identified in gnomAD that met the ACMG/AMP recommended population specific AF thresholds. Overall, low levels of somatic contamination were identified in the 6 HR genes evaluated here. In total, the gnomAD data for 8/22 (36%) variants failed one or more of our quality parameters. The remaining 14/22 (64%) variants were downgraded to benign. **Conclusions:** Here we found that 36% of variants in HR-associated genes that met ACMG/AMP AF thresholds failed the quality parameters instituted by our laboratory. The use of low quality data in population databases during variant classification may result in a false-negative result, preventing eligible patients from receiving appropriate treatment. This emphasizes the importance of appropriate laboratory quality filters in variant classification, even when using AF thresholds recommended by society guidelines.

2791T

Pululation screening for BRCA1 and BRCA2 mutations: Results from a feasibility study within a military treatment facility. *R.E. Ellsworth¹, L.A. Lovejoy², C.D. Shriver^{1,3}*. 1) Clinical Breast Care Project, Murtha Cancer Center, Bethesda, MD; 2) Chan Soon-Shiong Institute of Molecular Medicine at Windber, Windber, PA; 3) Uniformed Services University, Bethesda, MD.

Background: Although strategies have been developed to decrease risk of developing breast and ovarian cancer in women with germline mutations in BRCA1 or BRCA2, genetic testing in the United States (US) is offered only to women with a significant family history, thus excluding carriers that do not meet testing criteria. Identification of carriers only after a cancer diagnosis may be considered a failed opportunity for prevention and implementation of a population screening program at or around age 30 has been proposed to identify high-risk women in the US. In this pilot study, we evaluated the yield and utility of population screening within a cohort of women with health-care provided through a military treatment facility. **Methods:** BRCA status was determined for 368 women age 18–40 years who had benign biopsies at Walter Reed National Military Medical Center and who had genomic (blood) DNA available for research purposes. 78% of the women tested did not meet the NCCN guidelines for genetic risk evaluation. BRCA1 and BRCA2 were sequenced using next-generation sequencing technologies and variants were classified as pathogenic, likely pathogenic, variants of uncertain significance (VUS), likely benign or benign, according to ACMG criteria. **Results:** Pathogenic mutations were detected in 2% of women (BRCA1 n=5 and BRCA2 n=3); an additional three and 12 women harbored novel mutations and VUS, respectively. Two of the patients with pathogenic mutations did not have a significant family history and would not have been eligible for genetic testing under current guidelines. **Conclusions:** These data have important clinical. Based on 2010 US Census data, there were 41,141,300 women age 20–39; extrapolation from our pilot data suggests that >800,000 women in that age range harbor BRCA1/2 mutations including 223,000 who would not qualify for testing. Identification of women with germline mutations before the development of breast cancer will allow risk-reduction strategies to be implemented, improving the health of these women while decreasing medical costs. The opinions or assertions contained herein are the private ones of the author/speaker and are not to be construed as official or reflecting the views of the Department of Defense, the Uniformed Services University of the Health Sciences or any other agency of the U.S. Government.

2792F

Searching for variants by whole-exome sequencing in Brazilian women at high-risk for hereditary breast/ovarian cancers, negative for mutations in BRCA1/BRCA2 genes. *P.S. Felicio¹, R.S. Grasel¹, B.D.F. Barros², N. Campacci¹, H.C.R. Galvão³, C.P. Souza³, A.F. Evangelista¹, D.M. Carraro², E.I. Palmero¹*. 1) Molecular Oncology Research Center, Barretos Cancer Hospital, Barretos, São Paulo, Brazil; 2) Laboratory of Genomics and Molecular Biology, CIPE - A. C. Camargo Cancer Center, São Paulo, Brazil; 3) Department of Oncogenetics, Barretos Cancer Hospital, Barretos, São Paulo, Brazil.

Background: BRCA1 and BRCA2 testing is uninformative in approximately 80% of families with clinical features of inherited susceptibility to breast and ovarian cancers. Early identification of high-risk individuals may lead to improvements in screening and prevention strategies, and, potentially, improving overall cancer survival. **Aim:** To identify the missing heredity associated to cancer development in high-risk families, BRCA1/2 WT (BRCAx families), through whole-exome sequencing (WES) of constitutive DNA. **Methods:** We performed WES (Nextera Rapid Capture Exome and Expanded Exome Kit) in 45 BRCA1/2 WT individuals with high-risk for breast/ovarian cancer. The criteria for data analysis were: (i) coverage $\geq 10\times$, (ii) variants present in less than 15% (≤ 7 samples) and (iii) missense, in-frame and stop loss variants classified as pathogenic by at least three of six *in silico* prediction programs selected. **Results:** For each sample, an average of 122 genetic changes per sample (ranging from five to 196 variants) was identified. In addition, most of the identified variants were of the missense type, followed by the nonsense variants. Among the twenty most commonly altered genes, many of them have already been associated with hereditary predisposition syndromes to cancer, such as: *ATM*, *MLH1*, *MSH6*, *MUTYH* genes. Besides, pathogenic/probably pathogenic alterations were identified in DNA repair genes, such as *ERCC3*, *ERCC4*, *EXO1* and *FAN1* genes were identified. **Conclusion:** Our data show that significant genetic heterogeneity exists in BRCAx families. Large-scale collaborative efforts will be required to attain sufficient power to understand the missing heredity associated with breast/ovarian cancer development in high-risk families.

2793W

Circulating tumor DNA methylation haplotypes in plasma can accurately detect early-stage colorectal cancer and advanced adenomas in a single blinded study. J. Gole¹, A. Gore¹, J. Min¹, H. Wang¹, J. Dang¹, H. Shi¹, J. Lu¹, S. Mo², X. Lan³, W. Dai², R. Wang², Y. Gao¹, K. Zhang¹, L. Liang², G. Cai², R. Liu¹. 1) Singlera Genomics, La Jolla, CA; 2) Fudan University Shanghai Cancer Center, Shanghai, China; 3) Southern Medical University, Guangzhou, China; 4) University of California - San Diego, La Jolla, CA.

Background: Colorectal cancer (CRC) is one of the most common forms of cancer and is responsible for approximately 700,000 deaths per year worldwide. Because of this, colorectal cancer screening is recommended by the USPSTF, but patient compliance can be low due to the invasiveness of screening colonoscopy. While a non-invasive circulating tumor DNA (ctDNA) methylation screening test has been developed using the Septin9 gene, the sensitivity and specificity of the test are inadequate for use as a primary screening tool (especially for early-stage CRC). We set out to develop a highly accurate, non-invasive plasma based screening assay to identify colorectal cancer at an early stage using ctDNA methylation signatures. **Methods:** Blood samples from 88 healthy individuals, 150 individuals diagnosed with advanced adenoma by colonoscopy, and 526 individuals diagnosed with colorectal cancer by colonoscopy were collected in EDTA tubes and immediately separated into plasma. Plasma samples were then processed using the Singlera Genomics ACE assay, a targeted bisulfite sequencing method which identifies methylation haplotype patterns specific to early- and late-stage colorectal cancer. 217 samples were used to train a classification model based on colonoscopy results, and 218 samples were used as a test set to validate the model. An additional 329 samples were held out as a single blinded independent test set to further confirm the model accuracy and robustness. **Results:** In the single blinded independent test set, the ACE assay was able to show a sensitivity of 88.4% in advanced adenoma patients and 93.1% in colorectal cancer patients, with a specificity of 96.5% in healthy patients. **Conclusions:** We have shown that ctDNA methylation can be utilized to non-invasively screen for early-stage colorectal cancer and advanced adenoma with best in class sensitivity, paving the way for a primary blood-based colorectal cancer screening assay.

2794T

Multi-gene panel testing 54 breast cancer predisposition genes. C. Hata¹, H. Nakaoka¹, L. Wei², K. Zheng², H. You², I. Inoue¹. 1) National Institute of Genetics, Mishima, Japan; 2) Affiliated Cancer Hospital & Institute of Guangzhou Medical University, Guangzhou, China; 3) Department of Endocrine and Breast Surgery, The First Affiliated Hospital of Chongqing Medical University, Chongqing, China.

Breast cancer is the most common noncutaneous malignancy in woman and approximately 10-15% of cases are associated with germline mutations of DNA repair genes. In this study, we scrutinized germline mutations of 54 cancer predisposition genes including *BRCA1/2* in 672 breast cancer patients of Chinese origin. DNA samples, extracted from peripheral blood samples, were subjected to NGS sequencing based on a pool and capture method established by our previous study. After the quality control step, the sequence reads were aligned to the human reference genome (hg19). The aligned reads were processed for removal of PCR duplicated and erroneous reads by Picard tools. The local realignment and base quality recalibration conducted by GATK. Single nucleotide variants (SNVs) and insertions and deletions (indels) were detected with the GATK. Variants were classified according to pathogenicity as follows: First, variants were classified as 'benign' if their frequencies were greater than 1.0% in the public database in the 1000 Genomes or ExAC project. Second, nonsense, splice site mutations, and indels were classified as 'pathogenic'. Third, we explored missense variants with previously established pathogenic or benign effects based on ClinVar. Variants were classified as 'conflicting interpretations of pathogenicity' if reports were conflicted on ClinVar. Finally, variants with 'uncertain significance' were classified as 'VUS'. Variants without reliable information were classified 'NA'. We detected 81 pathogenic mutations (34 nonsense, 37 frameshift indels and 10 splice site mutations) in 672 breast cancer patients, accounting for the 12% causality. Furthermore, 5.5% of the patients harbored *BRCA1/2* mutations (37/672). The remaining 7% of the patients carried pathogenic mutations on 23 genes (44/672). Interestingly, we detected a novel deletion-insertion mutation on *PTEN*, where 69bp of mitochondrial DNA sequence was inserted into 46bp deletion site in the 1st exon of *PTEN*. Finally, we identified significant associations between mutation status and clinical variables. Taken together, these results suggest that screening of pathogenic mutations in breast cancer patients based on the multi-gene panel is clinically useful.

2795F

An analysis of tumor mutation burden in colorectal cancer has revealed signature damage patterns associated with multiple cancer types. J. Huang¹, R. Chaudhary¹, V. Mittal², W. Tom¹, F. Shen¹, J. Au-Young¹, F. Hyland¹.

1) Clinical Next-Generation Sequencing Division, Thermo Fisher Scientific, South San Francisco, CA; 2) Clinical Next-Generation Sequencing Division, Thermo Fisher Scientific, Ann Arbor, MI.

Introduction: Tumor mutation burden (TMB) is defined as the rate of somatic mutations within a tumor genome and has been shown in several cancer types to be a promising genomic biomarker that effectively indicates potential for successful immunotherapy. TMB analyses can reveal tumorigenesis signature patterns associated with specific damages through substitution type and context of somatic mutations. **Methods:** We estimated mutation burden (mutation count per Mb) from 50 FFPE tumor research samples of colorectal cancers (CRC), melanoma, and non-small cell lung cancer (NSCLC) with OncoPrint™ Tumor Mutation Load Assay, a PCR-based target enrichment panel that covers 409 cancer-related genes and 1.7 Mb of genomic regions. The workflow requires only 20 ng of input tumor DNA (no matched normal sample required). Sequencing was performed on Ion GeneStudio with ~500x coverage. Germline variants, along with background noise, are removed through application of filters based on variants prevalent in population databases.

Results: Initial results for a cohort of CRC FFPE tumors demonstrated that most of the samples had a high TMB and mutations in mismatch repair (MMR) genes. Further analysis revealed a series of signatures of tumorigenesis based on the substitution type and genomic context of somatic mutations. In addition, signatures for FFPE deamination were observed for quality evaluation. Characteristic signature patterns have been observed in multiple tumor types, including deamination of 5-methylcytosine in CRC tumors, tobacco damage in lung tumors, and UV damage in melanoma tumors. The proportion of respective signatures to the total somatic mutations was quantified. In a CRC tumor, 43.6% C:G>T:A mutations were observed at CpG sites representing spontaneous deamination of 5-methylcytosine. In a melanoma tumor, we found two novel signatures: high C:G>T:A at TpC, high C:G > T:A at CpC and CpC sites, totaling 88.6% mutations, consistent with DNA damage signatures of ultraviolet radiation exposure. Finally, lung tumor samples showed 48.3% C:G>A:T mutations consistent with tobacco damage. **Conclusions:** Tumor mutation burden analyses can be used to detect somatic mutations and reveal signature patterns correlating with the genomic damage characteristic of specific tumor types. This analysis will aid in elucidating properties of the tumor corresponding to known mechanisms of tumorigenesis and identify additional variants of different tumor origin.

2796W

Clinical application of the OncoKidsSM next generation sequencing platform for detection of germline and somatic alterations in pediatric brain tumors. J. Ji^{1,2}, K. Kaneva^{3,4}, M.H. Hiemenz^{1,2}, R.J. Schmidt^{1,2}, A. Margol^{3,4}, N. Robison^{3,4}, G. Dhall^{3,4}, D. Tian^{1,2}, D. Hawes^{1,2}, J.A. Cotter^{1,2}, A.R. Judkins^{1,2}, J.A. Biegel^{1,2}. 1) Department of Pathology and Laboratory Medicine, Children's Hospital Los Angeles, Los Angeles, CA; 2) Department of Pathology, University of Southern California; 3) Department of Pediatrics, Children's Hospital Los Angeles, Los Angeles, CA; 4) Department of Pediatrics, University of Southern California.

We developed a comprehensive DNA- and RNA-based next generation sequencing (NGS) panel, OncoKidsSM, to detect diagnostic, prognostic, and therapeutic markers across the spectrum of pediatric malignancies, including leukemias, solid tumors, and brain tumors. To date, 100 pediatric brain tumors have been prospectively analyzed using OncoKidsSM, including low and high-grade gliomas, choroid plexus tumors, embryonal tumors, and ependymomas. OncoKidsSM demonstrated pathogenic mutations in 64% (64/100) of cases. A total of 88 SNVs and small InDels of strong clinical significance (Tier I) or potential clinical significance (Tier II) in 29 genes were reported in 53 cases. Diagnostic and prognostic Tier I mutations included mutations in *H3F3A*, *HIST1H3B*, *FGFR1*, *ACVR1*, and *TP53* in high-grade gliomas; *BRAF* V600E, *FGFR1*, and *NF1* mutations in low grade gliomas; *KDM6A*, *PTCH1*, *SMARCA4*, and *TP53* mutations in medulloblastomas; *CTNGB1* exon 3 mutations in craniopharyngiomas; and *SMARCB1* mutations in rhabdoid tumors. A combination of *H3F3A* G34R, *ATRX*, and *TP53* Tier I mutations which characterizes a specific subtype of high-grade glioma was identified in four histologically challenging neuroepithelial tumors. Clinically significant RNA fusions were identified including *KIAA1549-BRAF* in 11 pilocytic astrocytomas, *PTPRZ1-MET* in two high-grade gliomas, *YAP1-MAML2* in one malignant neoplasm of the posterior fossa without further histologic classification, and *FGFR2-SHTN1* in one ganglioglioma. Amplifications of *EGFR*, *MYCN*, *MET*, and *MDM4* were detected in three high-grade gliomas, including one patient with amplification of *EGFR*, *MYC*, and *MDM4*. A high frequency of mutations was observed in two high grade-gliomas, consistent with the high tumor mutation burden that is seen in association with mutations in mismatch repair genes. One patient had a bi-allelic germline mutation in *MSH2* and the other patient had a bi-allelic germline mutation in *MSH6*. Potential targeted therapies with BRAF, FGFR, mTOR, MET, IDH1 or ALK inhibitors, were suggested in 34% (22/64) of cases with at least one Tier 1 variant. OncoKidsSM analysis also revealed potential germline mutations in cancer predisposition genes for 26% (26/100) of cases, including one patient with possible germline mosaicism for an *FGFR1* mutation. Our results demonstrate significant clinical utility of integrating NGS-based genomic profiling into routine clinical testing for pediatric brain tumors.

2797T

The Veterans Health Administration's Research Precision Oncology**Project: Integrating real-world data into a learning health system.** *B.R.**Johnson^{1,2,4}, D.C. Elbers¹, E. Pierce¹, F. Meng^{1,3,4,5}, N. Fillmore^{1,2,4}, S. Ayandeh¹, D. Chen³, L. Selva¹, N.B. Begley¹, A.J. Zimolzak^{1,3,4}, M.T. Brophy^{1,3,4}, N.V. Do^{1,3,4}.**1) MAVERIC- VA Boston, Boston, MA; 2) University of Buffalo, School of Medicine, Buffalo NY; 3) Boston University School of Medicine, Boston, MA; 4) BD-STEP program, NCI/VA, Washington DC; 5) UCLA, Los Angeles CA.*

The Veterans Health Administration's (VHA) "Veterans Integrated Service Network 1" collected targeted molecular diagnostic oncology panel data from multiple vendors as the standard of care. The program yielded 1967 reports from patients with advanced cancer from 2015 to December 2017. Notable among these reports, 785 (40%) of these patients had a lung cancer diagnosis, 140 (7%) had liver cancer and 101 (5%) were determined to have prostate cancer. So far, 113 of these patients have consented to have their data shared under the IRB approved Research Precision Oncology Program (RePOP) protocol. Patients were enrolled from 26 facilities throughout the VHA. Data has been aggregated for visualization from multiple streams, including; molecular diagnostics from multiple vendors and radiology. Over 90% of these consented patients were determined to have at least 1 of 327 unique clinically actionable variants. There were 27 (24%) patients with *KRAS* mutations found in the consented population. The large majority (>95%) of patients with lung cancer within the cohort were diagnosed with Non Small Cell Lung Carcinoma (NSCLC), with an almost even split among adenocarcinoma and squamous histologies. The VHA contains a wide array of data that may be utilized for precision medicine, predictive models and in creating a learning healthcare system. To reach this goal, we are preparing the consented cohort of data derived from REPOP for ingestion into the Precision Oncology Data Repository. This effort is making VHA oncology data more widely available to the research community. Internally, we are analyzing patterns of survival and treatment options to better understand current treatment modalities and outcomes. Additionally, Non-VHA researchers will be able to gain access for the establishment of predictive models and research purposes.

2798F

Clinicopathological and molecular analysis of 45 cases of mucinous**breast cancer.** *SY. Kang¹, MS. Ahn¹, YS. Jung², TJ. Park³, SH. Park⁴, JH.**Kim⁴, HE. Yim⁴.* 1) Hematology-Oncology, Ajou University School of Medicine, Suwon, South Korea; 2) Department of Surgery, Ajou University School of Medicine, Suwon, South Korea; 3) Department of Biochemistry and Molecular Biology, Ajou University School of Medicine, Suwon, South Korea; 4) Department of Pathology, Ajou University School of Medicine, Suwon, South Korea.

Purpose: Mucinous carcinoma of the breast is characterized by clusters of tumor cells floating in abundant extracellular mucin and can be divided into two main subtypes based on architectural and cytologic features: paucicellular (classic) and hypercellular variant. However, clinicopathological and molecular differences between these two variants has far not been investigated.

Materials and Methods: We retrospectively evaluated clinicopathologic features of 45 cases of surgically removed mucinous carcinoma of the breast at Ajou university hospital from April 2011 to April 2017. Among them, 8 cases were microdissected and subjected to whole exome sequencing for molecular analysis. SNPs and Indels were called from every alignment files by using HaplotypeCaller with GVCF mode which covers not only variable sites, and also same with reference genotype, and then all genomic vcf files were merged by using CombineVCF. To compared the allele frequencies between normal individual and the 8 breast cancers, we used Korean Reference Genome Common Variants (KRGCV) of 1,100 Korean people. We further analyzed the differences of the frequencies between two subtypes of mucinous breast cancer. **Results:** Of the 45 patients who underwent surgery, 30 were paucicellular subtypes and 15 were hypercellular subtypes. We found lymph node metastasis in 5 cases of hypercellular subtypes but no in paucicellular subtypes. Hormone receptors were positive in both groups and 2 patients were HER2 positive only in hypercellular subtypes. All patients were alive at the time of analysis and one of the hypercellular subtypes underwent reoperation due to local recurrence. By whole exome sequencing analysis, we found 1,523 SNPs and 812 INDELs significantly different between mucinous breast cancers and reference Genome Data Base (KRGCV). Among them, 133 SNPs were significantly different between two variants of mucinous carcinoma.

Conclusion: Our results suggested that hypocellular and hypercellular variant show different patterns of SNPs and hypercellular variant maybe associated with more frequent lymph node metastasis than paucicellular subtypes.

2799W

BRCA1/2 mutations, including large genomic rearrangements, among unselected ovarian cancer patients in Korea. D. Kim¹, N. Ryoo¹, D. Jeon¹, W. Lee¹, J. Ha¹, C. Cho², S. Kwon³. 1) Department of Laboratory Medicine, Keimyung University School of Medicine, Daegu, South Korea; 2) Department of Obstetrics and Gynecology, Keimyung University School of Medicine, Daegu, South Korea; 3) Department of Pathology, Keimyung University School of Medicine, Daegu, South Korea.

We performed small-scale mutation and LGR analysis without selection in multiple ovarian cancer subtypes. We also evaluated the usefulness of several mutation carrier prediction algorithms for predicting *BRCA1/2* mutations, including LGRs. Here, we report the results of a comprehensive *BRCA1/2* genetic study that determined the prevalence and the characteristics of mutations among unselected ovarian cancer patients. All ovarian cancer patients who visited a single institution between September 2015 and April 2017 were included. Sanger sequencing, multiplex ligation-dependent probe amplification (MLPA), and long-range polymerase chain reaction (PCR) were performed to comprehensively study *BRCA1/2*. The genetic risk models BRCAPRO, Myriad, and BOADICEA were used to evaluate the mutation analysis. In total, 131 patients were enrolled. Of the 131 patients, Sanger sequencing identified 16 different *BRCA1/2* small-scale mutations in 20 patients (15.3%). Two novel nonsense mutations were detected in two patients with a serous borderline tumor and a large-cell neuroendocrine carcinoma (LCNC). MLPA analysis of *BRCA1/2* in Sanger-negative patients revealed two LGRs. The LGRs accounted for 14.3% of all identified *BRCA1* mutations, and the prevalence of LGRs identified in this study was 1.8% in 111 Sanger-negative patients. The genetic risk models showed statistically significant differences between mutation carriers and non-carriers. These results partially support a previous study that showed a positive correlation between the proportion of mutations due to a LGR and the BRCAPRO score. The two patients with LGRs had at least one blood relative with breast or ovarian cancer. Twenty-two (16.8%) of the unselected ovarian cancer patients had *BRCA1/2* mutations that were detected through comprehensive *BRCA1/2* genetic testing. Patients with ovarian borderline tumors with aggressive characteristics and LCNC of the ovary should be considered for *BRCA1/2* genetic testing. Additionally, an above-average score on the BRCAPRO, Myriad, or BOADICEA assessments may also indicate the need for LGR testing in Sanger-negative patients. Ovarian cancer patients with Sanger-negative results should be considered for LGR detection if they have one blood relative with breast or ovarian cancer. The detection of more *BRCA1/2* mutations in patients is important for efforts to provide targeted therapy to ovarian cancer patients.

2800T

An amplicon panel for BRCA sequencing bundled with a user-friendly bioinformatics solution. K. Kushiro, A. Sharma, D. Fox, S. Piehl, A. Kothandaraman. Bioo Scientific, a PerkinElmer company, 7050 Burleson Rd Austin, TX 78744.

More and more labs are exploring next-generation sequencing for clinical research applications as adoption is being fueled by the prolific growth of personalized medicine. The identification of mutations in *BRCA1* and *BRCA2* tumor suppressor genes is a staple in many clinical research projects and as such, amplicon panels have proven to be very versatile and economical solutions to understand these genes in human cancers. Panels and bioinformatic pipelines have been designed to construct libraries from fresh and frozen tissues or FFPE samples, both of which are popular sample types for labs. The panel for *BRCA1*, *BRCA2*, *PALB2*, and *CHECK2* using fresh and frozen tissue is composed of 187 amplicons ranging from 92-235bp in size, not including primer regions, and exhibits 98% uniformity at 0.2x coverage and 100% uniformity at 0.1x coverage. The panel for *BRCA1* and *BRCA2* from FFPE tissue is composed of 261 amplicons ranging from 64-129bp in size, not including primer regions, and exhibits 97% uniformity at 0.2x coverage and 100% uniformity at 0.1x coverage. Aside from reliable and optimized library prep at the bench, amplicon panels facilitate a naïve user's transition into adopting next-generation sequencing as it also allows for a much narrower and targeted approach that also minimizes the amount of data generated compared to wider sequencing applications like DNA-Seq or hybrid-capture methods. This poster reviews the workflow, performance and sequencing metrics, and the bioinformatics analysis and output associated with these chemistries for amplicon panels. Research use only. Not for diagnostic procedures.

2801F

Validation of the GeneReader NGS system for *BRCA 1* and *BRCA 2* genotyping. E. Lee^{1,2}, N. Park², W. Lee², S. Chun², W. Min². 1) Hallym University College of Medicine, Hwaseong-Si, Korea, Republic of; 2) University of Ulsan College of Medicine and Asan Medical Center, Seoul, Korea, Republic of.

Background: *BRCA1/2* testing with NGS technology offers many advantages over Sanger sequencing, including high throughput, and automated analysis. However, multiple barriers still exist for its broader adoption in clinical laboratory, such as fragmented workflow and complex bioinformatics analysis. GeneReader NGS system includes all sample processing steps starting from nucleic acid extraction to an integrated bioinformatics solution that allows for direct access to real-time updates from clinical knowledge and evidence. We evaluated performance of the GeneReader NGS Platform for *BRCA 1* and *BRCA 2* genotyping. **Methods:** *BRCA1* and *BRCA2* sequence determination using GeneReader was performed in a total of 100 patients who were diagnosed breast or ovarian cancer and previously tested *BRCA1* and *BRCA2* analysis using Sanger sequencing. Libraries (253 amplicons covering 37.5 kb) were prepared using QIAGEN Library Kit. Emulsion PCR and bead enrichment steps were performed using GeneReader QIAcube and Clonal Amp Q Kit. All the procedures were performed according to manufacturer's protocols. QCI Analyze software performed read alignment, quality control, variant calling, and clinical report generation. The results generated by GeneReader were compared to those obtained by Sanger sequencing. For reproducibility of variant calling, we performed the process in triplicate using 28 samples and analyzed the results. We reviewed turnaround time (TAT) of total process and its hands on time. **Results:** Greater than 99.9% of the target regions showed read depths >100x and the percentage of reads in on-target regions was 90.5%. Variants generated from GeneReader achieved 100% (13/13) and 98.9% (722/730) concordance with those generated from Sanger sequencing in insertion/deletions (indels) and single nucleotide variations (SNVs), respectively. The annotation of the GeneReader system showed over 99.9% (744/745) concordance with databases in ClinVar. All of 272 SNVs and 7 indel variants showed 100% calling reproducibility. The GeneReader provided a truly integrated workflow and most process was composed of hands off time (3250 min, 87% of total TAT). **Conclusions:** The GeneReader NGS system for *BRCA 1* and *BRCA 2* genotyping showed great performance and takes much shorter hands on time compared with Sanger sequencing. With a full end-to-end solution with integrated sample preparation to bioinformatics interpretation, the GeneReader NGS System can be useful for clinical laboratory practices.

2802W

Optimization and validation of the ACMG/AMP variant interpretation guidelines recommended by ClinGen's Myeloid Malignancy Expert Panel.

X. Luo^{1,2}, J. Ross³, S. Jackson⁴, A. Bertuch^{1,2}, A. Brown⁷, A. Raimbault⁶, A. Agarwal⁸, C. Kesserwan¹⁰, C. Porter¹¹, C. DiNardo¹², G. Ryan¹³, J. Soulier⁸, K. Nichols¹⁰, L. Rawlings⁷, L. Malcovati¹⁴, M. Fleming¹⁵, M. Routbort¹², M. Chicka¹⁶, M. Walsh¹⁷, M. Luo¹⁸, N. Speck¹⁹, P. Baliakas²¹, S. McWeeney⁹, S. Keel⁴, T. Vulliamy²⁰, Y. Wang⁶, Z. Li³, S. Plon^{1,2}, D. Wu⁴, L. Godley⁹. 1) Baylor College of Medicine, Houston, TX; 2) Texas Children's Hospital, Houston, TX; 3) University of Chicago, Chicago, IL; 4) University of Washington, Seattle, WA; 5) University of North Carolina at Chapel Hill, Chapel Hill, NC; 6) GeneDx, Gaithersburg, MD; 7) University of South Australia, Adelaide, Australia; 8) University of Paris Diderot, Paris, France; 9) Oregon Health & Science University, Portland, OR; 10) St. Jude Children's Hospital, Memphis, TN; 11) Emory University, Atlanta, GA; 12) Univ of Texas MD Anderson Cancer Center, Houston, TX; 13) American Society of Hematology, Washington, DC; 14) University of Pavia, Pavia PV, Italy; 15) Boston Children's Hospital, Boston, MA; 16) Prevention Genetics, Marshfield, WI; 17) Memorial Sloan Kettering Cancer Center, New York, NY; 18) Children's Hospital of Philadelphia, Philadelphia, PA; 19) University of Pennsylvania, Philadelphia, PA; 20) Queen Mary College London, London, UK; 21) Uppsala university hospital, Uppsala, Sweden.

Clinical Genome Resource (ClinGen) is a National Institutes of Health (NIH)-funded resource dedicated to building an authoritative central resource that defines the clinical relevance of genes and variants for use in precision medicine and research. ClinGen has developed expert panels in different clinical domains to adapt the American College of Medical Genetics and Genomics/Association for Molecular Pathology (ACMG-AMP) guidelines for consistent and accurate variant classification of specific genes and diseases. Eighteen expert panels has been established to date under ClinGen and two of them (Cardiomyopathy and RASopathy) has published their recommendations of disease-specific variant classification rules. A new Myeloid Malignancy Expert Panel has recently started to focus on curating variants in genes associated with inherited risk for myeloid malignancies. This panel consists of expert clinicians, clinical laboratory diagnosticians, and researchers who enlist representatives from community-organized efforts to implement standardized protocols for sequence variant specific annotations of genes related to the specific disease domain. The optimization of the ACMG/AMP guidelines represent disease-/gene-informed specifications or strength adjustments of existing rules, including gene-specific population frequency cutoffs, recommendations for the use of computational/predictive data, functional data, and clinical evidence. Some rules will be deemed not applicable. This expert panel will consider variants in genes that cause life-long thrombocytopenia (e.g., RUNX1, ANKRD26, and ETV6) as well as those associated only with cancer risk (e.g., DDX41 and CEBPA). In summary, ClinGen Myeloid Malignancy Expert Panel provide recommendations to optimize ACMG/AMP criteria using for standardization of variant interpretation practices. This effort will facilitate more knowledgeable utilization of genomic variants in clinical diagnostics and cancer research, and provide a strong foundation for genomic medicine.

2803T

Multi-gene testing of an Indian ovarian cancer cohort: Prevalence of BRCA1/2 and non-BRCA mutations. A.U. Mannan, J. Singh, S. Deshwal, S. Sur, N. Thota, A.V. Kanthi, P. Preeti, K. Saxena, S. Sankaran, S. Katragadda, V. Veeramachaneni, R. Hariharan. Strand Life Sciences, Bangalore, Karnataka, India.

Ovarian cancer is one of most commonly diagnosed hereditary cancer and leading cause of deaths among gynecologic malignancies in India. This emphasizes on the need for a cost-effective method for early detection and prevention of the cancer. NGS (next-generation sequencing) based tests provide an attractive alternative to the traditionally employed sequential testing strategies as they allow massively parallel sequencing of multiple genes associated with a cancer efficiently and at a significantly lower cost as compared to the traditional methods. We sequenced 519 unrelated patients and families affected with ovarian cancer using the TruSight Cancer panel, which includes 11 genes strongly associated with risk of inherited ovarian cancer. Multi-gene sequencing was done on the Illumina MiSeq/NextSeq platform. Genetic variations were identified using the Strand NGS software and interpreted using the StrandOmics platform. We were able to detect pathogenic mutations in 38.7% (201/519) cases. In familial ovarian cancer cases, the detection rate increased to 52% (117/223) and in sporadic cases, the diagnostic yield was 28.4% (84/296). Overall distribution of mutations showed that 89.6% were detected in the *BRCA1/2* genes and 10.5% were detected in the non-BRCA genes. Among the non-BRCA genes, 2% of mutations were detected in the MMR genes and 3.5% in the *RAD51C/D* genes. We also identified 2 *BRCA1* recurrent mutations (likely to be founder mutations) in 41 cases in our cohort, which accounts for 20.4% of the total mutations. In our cohort, among the 201 mutations detected in ovarian cancer patients, the majority (89.6%) were in the *BRCA1/2* genes and they are likely to benefit from PARP [poly (ADP-ribose) polymerase] inhibitors as targeted therapy for treatment of cancer. Our study suggests that NGS based multi-gene panel increase the sensitivity of mutation detection and help in identifying patients with a high risk of developing cancer as compared to sequential tests of individual genes. Incorporation of such cost-effective and comprehensive test as a first-tier test will be helpful in optimizing resources for screening, prevention, treatment and disease management measures in the clinical practice for hereditary ovarian cancer.

2804F

Pathogenic germline variants in a pediatric cancer cohort and identification of new candidate cancer predisposition genes. K.E. Miller¹, D.C. Koboldt^{1,2}, B.J. Kelly¹, P. Brennan¹, V. Magrini^{1,2}, J.M. Gastier-Foster^{1,3}, P. White^{1,2}, E.A. Varga^{1,4}, C.E. Cottrell^{1,3}, R.K. Wilson^{1,2}, E.R. Mardis^{1,2}. 1) Institute for Genomic Medicine, Nationwide Children's Hospital, Columbus, OH, USA; 2) Department of Pediatrics, The Ohio State University College of Medicine, Columbus, OH, USA; 3) Department of Pathology and Laboratory Medicine, Nationwide Children's Hospital, Columbus, OH, USA; 4) Division of Hematology/Oncology/Bone Marrow Transplantation, Nationwide Children's Hospital, Columbus, OH, USA.

Previous studies report the frequency of pathogenic germline variants in cancer predisposition genes to be ~8-10% in pediatric cancer populations. Identification of new susceptibility genes is essential, as there are undoubtedly many genes that remain to be recognized as cancer-predisposing. We applied exome sequencing to a pediatric cancer cohort at Nationwide Children's Hospital (Columbus, OH, USA) to search for germline variants in *bona fide* cancer genes and to identify putative cancer predisposition genes. Our cohort (N = 45) included 38 patients with central nervous system (CNS) tumors (e.g., pilocytic astrocytoma, ganglioglioma), 5 patients with non-CNS solid tumors (e.g., osteosarcoma, rhabdomyosarcoma), and 2 patients with hematological cancers. Most patients were pediatric with median age at diagnosis 10.0 years (range: 6 months to 24 years). For each patient, we performed exome sequencing on both tumor (mean depth ~194x) and germline (mean depth ~215x) samples and used SnpEff plus custom in-house scripts to annotate single nucleotide variants (SNVs)/indels with gene, transcript, molecular consequence, population frequency, and damaging prediction scores. SNVs/indels were assessed for pathogenicity using the current ACMG/AMP clinical variant interpretation guidelines. Six patients (13.3%) harbored pathogenic germline variants in well-established cancer predisposition genes *TP53* (3 patients), *NF1* (2 patients), and *PALB2* (1 patient). All 3 patients harboring a *TP53* germline variant had a diagnosis of choroid plexus carcinoma; both patients harboring an *NF1* germline variant had a diagnosis of CNS astrocytoma. For the remaining 39 patients, we examined rare damaging or loss-of-function variants to identify possible predisposition alleles. Four patients harbored rare, likely damaging missense variants in previously implicated cancer genes *PMS2* (1 patient), *SMARCA4* (2 patients), and *ALK* (1 patient). To identify potential novel predisposition genes, we searched for genes in which at least two patients harbored a very rare, damaging or loss-of-function variant, which nominated four new candidate genes involved in chromatin remodeling / DNA repair (*KAT6B* and *POL1*, respectively) or cell signaling (*TXK* and *PTPRF*). Identification of new cancer predisposing genes and detection of germline pathogenic variants may increase our understanding of tumorigenesis, influence clinical management, and prompt further genetic testing for the patient and family members.

2805W

Cancer mutations in healthy admixed elderly: Can we improve pathogenicity interpretation? M.S. Naslavsky^{1,2}, G.L. Yamamoto^{1,3}, S.A.M. Ezquina¹, Y.A.O. Duarte^{4,5}, M.R. Passos-Bueno¹, M. Zatz¹. 1) Human Genome and Stem Cell Research Center, University of São Paulo, São Paulo, Brazil; 2) Hospital Israelita Albert Einstein, São Paulo, Brazil; 3) Department of Clinical Genetics, Children's Hospital, Medical School, University of São Paulo, São Paulo, Brazil; 4) Department of Epidemiology, Public Health School, University of São Paulo, São Paulo, Brazil; 5) School of Nursing, University of São Paulo, São Paulo, Brazil.

Population-based whole genome sequencing (WGS) initiatives have been taken in different countries and tens to hundreds of thousands of individuals are being studied without a prior clinical indication. While providing insights into novel biological mechanisms, ancestry and evolution, the actionability in public health and patient-oriented measures is still limited. Incidental findings in undiagnosed patients or normal controls impose an important diagnostic challenge particularly in large population-based projects with limited clinical and phenotypic information as well as follow-up and tracking-back obstacles. The American College of Medical Genetics and Genomics (ACMG) recommends reporting back to sequenced individuals pathogenic variants in specific genes with actionability. Among them, *BRCA1* and *BRCA2* germ-line mutations are the most commonly associated to familial susceptibility to breast-ovarian cancer, and genes involved in mismatch repair (MMR) *MLH1*, *MSH2*, *MSH6* and *PMS2* germ-line mutations associated to Hereditary non-polyposis colorectal cancer (HNPCC), a multisystem tumoral predisposition syndrome. Here we report the variant analyses of these 6 genes from a cohort of elderly (60 and older) from São Paulo, Brazil (n=1324) submitted to WGS. Among the over 12 thousand variants within these genes, 577 have coding or splicing consequences of which 35 potentially result in loss of function. Filtering for pathogenic assertions in ClinVar yielded 9 variants (7 curated) in 10 individuals, with frequencies of less than 0.02% (gnomAD). Among those, six subjects are asymptomatic, including a 93 year old woman with *BRCA1* p.Q563X stopgain carrier without family history of cancer. One subject carrying a *BRCA2* frameshift deletion reported lung cancer and another who carries a *MSH6* frameshift insertion reported intestinal cancer. Two subjects have not provided cancer history. These ten individuals are admixed, 7 are mostly non-European (<70% European ancestry). Our preliminary results suggest that genome and clinical data of elderly individuals are of utmost importance to improve variants pathogenicity interpretation and risks estimates in genetic counseling. Moreover, data on admixed individuals may provide insights to explain reduced penetrance and may lead to the identification of population specific genetic modifiers to known monogenic genes.

2806T

Novel *GREM1* enhancer region duplications and associated phenotypes. R. O'Connor, J. Van den Akker, S. Kim, G. Mishne, A. Zimmer, A. Leon, J. Ji. Color Genomics, Burlingame, CA.

Hereditary mixed polyposis syndrome (HMPS) is an autosomal dominant inherited cancer syndrome characterized by colorectal polyps of multiple or overlapping histological types. The syndrome is caused by duplications of the upstream regulatory region of the gremlin1 (*GREM1*) gene that include the 3' region of the adjacent *SCG5* gene. A duplication of 40kb was originally described in HMPS families of Ashkenazi Jewish descent (Jaeger 2012). Subsequently, duplications of 16kb (Rohlin 2016), 57kb (Venkatachalam 2011), and 24kb (McKenna 2018) have now also been described in patients with a personal or family history of colorectal cancer. These additional reported cases of *GREM1* duplications have indicated that testing of the 5' regulatory region of *GREM1* beyond the 40kb Ashkenazi founder is necessary to get a full picture of *GREM1*-associated HMPS. Here we report ten novel duplications or triplications of the *GREM1* upstream regulatory region identified in twelve individuals who were tested using a 30-gene panel for hereditary cancer risk at Color Genomics. Details of the duplications and the patient clinical histories will be presented. In summary, (1) duplications that include the 5' regulatory region of *GREM1*, and have both breakpoints located between the *SCG5* start codon and *GREM1* start codon, which have previously been demonstrated to increase the expression of the *GREM1* gene (Jaeger 2012, Rohlin 2016), correlate with strong clinical histories of colorectal cancer; (2) Larger duplications including the 5' regulatory region of *GREM1* with the entire *SCG5* gene and additional genes upstream correlate with weaker clinical histories and/or later onset of cancer; (3) duplications that include the 5' regulatory region and entire coding sequence of *GREM1* display the weakest clinical association. Each of these duplications was identified by NGS read depth analysis. NGS split read analysis revealed the breakpoints of 5 unique tandem duplications, which were confirmed by dedicated PCR and Sanger sequencing across the breakpoints. The remaining duplications were all confirmed by MLPA. This study extends the knowledge of the types of duplications occurring near *GREM1* in individuals of diverse ethnicities and reveals phenotypic variability associated with *GREM1* duplications.

2807F

Mutation spectrum identified by germline testing of hereditary cancers from 500 healthy Chinese individuals using an accessible 30-gene panel following ACMG guidelines. *D. Siu¹, L. Servais², Y. Okazaki³, S. Lam⁴, K. Law¹, J. Wu⁵, A. Gardner², A. Zhou².* 1) Rainbow Genomics, Hong Kong; 2) Color Genomics, Burlingame, CA; 3) Juntendo University Graduate School of Medicine, Tokyo, Japan; 4) Hong Kong Sanatorium and Hospitals, Hong Kong; 5) HKDNA laboratory, Hong Kong.

Driven by rapid reduction of DNA sequencing cost and increased access to testing, significant progress has been made in the characterization of hereditary cancer variant carriers in the U.S. and European populations. In contrast, relatively few studies have been conducted in the Chinese population, especially regarding asymptomatic carriers of germline pathogenic variants. With physician consultation and genetic counseling support, Hong Kong-based Rainbow Genomics has been providing hereditary cancer testing to both patients with cancer symptoms and healthy individuals, and the testing and reporting process follows current American College of Medical Genetics and Genomics (ACMG) guidelines. Here, we present the results of testing over 500 healthy Chinese individuals by next-generation sequencing using a 30-gene panel, in collaboration with Color Genomics (Color). The 30-gene panel detects pathogenic variants associated with increased risk for hereditary breast, ovarian, uterine/endometrial, colorectal, melanoma, pancreatic, prostate, and stomach cancer (*APC*, *ATM*, *BAP1*, *BARD1*, *BMPR1A*, *BRCA1*, *BRCA2*, *BRIP1*, *CDH1*, *CDK4*, *CDKN2A* (*p14ARF* and *p16INK4a*), *CHEK2*, *EPCAM*, *GREM1*, *MITF*, *MLH1*, *MSH2*, *MSH6*, *MUTYH*, *NBN*, *PALB2*, *PMS2*, *POLD1*, *POLE*, *PTEN*, *RAD51C*, *RAD51D*, *SMAD4*, *STK11*, and *TP53*). The testing was performed at Color's U.S. laboratory under CLIA and CAP compliance. English and Chinese clinical reports, localized pre-test consultation, and post-test genetic counseling in Mandarin or Cantonese were provided. Initial results indicate pathogenic or likely pathogenic variants were detected in over 8% of these individuals, higher than the 1% - 3% previously reported in the literature for healthy ethnic Chinese. The positive findings were mostly associated with colon and breast cancer variants. A significantly higher number of individuals also received reports with one or more variants of uncertain clinical significance (VUS). In this presentation, we will report the mutation spectrum of pathogenic and likely pathogenic variants, as well as the VUS detected. This Chinese variant distribution will be compared to those from other ethnic groups using data generated from Color's worldwide testing results.

2808W

Technical validation of EFIRM, a sensitive, plate-based, rapid, inexpensive, liquid biopsy platform for common *EGFR* variants in non small cell lung cancer. *C.M. Strom¹, F. Wei¹, J. Cheng¹, D. Chia¹, W. Liao², M. Tu², D. Wong¹.* 1) UCLA Head and Neck Cancer Center, Dental School, Los Angeles, CA., USA; 2) EZLife Bio, Los Angeles, CA.

Liquid Biopsy (LB) is the analysis of circulating tumor DNA (ctDNA) in patients with solid tumors. LB may be particularly useful in non small cell lung cancer (NSCLC) because of the frequent presence of activating mutations that convey sensitivity to particular chemotherapy drugs and the limited access to tumor tissue. The p.L858R and Exon 19 del variants in *EGFR* confer sensitivity to Tyrosine Kinase Inhibitors (TKI) and are found in up to 50% of NSCLC tumors in China and 10% of NSCLC tumors in the USA. Electric Field Induced Release and Measurement (EFIRM) is a novel plate based signal amplification platform that is capable of detecting ctDNA in both direct saliva and unprocessed plasma samples. Only 60 µl of direct plasma or saliva are needed for this assay. Two previous publications demonstrated near perfect correlation between EFIRM LB genotyping results and the tissue samples in patients with advanced NSCLC (stage 3 and 4) for the 2 most common *EGFR* somatic activating mutations, p.L858R and Exon19 del. Preliminary data has expanded this observation to early stage NSCLC patients. The current study presents the data from the technical validation of a qualitative EFIRM assay for 3 variants in *EGFR*, the common activating mutations, p.L858R and Exon19 del and a third variant, p.T790M that is seen when tumors develop resistance to tyrosine kinase inhibitors. We established reference ranges and determined limit of detection, intra and inter assay variability (0%) for each mutation. EFIRM has improved performance when benchmarked against published data from an available IVD for *EGFR* genotyping. The EFIRM assay is capable of detecting as few as 5 copies of mutant DNA and variants with a minor allele frequency of between 0.1% and 1%. Unlike other available LB modalities, EFIRM is capable of detecting fragments of single stranded DNA as small as 30 - 100 base pairs. Subcellular fractionation reveals that EFIRM is preferentially detecting variants in the exosomal fraction. In addition to the lack of sample manipulation, these observations may explain the ability of EFIRM to detect ctDNA using a small volume of plasma or saliva. Potential uses are early detection diagnosis screening, treatment selection, treatment monitoring, and stratification of individuals with indeterminate pulmonary nodules discovered on screening spiral CT scans. Work in underway to validate a quantitative *EGFR* assay for these 3 mutations. .

2809T

Association of lower expression of beta-catenin with pancreatic carcinogenesis and poorer prognosis. L. Suo¹, L. Green^{1,3}, J. Jiao², S. Lai^{1,3}. 1) Dept of Pathology, Baylor College of Medicine, Houston, TX; 2) Dept of Molecular & Cellular Oncology, MD Anderson Cancer Center, Houston, TX; 3) Dept of Pathology, Michael E. DeBakey VA Medical Center, Houston, TX.

Background: Pancreatic ductal adenocarcinoma (PDA) is the most common malignancy of the pancreas and the third most common cause of cancer related mortality with a 5-year survival of <5%. Promising cancer treatments depend on identification of biomarkers and targeted therapy. Beta-catenin is a dual function protein with an important role in cell-cell adhesion and gene transcription. It has been reported to contribute to tumor progression. However, its role in pancreatic cancer is yet uncharacterized. The aim of this study is to determine the expression and clinical significance of beta-catenin in relation to tumor progression and survival. **Design:** Tissue microarray samples were retrieved from archives from 2007-2015, including pancreatic tissue: 29 PDA, 16 neuroendocrine tumors (NET), 1 sarcomatoid carcinoma, 8 benign neoplasms (BEN), 24 normal parenchyma (N), and 13 nonpancreatic carcinomas (NPC). Immunohistochemistry was performed using anti-beta-catenin antibody and each section was scored (ranged from 0-300) by the sum of multiplying both intensity (0, 1+, 2+ and 3+) and extension values (0-100%). Related pathologic data and patients' demographic information were also obtained. **Results:** Differential membranous expression of beta-catenin was observed. Beta-catenin scores were similar in normal pancreatic acini and ducts. Beta-catenin scores were significantly lower in cancer groups (PDA, NET and NPC) compared with BEN and N groups. Decreasing trend of beta-catenin score was found in the more clinically aggressive pancreatic adenocarcinomas: N (227.5±19.5), BEN (200.5±24.8) and PDA (114.5±18.4). Importantly, lower beta-catenin expression was correlated with poorer survival in PDA group (p=0.02). **Conclusions:** Our findings support the theories that the low beta-catenin membranous expression is one of the possible mechanisms of pancreatic carcinogenesis and may be a prognostic factor for poor outcome in pancreatic ductal adenocarcinoma.

2810F

Development of SureSelect sequencing panels and algorithms to simultaneously detect mutations, copy number variations and DNA rearrangements in FFPE specimens. H. Tao, A. Ashutosh, A. Khare, G. Amparo, C. Lecocq, A. Vadapalli, L. Forsmark, L. Niu, J. Jin, J. Ghosh, M. Ruvalo, D. Roberts. Agilent Technologies, Santa Clara, CA., USA.

Introduction: We have developed custom SureSelect target enrichment sequencing panels for comprehensive profiling of single nucleotide variations (SNP/INDEL), gene level copy number variations (CNV), and DNA translocations (TL) in a single assay. SureSelect target enrichment baits were designed to cover exons of target genes to detect SNP/INDEL, spanning putative translocation regions to detect TL, or periodically spaced across genes being assessed for CNV. We leverage high sensitivity SureSelect (XTHS) reagents employing molecular barcodes to aid in detection of low frequency mutations. To evaluate detection limit of SNP, CNV and TL, a custom SureSelect cancer panel was utilized in SureSelect target captures on gDNA from cell lines and FFPE samples harboring CNVs (ERBB2 or MET) or fusion genes (ALK or ROS1). Detection limits were assessed by identifying the lowest detectable allele frequency in serial dilutions of 2 genomic DNA samples titrated against one another or tumor DNA titrated with its matched normal DNA. To further assess the sensitivity and accuracy of CNV and TL detection, multiple breast FFPE tumor samples previously determined to be Her2+ by IHC, and multiple lung FFPE tumor samples (ALK positive by FISH), were analyzed with the above SureSelect custom panel. Fusions and CNVs detected by NGS were compared with results from Fluorescence in situ Hybridization (SureFISH). **Results:** For SNP/INDEL, SureSelect panels coupled with our analysis algorithm can detect SNVs down to 1% allele frequency. For CNV, DNA containing 3 copies of ERBB2 or MET genes can be detected at 10% tumor fraction (overall 2.1 copies) in both cell line samples and FFPE samples. For TL, ALK/EML4 or SLC34A2/ROS1 fusion genes can be detected at ≤3% tumor fraction. In comparison with FISH, the positive percent agreement (PPA) for ERBB2 CNV detected by the custom cancer panel is 97.5% (40/41) and negative percent agreement (NPA) is 97.5% (40/41). For ALK TL detection in 52 lung FFPE samples, the custom cancer panel successfully identified EML4/ALK or KIF5B/ALK gene fusion in 37 (86%) out of 43 ALK FISH positive samples. **Conclusions:** We demonstrate a customizable SureSelect sequencing platform to detect all types of DNA mutations in FFPE samples with high sensitivity and accuracy. SureSelect target enrichment assays have great potential to be used as alternatives to or pre-screens for traditional FISH assays.

2811W

Molecular cytogenetic characterization of a case of a mixed myelodysplastic/myeloproliferative neoplasm: Chronic myelomonocytic leukemia-1 (CMML-1) with an abnormal karyotype with an apparent monosomy 7 resulting in rearrangements involving chromosomes 7 and 21. C.A.

Tirado^{1,2,3,4}, D. Chung^{2,5}, A. Reyes^{2,5}, A. Nguyen^{2,5}, K. Cunnien¹, L. Batholomaeus¹, D. Serk¹, J. Ahmad¹, K. Stieglbauer^{1,3}. 1) Cytogenetics, Allina Health Laboratory, Minneapolis, MN; 2) The International Circle of Genetic Studies, Los Angeles, CA; 3) HPA, Minneapolis, MN; 4) The University of Minnesota, Department of Laboratory Medicine and Pathology, Minneapolis, MN; 5) The University of California at Los Angeles, Los Angeles, CA.

We report here a 69 y.o. male with persistent anemia suspicious cells on blood smear review (by clinical history). Renal dysfunction raised the issue of a plasma cell neoplasm. However, on 08/02/2017 the LD, total was 529.00. No recent iron panel data or ferritin determination was seen in the Allina Laboratory database. This patient was then diagnosed with mixed myelodysplastic/myeloproliferative neoplasm: chronic myelomonocytic leukemia-1 (CMML-1). Chromosome analysis revealed an abnormal karyotype with an apparent monosomy 7 and the presence of two marker chromosomes in 19 of the 20 metaphase cells examined. FISH analysis of metaphases from distained G-banded slides revealed translocation of D7S486 to a derivative chromosome 21, and 2 copies of *RUNX1* located on an isoderivative chromosome 7. This karyotype was then reinterpreted as an abnormal male karyotype with rearrangements of chromosomes 7 and 21 [ider(7)(q10)t(7;21)(q11.2;q11.2), der(21)t(7;21)], resulting in loss of 7p and gain of 21q. And it was described as: 46,XY,ider(7)(q10)t(7;21)(q11.2;q11.2),der(21)t(7;21)[19]/46,XY[1]. ish ider(7)(D7Z1+,*RUNX1*++),der(21)(D7S486+).nuc ish(D7Z1,D7S486) x2[200].(*RUNX1*T1x2,*RUNX1*x3)[333/500] Extra copies of *RUNX1* and abnormalities of chromosome 7 are seen in myeloid disorders including MDS/MPN. Complex rearrangements as the ones present in this study suggest genomic instability which is usually associated with a poor prognosis. Clinicopathologic correlation of these results is recommended. This case pinpoints the importance of characterization of complex rearrangements using molecular cytogenetics techniques on metaphases previously G-banded.

2812T

Reference standards for the validation of myeloid sequencing assays. M. Wang, H. Child, J. Wickenden. Horizon Discovery, Cambridge, UK, Cambridgeshire, United Kingdom.

Introduction: Myeloid neoplasms are diseases of the hematopoietic stem cells. They can be categorised into five primary types based on WHO classification, with Acute Myeloid Leukemia (AML) occurring at the highest incidence. With the advent of more robust sequencing capabilities, pathogenic mutations have now been identified in genes from 5 main groups: signalling pathway proteins (e.g. *CBL*), transcription factors (e.g. *RUNX1*), epigenetic regulators (e.g. *ASXL1*), tumor suppressors (e.g. *TP53*) and components of the spliceosome (e.g. *SF3B1*). A recent influx of new myeloid gene panels for sequencing of these genes requires rigorous validation to ensure sufficient clinical accuracy, as errors can often occur in library prep, sequencing, analysis and interpretation of results. In order to support this effort, Horizon has developed a novel cell-line derived reference standard containing 22 variants across 19 genes involved in all of the main pathway groups, including *CBL*, *RUNX1*, *ASXL1*, *TP53*, *SF3B1*, *JAK2*, *FLT3*, *ABL*, *KRAS* and *DNMT3A*. **Methods:** Engineered cancer lines were single-cell diluted and investigated for their genomic status by Sanger sequencing at key loci identified as relevant to myeloid cancer in the literature. Target variants were then further characterised by ddPCR, to enable the variant allele frequency to be reliably determined. The material was sent to test labs for assessment across a range of mainstream myeloid NGS gene panels and compared to the gold standard ddPCR data. **Results:** Sequencing by both Sanger and ddPCR confirmed the presence of 22 variants across 19 genes relevant to myeloid cancer, at a blended ratio of 5-70% variant allele frequency. Multiple reproducible batches of DNA were tested producing consistent sequencing results. The reference material was found to have good coverage across some of the most commonly used clinical myeloid NGS gene panels, and performance-testing yielded accurate variant calling of the confirmed mutations at the expected allele frequencies as originally verified by ddPCR. **Conclusion:** This is the first fully cell-line derived Myeloid Reference Standard containing 22 relevant mutations across 19 genes implicated in the diagnosis, management and treatment of myeloid cancers. Results show it to be a consistent, commutable and high quality DNA reference standard that closely mimics DNA from real clinical samples to support the genetic testing of myeloid cancer patients.

2813F

Improving patient outcomes: Demonstrating competence in *BRCA1/BRCA2* variant classification for PARPi treatment stratification. *N. Wolstenholme¹, F. Khawaja², F. Moon², S. Patton¹, Z.C. Deans².* 1) EMQN, Manchester, United Kingdom; 2) UK NEQAS/GenQA, Edinburgh, United Kingdom.

Introduction The availability of PARP-inhibitor therapy for ovarian cancer patients with *BRCA1/BRCA2* mutations has introduced the requirement for testing laboratories to accurately classify the pathogenicity of variants detected. There is a worldwide approach to the implementation of standardised classifications many using ACMG/AMP guidelines and ENIGMA guidance. However, the success of application is variable and in order to drive standardisation and educate those performing variant classifications a web-based external quality assessment has been provided. **Methods** Using the NEQAS/GenQA online genomics training assessment competency tool (G-TACT) a module was developed to enable individuals to assess 15 *BRCA1/BRCA2* variants (SNVs, exonic deletions/duplications and splicing defects) and submit a classification of pathogenicity for each along with or without recommendation for PARP-inhibitor treatment. The variants ranged from class 1, benign to class 5, pathogenic. The submissions were scored, reported to each participant and a summary report was issued. This detailed the expected variant classifications and provided an overview of all submitted results. Participants were invited to join WebExs which summarised the evidence assessed, the expected variant classifications and addressed participant queries. **Results** Submissions were received from 271 individuals from 59 countries. The majority of classifications were correct with a third +/-1 class. As most participants with discordant classifications did not provide their evidence it is not possible to attribute reasons to these conflicting classifications. However five variants gave variable classifications and these will be discussed. In countries where mutation testing is required for PARP inhibitor treatment, this is recommended for likely pathogenic and pathogenic variants. Overall where participants had correctly classified the variant, the recommendations for PARP inhibition therapy were correct. **Conclusion** Overall there was consistency of variant classifications with most variability between benign or likely benign, and pathogenic or likely pathogenic. However where a variant of uncertain significance (VUS) was reported instead of a likely pathogenic/pathogenic classification this would have an impact on patient management and therefore the need for standardisation has been identified and through education, training and competency assessment this can be achieved.

2814W

Mutation analyses of four Japanese pedigrees with Hereditary Pheochromocytoma/Paraganglioma Syndrome (HPPS): Identification of a novel *SDHB* IVS2-2A>C mutation in two unrelated pedigrees. *M. Yamanaka^{1,3}, K. Shiga⁴, S. Fujiwara¹, Y. Mizuguchi¹, S. Yasuda^{1,3}, K. Ishizawa¹, Y. Saiki¹, K. Higashi^{1,2}, T. Ogawa^{1,2}, N. Kimura⁵, A. Horii¹.* 1) Department of Molecular Pathology, Tohoku University School of Medicine, Sendai, Miyagi, Japan; 2) Department of Otolaryngology-Head and Neck Surgery, Tohoku University School of Medicine, Sendai, Miyagi, Japan; 3) Exploring-Germination-and-Growth Program for Young Scientists, Tohoku University, Sendai, Miyagi, Japan; 4) Department of Otolaryngology and Head and Neck Surgery, Iwate Medical University School of Medicine, Morioka, Iwate, Japan; 5) Department of Clinical Research, Pathology Division, National Hospital Organization Hakodate National Hospital, Hakodate, Hokkaido, Japan.

Pheochromocytomas and paragangliomas are neuroendocrine tumors which arise from sympathetic and parasympathetic nerve. Hereditary cases afflicted by both or either pheochromocytomas and paragangliomas have been reported: these are called hereditary pheochromocytoma/paraganglioma syndromes (HPPS). Many cases of HPPS are caused by mutations of one of the *SDHB*, *SDHC*, or *SDHD* genes that encode subunits for the mitochondrial respiratory chain complex II. In this study, we investigated mutations of these genes in six HPPS patients from four Japanese pedigrees using peripheral blood lymphocytes and tumor tissues. Results showed that all of these cases harbored germline mutations in *SDH* genes. In two pedigrees, a novel splice-site mutation in *SDHB* was found, and the tumor RNA of the patient clearly showed frameshift caused by exon skipping. The remaining two pedigrees harbored reported mutations in *SDHB* and *SDHD*, respectively. All these mutations were heterozygous in constitutional DNAs, and two-hit mutations were evident in tumor DNAs. The novel mutation revealed by our study may contribute to improvement of clinical management for patients with HPPS.

2815T

Genetic features in children with acute lymphoblastic leukemia. *L. T. An, D. Q. Vu, X. H. Nguyen, D. N. Ngo, N. L. Bui, T. H. Le.* Human Genetic Department, National children hospital, Hanoi, Hanoi, Viet Nam.

Background: To study chromosome abnormalities and fusion oncogenes caused by chromosomal abnormalities play an important role in acute lymphoblastic leukemia (ALL) especially in children. Detecting the most common fusion oncogenes such as BCR-ABL, MLL-AF4, ETV6-RUNX1 and TCF3-PBX1, will play an important role the assessment of prognosis and indication of treatment regimen. **Methods:** 111 patients were collected in the descriptive study from Feb, 2017 to March, 2018 in the National Children's Hospital, who were recently diagnosed with ALL. Karyotype analysis from bone marrow was performed on all patients. 44/111 patients were done Fluorescence In Situ Hybridization (FISH) to detect the common fusion oncogenes. **Results:** 32/111 (28.83%) patients had been failed karyotype. The frequency of cytogenetic abnormalities, including numerical and structural arrangements was 39.24% (31/79). 16/79 (20.25%) patients had structural chromosomal abnormalities, the most common was t(1;19) in 8.86% (7/79) of the patients. 15/79 (16.46%) patients had hyperdiploid (47-68 chromosomes), five patients complex karyotypes with structural abnormalities. 2/79 (2.53%) patients had hypodiploid, one patient had t(1;19). Complex karyotypes (a finding of three and more chromosomal aberrations in a karyotype) were identified in 5 (6.33%) out of 79 patients. 3 patients had variable chromosome included inv9(p12q13), 9qh+ and 15ps+. Fusion oncogenes were detected in 5 (11.36%) out of 44 patients by FISH, included: 3 patients with ETV6/RUNX1 and 2 patients with MLL/AF4. 17/44 (38.64%) had extra or missed signals (1 to 3 signals) of the oncogenes. **Conclusion:** Applying FISH technique to detect the common fusion oncogenes and bone marrow karyotype analysis at diagnosis should be used to guide treatment decision and prognosis in children with acute lymphoblastic leukemia.

2816F

Recurrent independent cytogenetic characteristics of myelodysplastic syndromes (MDS) and associated IPSS-R risk. *BaniB. Ganguly.* MGM New Bombay Hospital, Navi Mumbai, Maharashtra, India.

MDS represents heterogeneous clonal pre-malignant condition of hematopoietic stem cells that are typically associated with peripheral blood cytopenia, ineffective hematopoiesis, hypercellular marrow with morphologically defined dysplasia and risk of progression to AML. Conventional cytogenetic anomalies of bone-marrow have been reported in 30% *de novo* MDS and 50% of therapy-related. Del(3q/5q/7q/11q/12p/20q), monosomy 5/7, trisomy 8/19, i(17q), -Y, are frequently detected. A number of anomalies indicate presumptive diagnosis in absence of morphological dysplasia, including del(5q)/-7/del(7q)/del(9q)/-13/del(13q)/del(11q)/t(11q)/del(12p)/t(12p)/-17/del(17p)/i(17q)/+19/t(19)/idic(Xq13); whereas, +8/del(20q)/-Y are not considered as presumptive evidence of MDS because of their global presence in other myeloid neoplasms. Therapeutic response and clinical evolution of chromosomal rearrangements are taken into account for classification and risk-stratification in International Prognostic Scoring System-Revised (IPSS-R). We present cytogenetic data on 314 MDS collected during 2013-2014 of different age groups with prevalence at older group (60-80 years). Conventional cytogenetic investigation was carried out following unstimulated bone-marrow culture. The major observation includes culture failure in 5.4%, and normal and abnormal karyotypes in 48.7% and 45.9% respectively, which is concordant with Schanz et al, 2012 except del(5q), which was detected in all complex aberrations in our study (Table). It has been observed that many of the abnormalities are not considered in IPSS-R classification. Therefore, compilation of cytogenetic data will throw some input in understanding the risk and therapeutic response, and also may propose new cytogenetic risk grouping.

2817W

Cytogenomic characterization of B-Acute Lymphoblastic Leukemia (ALL) with intrachromosomal Amplification of Chromosome 21 (iAMP21) reveals co-existing cytogenetic manifestations and outlines driver alterations that play a role in leukemogenesis. R. Garcia, K. Wilson, P. Koduru. UT Southwestern, Dallas, TX.

B-ALL with iAMP21 is a rare cytogenetic aberration accounting for 2% of cases. It is detected by FISH using the *RUNX1* probe. While amplified *RUNX1* is not considered pathogenic, it is consistently part of the critical region amplified on chromosome 21. In this study, we explored the relationship of other cytogenetic abnormalities underlying evolutionary pathways and identified driver alterations leading to disease progression in this subset of B-ALL. Cytogenetic data from B-ALL patients with iAMP21 from the literature was evaluated for recurring chromosome abnormalities (RCAs), constructed oncogenetic tree to identify evolutionary pathways, identified genes involved in these pathways and in disease progression. Finally oncogenetic signatures and signaling pathways associated with the driver genes were identified. Programs used for the above analyses were tm text analysis in the R environment, TRNCO, PCA, Driver Net Algorithm, and Enrich R. Results from this analysis were applied to classify four institutional cases (3B-ALL one AML) with iAMP21. A total of 412 B-ALLs with iAMP21 were retrieved from the literature; in these karyotypes 31 RCAs were identified. The timed oncogenetic tree analysis of these RCAs revealed seven major and two minor clusters with RCAs 21q-, +X, 20q-, 12p-, 7q-, 11q-, 9p-, +9 and 15q- as entry points, and these were driver alterations. Three other aberrations, 6q-, 13q- and 16q- were determined to be secondary and probably involved in clonal evolution. These driver alterations deregulated DGF-ERK, JAK2, PTEN, HOXA9, KRAS, RAF, MTOR, EGFR, KRAS and E2F1 pathways, and abnormal signaling in JAK-STAT, Toll like receptor and VEGF pathways. CMA data from 4 institutional cases was compared with results from the above analyses. Two cases were assigned to the 21q- pathway. This subset consistently deleted *PCNT* and *DIP2A* and correlated with PDGF-ERK and JAK2 downregulation ($p = .01$). A third case was assigned to the +X category containing *PDK3*, *KDM6A*, *CSF2RA*, *ARHGAP6* and *IL1RAPL1*, all associated in a number of oncogenic pathways including TP53, ATM, NOTCH, KRAS and RAF ($p > .05$). The AML case with iAMP21 was unclassifiable into any of the B-ALL clusters by our data. This analysis suggests B-ALL with iAMP21 is comprised of closely related subset of neoplasms that are characterized by co-existing cytogenetic abnormalities and by functional driver alterations that play a role in tumor progression. Key words: iAMP21, Driver genomic alterations, array.

2818T

Genetic heterogeneity in pediatric alveolar rhabdomyosarcoma tumors.

K. Gleditsch^{1,2}, J. Penas², D. Mercer², A. Umrigar², M. Stark², F. Tsiens², A. Hollenbach². 1) Division of Hematology/Oncology, Department of Pediatrics, Louisiana State University Health Sciences Center, New Orleans, LA; 2) Department of Genetics, Louisiana State University Health Sciences Center, New Orleans, LA; 3) Department of Pathology, Children's Hospital New Orleans, New Orleans, LA.

In the United States, more than 13,000 cases of soft tissue sarcomas are diagnosed per year, of which over 5,000 are fatal. Rhabdomyosarcoma is the most common pediatric type of soft tissue sarcoma affecting skeletal muscle. The alveolar subtype (ARMS) typically has a more aggressive clinical course and portends a poorer clinical prognosis than the embryonal subtype (ERMS). Therefore distinguishing between the two subtypes is necessary for appropriate treatment. ARMS is genetically characterized by one of three translocation states: the presence of a translocation between chromosomes 2 and 13 [$t(2;13)(p35;p14)$] leading to the production of the PAX3-FOXO1 fusion protein, the presence of a translocation between chromosomes 1 and 13 [$t(1;13)(p35;q14)$] leading to the production of the PAX7-FOXO1 fusion protein or a lack of translocation leading to fusion negative status. These oncogenic proteins have the potential to enhance the local invasive capacity of the tumor cells and increase metastatic potential, PAX3-FOXO1 more so than PAX7-FOXO1. Multimodality analyses, such as SNP array and RT-PCR, have previously shown that significant heterogeneity regarding the number of translocations exists within tumor tissue samples from rhabdomyosarcoma patients, leading to a variation in clinical outcome. The aim of this study is to investigate the heterogeneity of pediatric ARMS tumor specimens obtained from patients of the Children's Hospital of New Orleans Hematology/Oncology clinic. We have designed and validated a fluorescence in situ hybridization (FISH) probe in our laboratory that can simultaneously and distinctively identify $t(2;13)$ and $t(1;13)$ translocations. Thirty-six tumor samples from rhabdomyosarcoma patients ranging in age from 6 months to 18 years were analyzed to determine the presence of these clinically significant translocations and the extent of heterogeneity within each sample. In addition to identifying varying numbers of translocation events within nuclei, we also observed a large amount of heterogeneity with respect to the percentage (7-96%) of translocation positive nuclei within tumor samples, along with differences in the extent of aneuploidy within individual cells. Our analysis of the cytogenetic markers for the $t(2;13)$ and $t(1;13)$ demonstrates that these translocations exhibit significant variation within these rhabdomyosarcoma samples providing a potential method to improve diagnostic and prognostic indicators for these tumors.

2819F

Translocation(3;8)(q26.2;q24) in myeloid neoplasm: Report of two new cases and review of the literature. J. Liu¹, N. Benvenuto², G. Uppa³, S. Nedumakel¹, D. Foster¹, E. Johnson⁴, J. Gong⁵, Z. Wang⁴, S. Peiper⁵. 1) Clinical Cytogenomics Laboratory, Department of Pathology, Anatomy, and Cell Biology Thomas Jefferson University, 1025 Walnut Street, Philadelphia, PA 19107; 2) Sidney Kimmel Medical College, Thomas Jefferson University 1025 Walnut Street, Philadelphia, PA 19107; 3) Section of Hematopathology, Hematology and Flow Cytometry, Department of Pathology, Anatomy, and Cell Biology, Sidney Kimmel Medical College, Thomas Jefferson University, 125 South 11th Street, Philadelphia, PA 19107; 4) Molecular & Genomic Pathology Laboratory, Department of Surgery, Pathology, Anatomy, and Cell Biology, Thomas Jefferson University, 1025 Walnut Street, Philadelphia, PA 19107; 5) Department of Pathology, Anatomy, and Cell Biology, Thomas Jefferson University, 1020 Locust Street, Philadelphia, PA 19107.

Translocation (3;8)(q26.2;q24), involving the *MECOM* gene complex (*MDS1* and *EV11* Complex Locus), is an extremely rare chromosomal abnormality in myeloid neoplasms. At present, there are only 23 cases reported in literature, mainly found in AML, MDS, CMML and blastic phase of CML. Patients with t(3;8)(q26.2;q24) demonstrate overlapping features to those with other types of 3q26 abnormalities, such as elevated platelet counts, multilineage dysplasia, monosomy 7 and poor clinical outcomes. The underlying molecular pathological mechanism is not well understood mainly due to the rarity of this entity. Here we report the clinical, morphological, cytogenetics, molecular genetics and prognostic features in two AML cases with t(3;8). Available clinical data was reviewed, including age, gender, presentation, treatment, response and outcomes. Peripheral blood smears, bone marrow aspirate and core biopsy results were assessed. Bone marrow cellularity, blast morphology and percentage, background dysplasia were particularly studied. Immunophenotype by flow cytometry and/or immunohistochemistry and laboratory data were also collected. Patient 1 was a 74-year-old male with a history of CMML. The bone marrow biopsy showed significantly increased blasts (64%) and markedly hyperplastic atypical megakaryocytes. Cytogenetics and FISH studies demonstrated 46,XY,t(3;8)(q26;q24), -7 and three copies of *MECOM* consistent with a rearrangement. Next generation sequencing (NGS) technology based hematological malignancy gene panel test revealed pathogenic mutations in genes *ASXL1*, *FLT3*, *KRAS*, *PTPN11*, *SRSF2* and *TET2*. The patient had persistent poor response to chemotherapy and died four year later after the initial diagnosis. Patient 2 was a 54-year-old male with a history of MDS-RAB. The bone marrow biopsy showed dramatically increased blasts (~90%) and significant megakaryocytic dysplasia. Cytogenetics and FISH studies demonstrated 46,XY,t(3;8)(q26;q24), -7 and a c-MYC gene rearrangement. NGS based molecular test revealed pathogenic mutations in genes *IDH2* and *NRAS*. The patient was lost to follow up later on. In summary, two new cases with t(3;8) were reported, a thorough review of currently published clinical cases were also performed, and possible mechanisms in addition to overexpression of the *MECOM* gene are discussed. This is also believed to be the first time that NGS panel test was included to describe this unique entity.

2820W

Molecular cytogenetic characterization of Rh30 and Rh4 alveolar rhabdomyosarcoma (ARMS) cell lines. J.A. Peñas¹, K. Gleditsch^{1,2}, D. Mercer¹, A. Umrigar¹, Y. Li³, T.J. Chen³, A. Hollenbach¹, F. Tsien¹. 1) Department of Genetics, Louisiana State University Health Sciences Center, New Orleans, LA; 2) Children's Hospital of New Orleans, New Orleans, LA; 3) Hayward Genetics Center, Tulane School of Medicine, New Orleans, LA.

Rhabdomyosarcoma, the most common soft tissue sarcoma in children, is comprised of two subtypes: alveolar (ARMS) and embryonal (ERMS). ARMS makes up 30% of all rhabdomyosarcoma cases, is typically more aggressive, and has a poorer clinical prognosis. ARMS is characterized by the presence of the t(2;13)(q35;q14) or t(1;13)(p36;q14) translocations, leading to the creation of the PAX3-FOXO1 and PAX7-FOXO1 oncogenic fusion proteins, respectively, with the PAX3-FOXO1 variant being more aggressive. Up until now the only commercially available diagnostic fluorescence *in situ* hybridization (FISH) probe for detecting the presence of the translocation, a FOXO1 breakpoint probe, was not capable of differentiating between these events, making accurate diagnosis difficult. Therefore, we designed novel probes capable of simultaneously differentiating PAX3-FOXO1 and PAX7-FOXO1, allowing for a more accurate diagnosis. We used these probes to address the specific aims of this project, which is to determine the significant large-scale genetic similarities and differences between Rh4 and Rh30 cell lines, two of the most commonly used ARMS cell lines, which derive from pediatric patients clinically diagnosed with ARMS. We utilized G-banding, FISH, spectral karyotyping (SKY), and array comparative genomic hybridization (aCGH), followed by pathway analysis which utilizes bioinformatics to uncover affected cellular processes and networks. Initial FISH results demonstrate that Rh30 had a higher number of PAX3-FOXO1 fusions compared to Rh4 whereas Rh4 had significant amplification of *PAX3*, *PAX7*, and *FOXO1*. Consistent with our FISH results, G-banding demonstrated a significant difference in ploidy between Rh4 and Rh30. Initial SKY analysis of Rh4 shows large scale and highly complex chromosomal rearrangements and aneuploidy throughout the genome. Our aCGH data confirmed large-scale insertions and deletions in gene regions previously associated with ARMS. Bioinformatics analysis of genes present in these regions of insertion and deletion identified genes previously associated with ARMS while also identifying novel DNA markers that could potentially be used for clinical diagnosis and the development of targeted therapies. Therefore, the resulting cytogenetic profile and bioinformatics analysis of these cell lines may potentially be used as a reference for ARMS researchers with our novel probes providing greater specificity for more accurate diagnosis for ARMS patients.

2821T

T(3;8)(q26.2;q24) is commonly associated with therapy-related myeloid neoplasms and involves MECOM/MYC rearrangement. G. Tang, S. Hu, Z. Tang, G. Toruner, J. Medeiros, W. Xie. Hematopathology, University of Texas MD Anderson Cancer Center, Houston, TX.

BACKGROUND T(3;8)(q26.2;q24) is a rare recurrent cytogenetic abnormality and ~20 cases have been reported. Most cases reported were myeloid neoplasms and showed *MECOM* rearrangement. The clinicopathologic, cytogenetic and molecular features of myeloid neoplasm with t(3;8)(q26.2;q24) have not been well characterized. **METHODS** We retrospectively reviewed the cytogenetic archives in our hospital over the past 20 years (1998-2018) and identified 20 patients who had myeloid neoplasm with t(3;8)(q26.2;q24). A detailed chart review was conducted; FISH analysis for *MECOM* and *MYC* rearrangements using break-apart probes was performed in cases with BM material available. **RESULTS** The study group included 12 men and 8 women, with a median age of 61 years (range, 21-79). Eleven patients had therapy-related myelodysplastic syndromes (MDS, n=3) or acute myeloid leukemia (AML, n=8); 7 had blast phase of chronic myeloid leukemia (CML, n=4) or chronic myelomonocytic leukemia (CMML, n=3); 1 had primary MDS and 1 had primary AML. Four patients had t(3;8)(q26.2;q24) as the sole abnormality, 7 had one additional abnormality and 9 had 2 or more additional chromosomal abnormalities; monosomy 7 was detected in 5 patients. *MECOM* rearrangement was confirmed in 13/14 patients and *MYC* rearrangement in 12/14 patients by FISH. *MYC* expression was assessed in 14 patients by immunohistochemistry, all 14 patients had *MYC* expression, ranging from 20% to 70%, and the level was largely correlated with the percentage of blasts in bone marrow. Fourteen patients underwent gene mutation testing, including 6 patients with targeted next generation sequencing, 7 of 14 patients showed at least one gene mutation, including *ASXL1* (2/6), *TET2* (2/6), *SRSF2* (2/6), *NRAS* (2/12), *ABL1* (1/6), *CBFB* (1/6), *BRINP3* (1/6), *GATA2* (1/6), and *RUNX1* (1/6). *FLT3* mutation was not detected in any of 14 patients tested. Fourteen patients were treated with chemotherapy or hypomethylating agents, with or without tyrosine kinase inhibitors. At the last follow-up, 18 patients died and 2 were alive with persistent disease, the median overall survival was 6 months. **CONCLUSION** T(3;8)(q26.2;q24) often leads to *MECOM/MYC* rearrangement and is highly associated with therapy-related myeloid neoplasms and advantage stage of disease. Patients with this chromosomal abnormality often have very poor outcomes.

2822F

ETV6 deletion as a potential biomarker for blastic plasmacytoid dendritic cell neoplasm. Z. Tang, W. Wang, G. Tang, N. Pemmaraju, S. Hu, C. Yin, M. Konopleva, G.A. Toruner, L.J. Medeiros, J.D. Khoury. University of Texas MD Anderson Cancer Center, Houston, TX.

Background: As a rare and aggressive hematologic malignancy with very poor prognosis, over 50% blastic plasmacytoid dendritic cell neoplasm (BPDCN) cases present chromosomal abnormalities detected in bone marrow specimen, approximately 40% of them with an involvement of 12p, mostly a deletion of 12p or other structural abnormalities involving rearrangement of *ETV6* gene located on 12p. In this study, we investigate the potential application of *ETV6* gene as a biomarker for this group of BPDCN cases. **Methods:** Patients with confirmed BPDCN treated at our institution from January 2014 to April 2017 were included in this study. Conventional cytogenetic analysis was routinely performed as a workup and staging of BM specimen in all cases. Fluorescence in situ hybridization (FISH) and array comparative genomic hybridization (CGH) to assess *ETV6* or the 12p locus were performed in selected cases according to the physician's orders. Additional *ETV6* FISH tests were performed on archived BM specimen of 11 cases and skin biopsy specimen of 6 cases respectively. Clinical and laboratory data were collected by electronic medical chart review. **Results:** Twenty-seven men and 5 women, with a median age of 64 years (range, 2 to 84 years) were included in this study. Six out of 12 cases with an abnormal karyotype exhibited an apparent 12p involvement by conventional cytogenetics in the BM specimen. Notably, chromosome 12p aberrations resulting in *ETV6* rearrangement were detected in 2 BM samples without morphologic or immunophenotypic involvement by BPDCN. We then assessed 29 BM samples collected at different time points from 11 BPDCN patients by FISH and identified 3 cases with *ETV6* deletion that were not detectable by karyotype analysis previously. Very interestingly, the detection of *ETV6* deletion was consistent with the clinical course of BPDCN diseases in these patients. Notably, *ETV6* deletion is detected in all 6 skin biopsy specimen involved by BPDCN. **Conclusions:** Chromosomal aberrations involving *ETV6* (including *ETV6* deletion and rearrangement) are highly prevalent in BPDCN. The rate of *ETV6* aberrations seems highest in disease lesions involving the skin, a feature of potential diagnostic utility at presentation. In some instances, aberrations involving the *ETV6* locus may be identified in the BM in the absence of detectable involvement by BPDCN. The *ETV6* can be a potential biomarker for BPDCN.

2823W

Duplication of the band q21q27 on the long arm of chromosome 3 : A rare cytogenetic event in B-Chronic Lymphocytic Leukemia (B-CLL). D. Zhao^{1,2}, K. Liu^{1,2}, A. Nguyen^{1,2}, D. Chung^{1,2}, J. Reinartz^{3,4}, D. Serk⁵, J. Ahmad⁵, L. Bartholomaeus⁵, K. Stieglbauer^{3,4}, A. Reyes^{1,2}, C.A. Tirado^{2,3,4,5}. 1) UCLA, Los Angeles, CA; 2) The International Circle of Genetic Studies, Los Angeles, CA; 3) Allina Health, Inc. Minneapolis, MN; 4) HPA, Minneapolis, MN; 5) The University of Minnesota, Department of Laboratory Medicine & Pathology, Minneapolis, MN.

We present the case of an 83-year-old female with a long history of B-CLL followed by observation only. Twelve years after her diagnosis of CLL, routine follow-up chromosome analysis of peripheral blood revealed an abnormal metaphase with a dup(3)(q21q27) in 18 of 20 metaphase cells. To further characterize the abnormal chromosome 3, fluorescence in situ hybridization (FISH) was performed using the Abbott BCL6 probe for 3q27. An additional BCL6 signal was observed in 303 of the 500 interphase nuclei examined. The ISCN was reported as 46,XX,dup(3)(q21q27)[1]/46,XX[2]. nuc ish(5'BCL6, 3'BCL6)x3(5'BCL6 con 3'BCL6x3)[303/500]. This abnormality was seen again in one of three available metaphases in a follow up peripheral blood study 5 years later, consistent with persistent disease. Molecular genetic analysis identified the presence of somatic hypermutation of the immunoglobulin heavy chain gene variable region (IGH-V), which is recognized as an independent favorable prognostic marker in CLL. FISH analysis was negative for loss of SEC63 (6q21), amplification of MYC (8q24), loss of ATM (11q22.3), trisomy 12, loss of D13S319 (13q14.3), loss of TP53 (17p13.1) and CCND1/MYEOV/IGH rearrangement [t(11;14) (q13;q32.30)]. Partial trisomy 3 is a relatively rare event seen in B-CLL, with commonly overrepresented segments including the q21-23 region and the q25-29 region of the long arm of chromosome 3, as well as changes leading to gains of 3q26-q27. The clinical significance of this finding in B-CLL is uncertain; however, our patient remains well and has not required therapy 17 years after her initial diagnosis. .

2824T

Mosaic Epigenetic Susceptibility Syndrome (MESS) to Wilms tumor.

E. Fiala^{1,2,3}, J. Kennedy^{1,2,3}, M. Harlan Fleischut^{1,2,3}, K. Duffy^{4,5}, M. Ortiz¹, N. Shukla¹, P. Steinherz¹, A. Kentsis¹, A. Latham Schwark^{2,3}, M. Dubard Gault^{2,3}, Z. Stadler^{2,3}, T. Heaton¹, M. LaQuaglia¹, A. Kung¹, D. Friedman¹, K. Offit^{2,3}, J.M. Kalish^{4,5}, M. Walsh^{1,2,3}. 1) Department of Pediatrics, Memorial Sloan Kettering Cancer Center, New York, NY; 2) Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY; 3) Niehaus Center for Inherited Cancer Genomics, Memorial Sloan Kettering Cancer Center, New York, NY; 4) Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA; 5) Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

Sixteen thousand children are diagnosed with cancer annually in the US. Recent molecular studies show that approximately 10% of children with cancer are born with a pathogenic mutation in a known cancer predisposition gene, however this is considered a low estimate. Gene discovery efforts and cancer susceptibilities revealed through non-DNA sequencing assays will add to the understanding of predisposition. Wilms tumor represents less than 5% of pediatric cancer. Predisposition to Wilms tumor is reported in 10-15% of patients, and 1% of patients have a family member with a history of Wilms tumor. Overt syndromes, such as Beckwith-Wiedemann Syndrome (BWS; characterized by overgrowth, glucose dysregulation, omphalocele, ear anomalies, etc.), familial Wilms tumor, and isolated cases of Wilms tumor are each reported having underpinning germline susceptibility. Cancer susceptibility resulting from mosaic DNA sequence mutations in cancer predisposition genes such as *TP53* and *RB1* are recognized, however parallel examples resulting from mosaic epigenetic alterations in non-overtly syndromic populations are less appreciated. Here we describe test results for six patients entering our predisposition clinic with a diagnosis of Wilms tumor. Four patients had bilateral tumors (stage V) while two had unilateral (one stage III, one stage IV). Age at diagnosis ranged from 2.1 to 6.5 years. Only two patients had other clinical manifestations of BWS (one with hemihyperplasia, one with mild neonatal hypoglycemia). All six underwent agnostic methylation testing of the BWS locus at 11p15. Genomic DNA extracted from blood was treated with sodium bisulfite followed by quantitative methylation sensitive PCR to measure the degree of methylation at the IC1 and IC2 regions on 11p15. The degree of methylation detected at IC2 ranged from 53.81% - 56.28% (normal range 50.00% +/- 1.54%), consistent with low-level mosaic gain of methylation. These results suggest that more patients are potentially mosaic for imprinting disorders than previously recognized, and routine sequencing currently deployed by many oncologists is insufficient to identify cancer susceptibility in all pediatric cancer patients.

2825F

Assessment of *BAP1* germline alterations in uveal melanoma. R. Pilarski¹, M.H. Abdel-Rahman^{1,2}, G. Boru², T. Grose², F.H. Davidorf², C.M. Cebulla².

1) Division of Human Genetics, Dept of Internal Medicine, Comprehensive Cancer Center, The Ohio State University, Columbus, OH; 2) Department of Ophthalmology and Visual Science, The Ohio State University.

Objective: The aim of this study is to assess the frequency of germline *BAP1* mutation and deletion/duplication in uveal melanoma (UM) patients with strong personal and family histories of cancers. **Methods:** Germline mutations in *BAP1* were assessed in 172 UM patients with moderate to high risk for hereditary cancer predisposition by direct (Sanger) sequencing. Germline large deletions and duplications were assessed by multiplex ligation probe amplification (MLPA) and epigenetic inactivation by pyrosequencing. The expression of *BAP1* was assessed in 17 non-tumor choroids by qRT-PCR. **Results:** We identified 28 patients with one or more germline variants in *BAP1*, 7 of these were deleterious (4 nonsense and 3 frame shift deletions). The remaining 21 patients had benign/likely benign variants. Germline large deletion/duplication was successfully assessed in 140 patients and in one patient a whole gene deletion was detected. The proband with whole gene deletion was diagnosed with UM at age 16 with no personal or family history of other cancers. The highest frequency of pathogenic *BAP1* alterations - 6/17 (35.3%) - was in patients with UM and personal/family history of ≥ 2 *BAP1*-related cancers followed by patients with young age of onset 4/21 (19%) and familial UM 6/34 (17.6%). No significant evidence of constitutional promoter hypermethylation was detected in the samples tested. Two out of the 17 non-tumor choroid tissues showed decrease in *BAP1* RNA expression compared to normal controls. No detectable germline genomic alteration or aberrant promoter methylation was identified suggesting other mechanisms for epigenetic *BAP1* inactivation. **Conclusions:** *BAP1* is an important candidate for hereditary predisposition to UM but additional candidate genes exist. Assessment for large deletions should be included in clinical testing. In UM patients a strong personal or family history of cancer associated with *BAP1*, early age of onset and familial UM are each risk factors germline mutations in *BAP1*.

2826W

Differential association of Lynch syndrome genes with colorectal and breast cancer. A.D. Zimmer, J. van den Akker, C.L. Neben, A.Y. Zhou. Color Genomics, Burlingame, CA.

Background: Lynch syndrome (LS) is an inherited condition that greatly increases the risk of multiple cancers, including colorectal, endometrial, and ovarian cancer. Pathogenic variants in *MSH2* and *MLH1* are identified in 70-90% of LS families whereas *MSH6* and *PMS2* are estimated to account for 10-30% and *EPCAM* for 1-3%. The spectrum of cancer risks associated with variants in these genes has been reported to vary by gene, with *MLH1* and *MSH2* conferring a higher risk for colorectal and endometrial cancer and an earlier age of onset. Recent evidence has also suggested that some of these genes differentially increase breast cancer risk. **Methods:** We analyzed a cohort of 546 individuals who received NGS panel testing for hereditary cancer risk from Color Genomics and were identified to have a pathogenic or likely pathogenic (hereafter referred to as pathogenic) variant in *MLH1*, *MSH2*, *MSH6*, *PMS2*, or *EPCAM*. Individuals with more than one pathogenic variant or a homozygous variant were excluded from the analysis. Familial relationships were not assessed. Personal history of cancer was based on information reported by the individual or ordering physician, and phenotype analysis was limited to individuals with sufficient information. **Results:** The pathogenic variants in this cohort were evenly distributed across *MLH1* (101), *MSH2* (137), *MSH6* (156), and *PMS2* (152). Colorectal cancer was more closely associated with *MLH1* (28/73, 38%) and *MSH2* (22/98, 22%) than with *MSH6* (5/132, 4%) and *PMS2* (7/117, 6%) ($p < 0.01$), consistent with previous studies. Similarly, the mean age of colorectal cancer onset was lower for *MLH1* (40.3 ± 13.0 yrs) and *MSH2* (42.1 ± 8.3 yrs) than for *MSH6* (50.4 ± 19.2 yrs) and *PMS2* (49.4 ± 10.0 yrs,) (*MLH1/MSH2* vs *MSH6/PMS2*, $p = 0.03$). Among females in the cohort, there were no differences in reported personal history of breast cancer in *MSH2* (9/51, 18%), *MSH6* (10/77, 13%), and *PMS2* (12/69, 15%), ($p = 0.70$). Intriguingly, breast cancer was less reported with *MLH1* (2/45, 4%), but did not reach statistical significance ($p = 0.10$ vs *MSH2/MSH6/PMS2*). The age of breast cancer onset was not different between genes. **Conclusions:** The spectrum of cancers in this cohort supports previous data that *MLH1* and *MSH2* have greater association with colorectal cancer than *MSH6* and *PMS2*. However, we observed that *MSH2*, *MLH6*, and *PMS2* were more associated with breast cancer than *MLH1*, providing evidence for a differentiation of breast cancer risk amongst LS genes.

2827T

Genotyping ccfDNA in cerebrospinal fluid as a new source of liquid

biopsy. O. Sonmezler¹, I. Boga¹, C. Mujde¹, A. Bisgin^{1,2}. 1) Cukurova University AGENTEM (Adana Genetic Disease Diagnosis and Treatment Center), Adana, Turkey; 2) Cukurova University, Faculty of Medicine, Medical Genetics Dept of Balcali Hospital and Clinics, Adana, Turkey.

Objective: Central nervous system (CNS) involvement is clinically important and a major challenge in oncology due to high morbidity and mortality rates. Thus, liquid biopsy of cerebrospinal fluid (CSF) derived ccfDNA might be effective in tracking clinically relevant genomic alterations in cancer patients with CNS metastasis. Where in applied next generation sequencing of CSF and peripheral blood derived ccfDNA and compared with the FFPE tissue sample from the primary tumor by a targeted multi-gene panel in a gastric cancer patient. **Method:** FFPE tissue sample from primary localized tumor, liquid biopsy sample from peripheral blood and also CSF were obtained from the patient. gDNA (genomic DNA) isolated from FFPE tissue sample while ccfDNA (circulating cell-free DNA) isolated from both peripheral blood and the cerebrospinal fluid. We measured the concentration of DNA using our usual protocol and each of them compared. Multi-gene panel including *AKT1*, *ALK*, *BRAF*, *DDR*, *ERBB2*, *ESR1*, *KIT*, *KRAS*, *MAP2K1*, *NRAS*, *NTRK*, *PDGFRA*, *PIK3CA*, *PTEN*, *ROS1*, *RICTOR*, *EGFR*, *MET* and *FGFR1* genes and copy number changes in *RICTOR*, *EGFR*, *MET*, *FGFR1* and *ERBB2* genes were sequenced and analyzed via GeneReader NGS System. **Results:** Unique R776H variant in *EGFR* gene which reported as “panitumumab” drug sensitivity for gastric cancer was captured in CSF ccfDNA when compared with plasma and primary tissue. Also *FGFR1* amplification was identified in plasma, while non in FFPE tissue sample. **Conclusion:** Liquid biopsy of CSF for the precise characterization and molecular diagnosis of tumors with CNS metastasis can be used to both complementary and/or more efficiently facilitate diagnosis and prognosis.

2828F

Clinical relevance of circulating tumor DNA assessed through amplicon-based next-generation sequencing in cancer. H. Zembutsu¹, H. Osumi², K. Yamaguchi², E. Shinozaki², M. Ozaka³, T. Sasaki³, N. Sasahira³.

1) Cancer Precision medicine Center, Cancer Institute, TOKYO, JAPAN; 2) Department of Gastroenterological Chemotherapy, Cancer Institute Hospital, Japanese Foundation for Cancer Research, TOKYO, JAPAN; 3) Department of Gastroenterology, Cancer Institute Hospital, Japanese Foundation for Cancer Research, TOKYO, JAPAN.

The development of circulating tumor DNA (ctDNA) panel covering hundreds of mutations is important to establish high sensitive diagnostic system for cancer. We enrolled 101 and 76 patients with metastatic colorectal (CRC) and pancreatic cancer (PC), respectively. The genomic profiling of 14 genes in plasma were performed to evaluate the feasibility of this panel. Somatic mutations were detected in 87.1% and 90.8% of patients with CRC and PC in their plasma, respectively. The Mutations in *TP53*, *KRAS* and *APC* genes were detected in 69.3%, 38.6% and 23.7% in CRC patients, and 64.3%, 60.5%, 13.2% in PC patients. ctDNA level were significantly associated with metastasis (liver, $P<0.00001$, lymph node, $P=0.008$, number of metastatic organs, $P<0.0001$), tumor markers (CEA and LDH, $P<0.0001$, CA19-9, $P=0.006$), and tumor diameter (maximum and sum of diameter, $P<0.0001$). The overall concordance rate of *RAS* status between ctDNA and matched CRC tissue was 77.2%. PC patients with lower ctDNA level showed significantly longer overall survival ($P=0.00000026$). These results suggest that profiling of mutated genes in the panel could be biomarkers to monitor changes in mutational status and tumor burden.

2829W

Rapid detection of genomic rearrangements and gene fusions in human leukemias using Oxford nanopore sequencing. *P. Zuzarte¹, T. Murphy², K. Garimella³, M. Minden², J. Simpson^{1,4}.* 1) Genome Technologies, Ontario Institute for Cancer Research, Toronto, Ontario, Canada; 2) Princess Margaret Cancer Centre, University Health Network, Toronto, Canada; 3) Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, UK; 4) Department of Computer Science, University of Toronto, Toronto, Canada.

Genetic rearrangements are a hallmark of many human hematological disorders. The correct identification of genetic rearrangements plays an important role in the diagnosis and treatment of these disorders. Although cytogenetics and FISH are considered the gold standard for identifying genetic arrangements in human Leukemias, these tests can take several days to complete and often miss small or cryptic genetic rearrangements. The Oxford Nanopore MinION is a small handheld sequencer that is capable of sequencing long fragments of DNA, often in excess of 10kbp, or nascent full-length RNA directly. In this proof of concept study, we show that nanopore sequencing can be used to rapidly detect and identify novel rearrangements in clinical samples from 6 patients who have been diagnosed with various forms of leukemia. Nanopore sequencing detects cryptic chromosomal rearrangements that are below the detectable level of resolution for cytogenetics (5Mb) and is capable of performing real-time basecalling, enabling the the rapid detection of rearrangements, often within 24 hours of initiating a sequencing run. Chromosomal breakpoints that were identified through DNA sequencing were validated by PCR amplification. Single molecule sequencing of full length nascent poly-adenylated RNA was also used for rapid validation of transcribed fusion gene products and may also offer other insight into fusion gene transcription. .

2830T

Validation of a targeted massively parallel sequencing panel for the clinical diagnostic testing of fusion genes associated with hematological malignancies. *M.R. Avenarius¹, E. Stonerock¹, S.C. Reshmi^{1,2,3}, R. Pfau^{1,2,3}, V. Magrini^{1,3}, C.E. Cottrell^{1,2,3}, J. Gastier-Foster^{1,2,3}, E. Zmuda^{1,2,3}.* 1) Institute for Genomic Medicine, Nationwide Children's Hospital, Columbus, OH; 2) Department of Pathology, The Ohio State University College of Medicine, Columbus, OH; 3) Department of Pediatrics, The Ohio State University College of Medicine, Columbus, OH.

Hematological malignancies are derived from oncogenic mutations in the proliferative, blood-forming cells and include leukemias, lymphomas and myelomas. Recurrent gene fusions drive oncogenesis in a significant proportion of these hematologic diseases and their presence often carries diagnostic, prognostic or therapeutic ramifications. To this end we clinically validated a customized Archer® FusionPlex® Pan-Heme assay that targets 148 genes for the detection of both known and novel fusion events. Through this work we optimized assay performance on a range of Illumina instrumentation, identified strategies to systematically assess assay quality, and developed bioinformatic tools beyond the standard Archer pipeline to ensure robust assessment of clinical results. Experimentation to test assay performance included the establishment of platform sensitivity and specificity, reproducibility and definition of the limits of detection. Through these experiments we also identified inconsistencies in the ability to detect immunoglobulin rearrangements and thus developed improved analytical methods for their detection. After meeting our internal standards for assay performance, a focused version of this assay, limited to genes associated with Acute Lymphoblastic Leukemia rearrangements, was launched. To date, we have completed testing on 93 cases. 54% of these cases were children/young adults (<30) while the remaining were adult patients (46%). Our cohort was primarily comprised of presumed Acute Lymphoblastic Leukemia diagnoses, however, a limited number of other hematological diseases were represented (ie myeloproliferative disorders). Of the 93 cases analyzed, 37.5% (n=35) tested positive for a verified, clinically meaningful fusion event. The majority of fusion positive cases harbored recurrent fusion events with previous associations to hematological malignancies (91.5%), while 8.5% (n=3) harbored novel fusion events. Compared to other fusion detection methods such as RT-PCR or FISH, our validation of the custom FusionPlex® Pan-Heme assay demonstrates the ability to detect novel gene fusion events with increased sensitivity ($\geq 10\%$ blast), specificity (100%), and precision (100% of fusions tested reliably detected within/between runs). In addition, the actionable fusions detected by this assay have significant diagnostic, prognostic or therapeutic implications and illustrates the increased clinical utility of this diagnostic test.

2831F

Comprehensive evaluation of data quality in public population databases guides pathogenic allele frequency threshold determination for variant classification downgrades in hereditary cancer genes. *M. Jones, H. Cox, L. Esterling, J.P. De La O, T. Judkins, S. Manley, K. Bowles, D. Mancini-DiNardo, B. Coffee.* Myriad Genetic Laboratories, Inc., Salt Lake City, UT.

Background: Data in research grade population databases, such as gnomAD, are often used by clinical testing laboratories as evidence in variant classification as recommended by ACMG/AMP guidelines. Specifically, a variant allele frequency (AF) higher than the expected disease prevalence associated with pathogenic variants in that gene may be used as evidence that the variant is benign. However, careful evaluation of the data available in such databases is necessary to avoid misclassification of variants. Here we report on a comprehensive quality assessment of gnomAD data for germline variants in 24 cancer-predisposition genes. **Methods:** Quality of the gnomAD data was analyzed for germline variants in 24 genes included on our laboratory's hereditary pan-cancer panel at every genomic position within the laboratory review window. Quality metrics analyzed included coverage, sequence complexity, NGS read frequencies, mapping quality, pseudogene homology, and somatic contamination. Pathogenic AF thresholds for each gene were determined by assessing variants identified by our laboratory in the course of clinical testing of >500,000 patients that are also present in gnomAD, along with estimates on penetrance, allelic and locus heterogeneity. **Results:** Quality issues were identified for germline variants in all 24 genes that would result in inaccurate AFs, including artifacts caused by homopolymer tracts, presence of somatic variants likely due to clonal hematopoiesis, and variants identified in pseudogenes that incorrectly map to the functional gene. As a result of this analysis, gnomAD data were fully excluded as evidence for variant classification for two genes and partially excluded for three genes due to inadequate quality. After assessment of quality, a maximum expected pathogenic AF threshold was determined. These AF thresholds were used to reclassify 40 variants in 12 genes from uncertain clinical significance to benign, impacting over 6,000 patients. **Conclusions:** Although public population databases can be an excellent resource for variant classification in research applications, the data require thorough review prior to integration into clinical applications to avoid potential variant misclassification. Use of vetted data that only includes high quality regions can result in greater accuracy in setting pathogenic AF threshold cutoffs for use in downgrading variants.

2832W

Benefits in the surveillance of the copy number of *RUNX1* in childhood B-ALL patients from targeted gene sequencing. *M. Wang, P. Cao, Y. Zhang, F. Wang, Y. Zhang, H. Liu, T. Wang, H. Liu.* Pathology & Laboratory Medicine Division, Hebei Yanda Lu Daopei Hospital, Si Pu, Beijing, China.

Objective Intrachromosomal amplification of chromosome 21 (iAMP21) defines a distinct cytogenetic subgroup of childhood BCP-ALL and have a dismal outcome when treated as standard risk. It is defined by multiple copies of the *RUNX1* gene (3 or more additional *RUNX1* signals) by FISH and is difficult to be detected by karyotyping. However, many limitations may have led to the underestimated diagnosis of patients with iAMP21. **Methods** We retrospectively analyzed the next generation sequencing data (targeted 58 genes involved in hematological malignancies) of 510 patients with B-ALL by ONCOCNV to calculate the copy number variation of *RUNX1* gene. **Results** A total of 49 (9.6%) cases were identified with *RUNX1* copy number gain. Chromosome karyotyping results are available in 30 cases. Among them 11 cases have duplication of chromosome 21, the remaining 19 cases have chromosome abnormalities not involved in chromosome 21 duplication and the median copy number gain was 3.5 copies (rang 3-5.5 copies). Analysis of the prognosis of the 19 patients found that 9 (47.4%) cases relapsed. Three of 19 cases performed FISH and were diagnosed with iAMP21, the copy number of *RUNX1* gene from sequencing data are concordant with FISH signals for all the three patients. The *RUNX1* copy number gain analyzed from next generation sequencing may suggest the diagnosis of iAMP21 and need to be confirmed by further testing. In addition, it is noteworthy that these patient concomitants with unique spectrum of secondary genetic abnormalities, including chromosomes gains (X, 6q and 10) or deletions (7q and 11q), which may also be used to improve the diagnosis of iAMP21. **Conclusion** In summary, we developed a potential strategy to improve the identification of iAMP21, thus improving the prognosis of patients within this high-risk group. .

2833T

Cis-acting genetic causes of *MLH1* promoter methylation in Lynch syndrome patients with secondary constitutional epimutations. J. Leclerc¹, C. Flament², T. Lovecchio², L. Delattre², E. Ait Yahya², S. Baert-Desurmont¹, N. Burnichon², M. Bronner², O. Cabaret¹, S. Lejeune², R. Guimbaud², G. Morin¹⁰, J. Mauillon¹¹, P. Jonveaux¹², P. Laurent-Puig¹³, T. Frébourg¹, N. Porchet¹, M.P. Buisine¹, other members of the GGC consortium. 1) Inserm UMR-S 1172, JPA Research Center, Lille University, and Lille University Hospital, Department of Biochemistry and Molecular Biology, Lille, France; 2) Lille University Hospital, Department of Biochemistry and Molecular Biology, Lille, France; 3) Lille University Hospital, Bioinformatics Unit, Molecular Biology Facility, Lille, France; 4) Rouen University Hospital, Department of Genetics, and Normandie Univ, UNIROUEN, Inserm U1245, Normandy Centre for Genomic and Personalized Medicine, Rouen, France; 5) Inserm UMR970, Paris-Cardiovascular Research Center, and Georges Pompidou European Hospital, Department of Genetics, Paris, France; 6) Nancy University Hospital, Department of Genetics, Vandœuvre-lès-Nancy, France; 7) Gustave Roussy Institute, Department of Biology and Pathology, Villejuif, France; 8) Lille University Hospital, Department of Genetics, Lille, France; 9) Claudius Regaud Institute, IUCT-Oncopôle, Department of Oncogenetics, Toulouse, France; 10) Amiens University Hospital, Department of Genetics, Amiens, France; 11) Rouen University Hospital, Department of Genetics, Normandy Centre for Genomic and Personalized Medicine, Rouen, France; 12) Nancy University Hospital, Department of Genetics, and Inserm U954, University of Lorraine, Vandœuvre-lès-Nancy, France; 13) Georges Pompidou European Hospital, Department of Biology, Paris, France.

Constitutional epimutations are an alternative to genetic mutations in the etiology of genetic diseases. These epimutations can be primary, defined as purely epigenetic events labile in the germline. They can also be secondary, caused by a *cis*-acting genetic defect transmitted following a Mendelian inheritance pattern, with re-establishment of the epigenetic change in the offspring. Constitutional epimutations of *MLH1* have been identified as a rare etiology of Lynch syndrome. Patients with primary *MLH1* epimutations, as well as a few families with secondary *MLH1* epimutations, have been reported. A secondary epimutation is usually suspected when 2 epimutation-carriers are diagnosed within a family and validated when the *cis*-acting genetic cause is identified. This is crucial to determine the risk of transmission in the family and to offer accurate genetic counseling to relatives. We have recently reported a first series of Lynch syndrome patients with a broad spectrum of genetic defects associated with their *MLH1* epimutation (Genetics in Medicine, 2018). We present here these results extended to a second series of patients, representing a total of 28 probands. Various types of variations were found to be associated with *MLH1* secondary epimutations, including single nucleotide variations, copy number variations and other structural variants such as insertion of an *Alu* element. We provide further insight into the complexity of molecular mechanisms leading to secondary epimutations and highlight the need for extensive *MLH1* gene analysis in Lynch syndrome patients with constitutional epimutations.

2834F

Development of a novel, rapid, ultra-sensitive and intraoperative IDH mutation detection method. T. Avsar^{1,2,3}, A. Sursal^{2,3}, G. Turan^{2,3}, BN. Yigit^{2,3}, FA. Mazi³, D. Altunsoy^{2,3}, K. Cantasir⁴, G. Duyu⁴, T. Kilic⁵, M. Acar¹. 1) Bahcesehir University, School of Medicine, Department of Medical Biology, Istanbul, Turkey; 2) Bahcesehir University, Health Sciences Institute, Neuroscience Program, Istanbul, Turkey; 3) Bahcesehir University, School of Medicine, Neuroscience Research Laboratory Istanbul, Turkey; 4) Bahcesehir University, School of Medicine, Turkey; 5) Bahcesehir University, School of Medicine, Neurosurgery Department, Istanbul, Turkey.

IDH1/2 mutations are indicative of good prognosis upon treatment and they may have intraoperative importance for disease management. Intraoperative diagnosis became a new technical approach in glioma management. If a neurosurgeon would know the mutation status in IDH genes, he/she would change the extent of resection in tumor removal. Because mutated tumors responsive to chemo/radiotherapy. We aimed to conduct a comprehensive IDH mutation analysis by using PCR based methodology during surgery. Our aim is to provide rapid, highly sensitive and specific intraoperative mutation detection system with robust accuracy. To this aim Amplification refractory mutation system (ARMS) was modified and optimized. The novelty of our methodology is adding an extra mismatch to increase specificity. ARMS was used to detect the presence of *IDH1*R132H, *IDH1*R132K and *IDH1*R132L mutation in 236 glioma sample selected from our tumor bank followed by Sanger sequencing. All results were also evaluated by immunohistochemistry (IHC) for IDH1 R132H. We also evaluated the diagnostic and screening potential of the test. By addition of the third mismatch to conventional ARMS we successfully detect IDH1 R132H mutation by 98.22% coherency with Sanger sequencing. Comparison between IHC and 3 mismatch-ARMS showed, 3-mismatch ARMS is superior to IHC in terms of accuracy. Based on our results, frequencies of *IDH1*/2 mutations among different glioma grades are, 6.67%, 69.84%, 65.08% and %8.75 for grade I, II, III and IV respectively. 3-mismatch ARMS is a novel methodology with great potential to be used as an intraoperative diagnostic test. It can detect *IDH* mutations in 67 minutes without the need of sequencing with the ability to be adapted for detection of various mutations through easy primer design.

2835W

Validation of an amplicon-based hematopoietic gene panel to characterize aberrant clonal expansions/clonal hematopoiesis. D. Castillo, T.P. Slavin, S. Sand, R. Mejia, J. Wang, X. Wu, J. Longmate, T.F. Springer, J. Herzog, B. Nehoray, J.N. Weitzel. City of Hope, Duarte, CA.

Purpose: Clonal hematopoiesis (CH) is an aberrant clonal expansion (ACE) in individuals with or without overt hematologic disease (with the latter referred to as CHIP), often marked by a pathogenic variant (PV) in hematopoietic pathway gene(s) at a minor allele frequency (MAF) < than expected for a heterozygous germline finding. The prevalence increases with age, and we recently demonstrated that 20% of commercial lab multigene panel-detected pathogenic *TP53* variants represented ACE, rather than a germline finding. The natural history of this phenomenon, and whether there is concomitant or sequential acquisition of PVs in other hematopoietic pathway genes is unknown, so we created and validated an economical next-generation sequencing (NGS) assay (81 gene QIAseq amplicon based panel) to assess reproducibility of the MAF (test/retest), compared to a standard Agilent targeted capture library.

Methods: We evaluated DNA from 14 participants with known *TP53* PV-associated ACE. All individuals were consented and enrolled in the Clinical Cancer Genomics Community Research Network prospective registry. Libraries were prepared in triplicate using separate aliquots of DNA from blood and other tissues, and were sequenced on an Illumina HiSeq 2500 (Illumina Inc., San Diego). Unique molecular indexes (UMI) tag fragments in the original sample were used to reduce the chances of annotating amplification error. Variant call format files were produced and annotated using American College of Medical Genetics variant calling guidelines through Ingenuity Variant Analysis version 4 (Qiagen Inc, Alameda, CA). **Results:** All 14 *TP53* PVs were detected; average read depth of 275X; MAF ranged from 10.55%- 45.56%[ST1] [CD2] . Within-subject standard error was 3%, with no significant excess variability ($P = 0.17$). Five of 14 (36%) subjects had additional CH-associated gene PV(s). These included: 1) *TET2* c.1360A>T (MAF=21.11%), 2) *TET2* c.3467delA (1.2%), 3) *ATM* c.8096C>T (3.42%), 4) *TP53* c.669delA (0.95%), 5) *TP53* c.350G>T (14.91%) and 6) *ATM* c.8096C>T (3.42%).[NB3] **Conclusions:** Our assay successfully identified all expected *TP53* PV(s), demonstrated reproducible MAF (minimal intra-case variability), and concurrently identified other CH-associated PV(s). This indicates that our custom panel can reliably detect and quantify ACE as it evolves, at a low per sample cost for future studies. The assay can further be extended in the future to other tissue types. .

2836T

Next generation sequencing of hereditary paraganglioma-pheochromocytoma syndrome. W. Lee, W. Min, S. Chun, H. Kim, S. Ahn. University of Ulsan College of Medicine, Seoul, Korea, Republic of.

Introduction Hereditary paraganglioma-pheochromocytoma(PGL/PCC) syndrome is an inherited condition characterized by the growth of noncancerous tumors in paraganglia. The diagnosis of this syndrome is based on several methods including genetic testing. SDHA, SDHB, SDHC and SDHD are four nuclear genes that closely related to the syndrome. Also SDHAF2, MAX, NF1, RET, TMEM127 and VHL genes are known to be relevant to this condition. In this research, we were trying to detect and validate genetic variants related to PGL/PCC syndrome from the pheochromocytoma patients using the Hereditary Tumor Next Generation Sequencing (NGS) panel which covers 128 genes. **Method** We analyzed the pheochromocytoma patients suspected as inherited nature during 6 monthes, who were referred for Hereditary Tumor NGS analysis in Asan Medical Center. Their diagnosis, variant gene, variant types, and inheritance pattern were investigated. The genetic variants were classified as Pathogenic (P), Likely Pathogenic (LP), and Uncertain significance (US) according to the ACMG STANDARDS AND GUIDELINES (2015). **Result** During the study period, total 30 pheochromocytoma patients were referred. Genetic variants related to the PGL/PCC syndrome were discovered from 16 patients. The genetic variants discovered from the patients were varied across the whole genes related to the PGL/PCC syndrome. A pathogenic variant was found in TP 53, and 8 LP variants were found in ACD, SDHB, TMEM127 and VHL. Four US variant were found in CDKN1B, MSH2, RAD50 and SDHA. Interestingly, pathogenic/likely pathogenic variants were detected in the genes (ACD and TP53), not well known to be related to the PGL/PCC syndrome. **Conclusion** Through this study, we discovered various genetic variations associated to PGL/PCC syndrome. Almost half of the pheochromocytoma patients showed P/LP/US variant, therefore, NGS work which covers many candidate genes for pheochromocytoma patients could be helpful for diagnosis of the disease when compared with the existing single-gene targeting method. NGS can help to find more meaningful genetic variants from possible PGL/PCC syndrome patients. In addition, two P/LP variants were discovered in the genes whose direct relations to PGL/PCC syndrome were not known. Further study with this panel could expand our knowledge on hereditary paraganglioma-pheochromocytoma(PGL/PCC) syndrome.

2837F

Performance of an NGS multi-gene panel for detection of hotspot variants in colorectal cancer patients. I. Boga¹, C. Mujde¹, S. Tug Bozdogan^{1,2}, O. Sonmezler¹, A. Bisgin^{1,2}. 1) Cukurova University, Adana, Turkey; Cukurova University AGENTEM (Adana Genetic Disease Diagnosis and Treatment Center), Adana, Turkey; 2) Cukurova University, Faculty of Medicine, Medical Genetics Dept of Balcali Hospital and Clinics, Adana, Turkey.

Objective: Using Next generation sequencing (NGS) to identify somatic mutations in cancerous tissues has become accepted as routine in molecular diagnostics in the name of "Precision Medicine". To gain insight into the issues of being continually asked to evaluate more genes with limited sample volume. We investigated the success rate of AIT Panel (Qiagen) targeted ("hotspot") cancer NGS test for genomic DNA derived from FFPE tissue across colorectal cancer. **Methods:** FFPE tissue samples were collected and DNA isolations were performed from 100 colorectal cancer patients. Samples with limited tissue or low amount of extracted DNA were still analyzed by NGS due to our institutional optimizations. During the study *KRAS*, *NRAS*, *BRAF*, *PDGFRA*, *KIT*, *PIK3CA*, *EGFR*, *ALK*, *RAF1*, *ESR1*, *ERBB2* and *ERBB3* genes were next generation sequenced via GeneReader NGS system (Qiagen). Bioinformatics' analysis was made accordingly with patients' clinical findings. Then, treatment modalities related to patients' status were determined in terms of genetic alterations related to therapy resistance and/or sensitivity for clinical reporting. **Results:** As a result, total of 77 variants were detected in 57% (n=57) of the patients. Unfortunately, 43% (n=43) patients had no actionable mutations. Our analysis showed that this NGS system with AITP targeted cancer panel have shown high success rate with no needed. We detected all the actionable variants in 8 different genes which are *KRAS*, *NRAS*, *KIT*, *PIK3CA*, *BRAF*, *RAF1*, *EGFR* and *PDGFRA*. Among all these detected variants, 22 (%28.6) were related to the sensitivity to certain therapies, while %32.5 (n=25) were related to drug resistance in literature. Furthermore, 20.8% (n=16) of the variants were reported as related with both drug resistance and sensitivity. Moreover, we detected 6.5% (5) variants of uncertain significance and 4 (5.2%) variants related to an ongoing clinical trial. Rest of the variants (6.4%) indicated disease prognosis. **Conclusion:** Within the context of precision medicine, we reviewed our results on colorectal cancer samples that were not triaged by sample size, type or DNA quantitation prior to NGS for targeted "hotspot" multi-gene panel. By the virtue of NGS, multi-gene panel studies for somatic variant detection become more achievable, and this kind of studies by commercial hotspot panels may facilitate further improvements for targeted personalized therapies.

2838W

Germline reporting in the National Cancer Institute - Children's Oncology Group (NCI-COG) Pediatric MATCH National Precision Oncology Trial. S.E. Plon^{1,2}, S. Scollon^{1,2}, S. Joffe³, J.A. Biegel⁴, S. Kulkarni⁵, G. Miles⁶, L. Diller⁷, C.V. Fernandez⁸, D. Malkin⁷, J.A. Oberg⁸, M.V. Relling⁹, J.D. Schiffman¹⁰, L.A. Schwartz^{3,11}, D.R. Stewart¹², N.L. Seibel¹³, D.W. Parsons^{1,2}. 1) Baylor College of Medicine, Houston, TX; 2) Texas Children's Hospital, Houston, TX; 3) University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; 4) Children's Hospital Los Angeles, Los Angeles, CA; 5) Dana Farber Cancer Institute, Boston, MA; 6) Dalhousie University, Halifax, Nova Scotia, Canada; 7) The Hospital for Sick Children, Toronto, Ontario, Canada; 8) Columbia University Medical Center, New York City, NY; 9) St. Jude Children's Research Hospital, Memphis, TN; 10) Huntsman Cancer Institute at the University of Utah, Salt Lake City, UT; 11) Children's Hospital of Philadelphia, Philadelphia, PA; 12) Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD; 13) Cancer Therapy and Evaluation Program of the National Cancer Institute.

Background: Genomic changes identified by tumor testing are increasingly used to inform treatment decisions and determine eligibility for clinical trials. Although at least 10% of cancer patients carry germline cancer susceptibility pathogenic variants, precision oncology trials have generally not incorporated germline reporting. We describe here the development of a process for return of matched germline results during the genomic screening portion of the national COG-NCI Pediatric MATCH trial for children with relapsed tumors. **Design/Method:** A multidisciplinary germline reporting committee and MATCH study leadership developed the germline reporting process. At entry, each subject provides a blood sample in addition to the required tumor sample (expected enrollment n=1000). Tumor and blood genomic analyses are performed on the same Pediatric MATCH Ion Torrent panel but reported independently. The committee selected 43 of 159 genes on the panel with clearly documented adult or pediatric cancer susceptibility for germline reporting. Germline reports are generated for each subject by experienced hereditary cancer genetic testing laboratories and include potential limitations of the germline analysis. Data on each subject are reviewed weekly by the testing laboratory, study leadership, and germline reporting teams. We also provide a genetics resource for the participating COG oncologists including genetic counselors, geneticists and oncologists with surveillance expertise. **Results:** Pediatric MATCH opened in 2017 and is now open at over 100 clinical sites in the U.S. Blood samples for genomic testing have been obtained from >95% of subjects enrolled to date. As expected, many somatic findings in genes like *TP53* were seen in the tumor but not the germline. Germline interpretation scenarios encountered to date represent a diversity of findings: clinically significant germline variants in patients with known genetic diagnoses (e.g. *NF1*, *MLH1*), findings in genes typically associated with adult cancer risk (e.g. *CHEK2*, *BRCA1*), and germline pathogenic variants lost in the tumor (*NF1*, *BRCA1*). **Conclusions:** Coordinated germline and tumor reporting has the potential to maximize the clinical value of genomic testing, including for patients in whom a biomarker for a molecularly-targeted treatment is not identified. The germline reporting process developed for Pediatric MATCH can serve as a model for germline reporting in multi-institutional precision oncology trials.

2839T

NEBNext Direct Custom Ready Panels overcome challenges associated with targeted re-sequencing. A. Barry¹, K. Patel², S. Bowman², A. Emeric², S. Adams², B. Desmond², J. Dunn², S. Corbett², C. Elfe², E. Mauceli², E. Dimalanta¹, F. Stewart¹, T. Davis¹, C. Hendrickson². 1) New England Biolabs, Ipswich, MA; 2) Directed Genomics, Ipswich, MA.

Efficient utilization of targeted gene panels for clinical research is challenged by the wide variation in gene constituents specific to a given study. While focused gene panels efficiently provide the necessary depth of coverage for low frequency variant detection, the high costs and design challenges associated with panel design present challenges. Furthermore, inherent issues with amplification-based enrichment include low uniformity of target coverage, and challenges when modifying content of a given panel. NEBNext Direct Custom Ready Panels employ a novel approach to selectively enrich nucleic acid targets ranging from a single gene to several hundred genes, without sacrificing specificity. The approach rapidly hybridizes both strands of genomic DNA with biotinylated probes prior to streptavidin bead capture, enzymatic removal of off-target sequence, and conversion of captured molecules into sequencer-ready libraries. This results in a unique read coverage profile that results in uniform coverage across a given target. Unlike alternative hybridization methods, the approach does not necessitate upfront library preparation, and instead converts the captured molecules into dual-indexed illumina sequencer compatible libraries containing an 8 basepair sample ID and a 12bp Unique Molecule Index (UMI). The result is a 1-day protocol that enables the preparation of sequence-ready libraries from purified genomic DNA specific to the gene content included in the panel. We have designed and optimized baits specific to the full exonic content of >800 genes associated with cancer, neurological disorders, autism, cardiovascular disease, and other conditions. These are designed, balanced, and pooled on a per gene basis, and can be combined into NEBNext Direct Custom Ready Panels, allowing rapid turn-around of specific custom gene subsets. Here, we present the ability to rapidly deploy custom gene panels across a variety of panel sizes and content, while maintaining high specificity, uniformity of coverage across target content, and sensitivity to detect nucleic acid variants from clinically relevant samples.

2840F

Systematic “minor genes” testing in inherited cardiac diseases: Friend or foe? V. Novelli¹, F.D. Tiziano¹, M.C. Russo¹, G. Pelargonio², F. Crea², P. Zeppilli², M. Genuardi³. 1) Institute for Genomic Medicine, Fondazione Policlinico A. Gemelli, Rome, Rome, Italy; 2) Cardiology Department, Fondazione Policlinico A. Gemelli, Rome, Rome, Italy; 3) Sport Unit, Fondazione Policlinico A. Gemelli, Rome, Rome, Italy.

Introduction: Genetic testing in inherited cardiac diseases has been revolutionized by NGS technology: the NGS-based panel approach, indeed, has enabled to test a large number of genes simultaneously, differently from the traditional Sanger, limited to the most prevalent and characterized genes. Therefore, recently, so called “minor genes” have been included in diagnostic NGS panels, generating remarkable interpretative problems, mostly related to the high prevalence of variants of unknown significance (VUS). Here, we report our experience with expanded gene panels for genetic testing. **Materials and Methods:** 103 consecutive patients, 26 suspected Hypertrophic Cardiomyopathy (HCM), 33 Arrhythmogenic Cardiomyopathy (ACM), 18 Long QT syndrome (LQTS), 19 Brugada syndrome (BrS) and 7 Idiopathic Ventricular Fibrillation (IVF) patients have been analyzed using 3 different Ion custom sequencing panels, including 11 genes for HCM, 9 for ACM, 12 for LQTS/IVF/BrS. **Results:** Thirty probands carried at least one potentially pathogenic variant (30%) have been identified. According to the ACMG guidelines, the majority of the identified variants (70%) were classified as VUS, while 18% were Likely Pathogenic (LP) and 12% Pathogenic (P). Looking at the VUS, the 24% of these have been identified in minor genes, mostly in genes related to inherited arrhythmias diseases. On the other hand, except for one variants classified as LP, all the other variants reported in minor genes did not reach the pathogenicity due to lack of functional studies or limited data reported in the literature. **Conclusions:** These data showed that the inclusion of minor genes in the genetic screening, even if restricted to a limited number of genes, gives a very limited contribution, increasing the rate of VUS and making the tests clinically irrelevant.

2841W

Enabling high quality hypertrophic cardiomyopathy genetic testing using multiplexed patient-like reference material. Z.C. Deans¹, F. Khawaja¹, F. Moon¹, D. Brudzewsky², C. Huang², R. Santhanam². 1) GenQA/UKNEQAS, Edinburgh, United Kingdom; 2) SeraCare Life Sciences Inc, Milford, MA, USA.

INTRODUCTION UK NEQAS for Molecular Genetics promotes high quality accurate genetic testing through external quality assessment (EQA) and educational exercises. UK NEQAS provided participants a multiplexed reference sample (SeraCare) with known pathogenic and non-pathogenic variants associated with hypertrophic cardiomyopathy (HCM) in the genes *MYBPC3*, *MYH7*, *TNNI3*, *TNNT2* and *TPM1*. Participants reported all detected variants as well as the variant classification. **METHODS** Laboratories participating in the UK NEQAS HCM 2017 EQA received the reference sample and applied routine testing strategy and variant classification for HCM. Variant type ranged from substitutions and small insertions to large deletions within repetitive regions. This was an educational exercise and no scores were assigned. Details of the variants present in the reference sample, collation of the submitted results, methodologies performed, and general feedback comments were provided to the participants. **RESULTS** The reference sample was distributed to 17 laboratories from eight countries; the majority from UK (53%), followed by Australia (12%). A next generation sequencing approach was used by all but one which performed Sanger sequencing. There was considerable variation in the number of genes each laboratory analyzed ranging from 4 to 141 genes, all included the five genes supplied with known variants except one laboratory did not test *TPM1*. The standard of variant detection was generally very high with 16/17 laboratories detecting $\geq 70\%$ of variants included in this reference sample. The large deletion situated within a repetitive region was detected the least, independently of testing methodologies. The laboratories reported expected classifications for most of the variants and these will be discussed. Three laboratories reported additional variants; two of which were confirmed by SeraCare and one variant of uncertain significance in the *PRKAG2* gene which was only reported by a single laboratory. **CONCLUSION** HCM mutation testing is not standardised, with laboratories testing different gene targets and variable limitations of the testing approaches. The use of reference material with known clinically relevant and challenging variants enables laboratories to determine the limitations of the test and review the scope of other tests performed. Such benchmarking exercises demonstrate the benefits of educational elements of external quality control in helping to promote high quality molecular testing.

2842T

A targeted enrichment sequencing assay for inherited cardiac condition genes. K. Patel¹, S. Bowman¹, A. Emerman¹, C. Hendrickson¹, S. Adams¹, B. Desmond¹, J. Dunn¹, A. Barry², T. Davis², S. Corbett¹, C. Elfe¹, E. Mauceli¹. 1) Directed Genomics, Ipswich, MA; 2) New England Biolabs, Ipswich, MA.

Genetic information is useful for the diagnosis, prognosis, and management of cardiac disorders. In addition to screening the patient, screening of the patient's family members may help determine whether regular cardiac testing is recommended. To this end, next generation sequencing technologies are rapidly being adopted for the screening of loci that contribute to a variety of inherited cardiac conditions. Target enrichment sequencing assays are ideal for this purpose, as greater coverage of the genomic regions of interest can be achieved for reliable variant detection, with reduced sequencing costs and data analysis burden compared to whole genome/exome sequencing. The NEBNext Direct cardiac disease focused panel was designed to target 87 genes associated with cardiomyopathy, inherited arrhythmia, familial hypercholesterolemia, and pulmonary arterial hypertension. This panel was developed to provide maximum coverage of all protein coding regions (~390 kb) with 10 base pair padding to also enable the detection of splicing variants. Individual baits were designed to capture both strands of the targeted molecules, followed by balancing of the baits by performance to increase or decrease coverage where necessary, resulting in uniform coverage that is superior to that typically seen with amplification based assays. Greater than 95% of the reads mapped to targeted regions and greater than 99.9% of target bases were covered at greater than 20X read depth. The NEBNext Direct target enrichment protocol is a single day, hybridization-based assay that produces a sequencer-ready library with an 8 base pair sample index and a 12 base pair Unique Molecular Identifier (UMI).

2843F

Clinical characterization of hypertrophic cardiomyopathy caused by the ancestral mutation p.Q1233* in the MYBPC3 gene. M. Polyak¹, G. Radghabova¹, Yu. Surikova¹, N. Chakova², S. Komissarova², O. Blagova², A. Motreva³, O. Kalmykova³, M. Bos⁴, M. Ackerman⁵, O. Chumakova⁶, M. Golubenko⁷, J. Ingles⁸, C. Semsarian⁹, S. Dzemeshevich¹, E. Zaklyazminska-ya^{1,9}. 1) Petrovsky National Research Centre of Surgery, Moscow, Russian Federation; 2) Institute of Genetics and Cytology of NAS of Belarus, Minsk, Belarus; 3) Sechenov First Moscow State Medical University of the Ministry of Health of the Russian Federation; 4) Federal Center for Cardiovascular Surgery, Astrakhan, Russian Federation; 5) Mayo Clinic, Rochester, MN, USA; 6) City Clinical Hospital No 17, Moscow, Russian Federation; 7) Tomsk National Research Medical Center of the Russian Academy of Sciences, Tomsk, Russian Federation; 8) Sydney Medical School, The University of Sydney, Australia; 9) Pirogov Russian National Research Medical University, Moscow, Russian Federation.

Introduction. Hypertrophic cardiomyopathy is a genetic-based disorder, characterized by thickening of left ventricular (LV) and interventricular septum, which can't be explained by other cardiac or systemic disease. Mutations in over a dozen genes are found to cause HCM. The clinical spectrum and phenotype-genotype correlations have not been fully investigated. The aim of our study is to characterize a natural history of the HCM caused by ancestral mutation p.Q1233* in the MYBPC3. **Methods.** Patients with the p.Q1233* mutation in the MYBPC3 underwent detailed clinical evaluation. Clinical and genetic data were collected from centers in Russia, USA, Australia, and Belarus. Clinical information was collected from 51 mutation carriers (38 probands and 13 relatives; M:F ratio 1:1) aged 1 to 71. Genetic screening included NGS of the 78 genes (25 probands), and Sanger sequencing of the MYBPC3 gene (12 probands). **Results.** Mutation p.Q1233* was co-inherited with the polymorphism p.R326Q in all the cases. Co-segregation of these two variants and the same haplotype were confirmed by cascade familial screening in familial cases. Twenty-six probands (68%) had self-reported Slavonic origin. This mutation was found in 11% of HCM patients of Slavonic origin in the HCM cohort tested in Petrovsky Russian Research Centre of Surgery (Moscow, Russia). The majority of mutation carriers manifested with cardiac hypertrophy. One proband (47 y.o., female) had LVNC and severe arrhythmias without obvious myocardial hypertrophy, she also carried p.R237W variant in MYH7 previously assigned to be causative for DCM. The remaining 50 mutation carriers were diagnosed with HCM, mainly with sigmoidal hypertrophy of interventricular septum. Five out of 50 (10%) met LVNC criteria. Mean age at the diagnosis was 41 years, and the mean age of manifestation (or typical complains) was 34 y.o. Seventeen out of 50 carriers had a LVOT obstruction, and probability to have an LVOT was increasing with age. Ten patients underwent surgical treatment (extended myectomy or alcohol ablation), and 1 proband was transplanted. Ventricular arrhythmias were noted in 36% of carriers, and SCD occurred in 7 families, at a mean age of SCD of 51 years. **Conclusion.** We suggest the mutation p.Q1233* had risen on the background of p.R326Q-allele. This mutation is the most common in HCM patients of Slavonic origin, and is highly penetrant. Mutation carriers have a high probability to meet criteria for early surgical treatment.

2844W

Targeted gene pannel screening in sudden cardiac death: Challenges and utilities, to be or not to be? S.G. Temel¹, A. Akcay², I. Deniz³, B. Turk-genc⁴, M.C. Ergoren⁵, C. Yakicier⁶. 1) Uludag University, Faculty of Medicine, Department of Medical Genetics, Bursa, Turkey; 2) Ministry of Justice, Council of Forensic Medicine, Mortuary Department, Istanbul, Turkey; 3) Nicosia Nalbantoglu Government Hospital, Department of Forensic Medicine, Nicosia, Cyprus; 4) Acubadem University Hospital, Genetic Diagnosis Centre, Istanbul, Turkey; 5) Near East University, Faculty of Medicine, Department of Medical Biology, Nicosia, Cyprus; 6) Acibadem University, Faculty of Sciences & Arts Department of Molecular Biology & Genetics, Istanbul, Turkey.

Sudden cardiac death (SCD) is one of the major causes of mortality worldwide and is responsible for a large proportion of sudden deaths in young individuals. In forensic medicine up to 30% of all sudden death cases remain unexplained after routine postmortem autopsy, with no definite cardiac etiology identified after gross and microscopic inspection of the heart. Two groups of inherited heart diseases are responsible for SCD: cardiomyopathies and channelopathies. Our research group focus on cardiac channelopathies, which are characterized by lethal arrhythmias in the structurally normal heart, incomplete penetrance, and variable expressivity. Arrhythmias in these disease group result from pathogenic variants in genes encoding cardiac ion channels or associated proteins. Due to a lack of structural changes in the heart; it can not be detectable during conventional forensic autopsy investigation, channelopathies are often considered as potential causes of death in unexplained forensic autopsies. The asymptomatic nature of channelopathies is cause for concern in family members who may be carrier, making the identification of these genetic factors of significant clinical importance. Targeted sequencing was performed to the forensic SUD cases and family members. Postmortem genetic testing, challenges and utilities of the genetic testing will be discussed with unique SCD cases in our cohort. Molecular autopsy, post-mortem genetic testing and detailed premorbid clinical and family history can survive family members of SCD cases. .

2845T

Molecular dissection of the dilated cardiomyopathy in pediatric patients in Russia. E. Zaklyazminskaya^{1,2}, A. Bukaeva¹, M. Polyai¹, A. Shestak¹, V. Mikhailov¹, O. Blagova³, Yu. Frolova¹, I. Povolotskaya⁴, N. Kotlukova², S. Dzemeshevich¹. 1) , Russian Federation; 2) Pirogov Russian Research Medical University, Moscow, Russian Federation; 3) Sechenov State Medical University, Moscow, Russian Federation; 4) Centre of Genetics and Reproductive Medicine GENETICO, Moscow, Russian Federation.

Aim: To determine mutation spectrum in patients with early manifestation of the dilated cardiomyopathy (DCM), and to propose optimal way of DNA diagnostics in pediatric DCM cohort. **Materials and Methods.** Complete clinical and instrumental examination of the 35 pediatric DCM patients with DCM was performed in expert cardiology departments. Molecular genetic study had included the semiconductor sequencing based on the IonTorrent platform of the coding and adjacent regulatory areas of the 90 genes (for 25 isolated cases); and whole exome sequencing *trios* on the Illumina platform (for 10 family cases). **Results and discussion.** The group included 35 probands diagnosed with DCM before the age of 17, the average age of diagnosis was 6.5 years old. The gender ratio (M: F) was 17: 7. Sporadic cases of DCM were 27, familial cases - 8 (when the diagnosis of DCM was found in at least one parent and/or sibling). Pathogenic and likely pathogenic genetic variants were found in 23 probands (66%). *De novo* origin of mutation was confirmed in 7 probands (20%). Complex genotype (one homozygous mutation or heterozygous compounds were found in 6 probands (17%). Surprisingly, no mutation leading to the premature stop codon, or any other potentially pathogenic substitutions were found in the *TTN* gene. Half of detected mutations were found in the genes encoding sarcomeric proteins. **Conclusion.** The combination of the large target gene panel and whole exome sequencing allows to identify genetic cause of DCM in >60% probands manifested before 18 y.o.. *De novo* mutations and "double-hit" cases accounts significant portion of patients. The results of this study do not allow to consider the *TTN* gene as the first line of DNA diagnostics for DCM in the pediatric group, despite the publication of a high incidence of mutations in this gene in whole DCM cohort. The optimal strategy of the DNA diagnostics is the whole-exome sequencing.

2846F

Validation of targeted next-generation sequencing panel for thoracic aortic aneurysms and aortic dissections. C. Seo^{1,2}, B. Lee^{2,3}, H. Yoo^{2,3}, S. Cho², C. Park¹, E. Seo^{1,3}. 1) Laboratory Medicine, University of Ulsan College of Medicine and Asan Medical Center, Seoul, South Korea; 2) Pediatrics, University of Ulsan College of Medicine and Asan Medical Center, Seoul, South Korea; 3) Medical Genetics Center, University of Ulsan College of Medicine and Asan Medical Center, Seoul, South Korea.

Background: Thoracic aortic aneurysms and aortic dissections (TAAD) are associated with genetic disorders such as Marfan syndrome, Loeys-Dietz syndrome, and other heritable thoracic aortic diseases. We developed and validated amplicon-based targeted next-generation sequencing (NGS) panel for TAAD. **Methods:** The NGS panel comprised 16 genes including ACTA2, COL3A1, FBN1, MAT2A, MYH11, MYLK, NOTCH1, PRKG1, SKI, SLC2A10, SMAD3, SMAD4, TGFB2, TGFB3, TGFB1, and TGFB2 (518 amplicons, 95kb). We used NA12878 as a reference material (Coriell, Camden, NJ, USA) and positive control samples from 5 patients with known variants as follows: FBN1:c.5683T>C, FBN1:c.8165del, FBN1:c.4684_4697del, FBN1:c.4839_4840dup, and TGFB2:c.1609C>T. The primary NGS platform was Ion OneTouch and Ion PGM Dx (Thermo Fisher Scientific, Waltham, MA, USA). And we re-validated the panel when the platform was upgraded to Ion Chef and Ion S5 XL (Thermo Fisher Scientific). Bioinformatics tools were Torrent Suite 5.6 and Ion Reporter 5.6 (Thermo Fisher Scientific) and Biomedical Genomics Workbench (BxWB) 5.0.1 (Qiagen, Hilden, Germany). We compared quality control (QC) metrics between Ion PGM Dx and Ion S5 XL. The performance evaluation included accuracy, precision, analytical sensitivity, specificity, reportable range, and reference range. We also performed this panel testing for 8 TAAD patients who was negative for FBN1 Sanger sequencing. **Results:** Ion S5 XL showed better QC metrics compared to Ion PGM Dx, especially for mean read length (208 vs. 169bp), and percent end-to-end reads (89.5 vs. 16.2%). The performance evaluation using NA12878 showed higher accuracy, precision, analytical sensitivity, and specificity in Ion S5 XL compared to Ion PGM Dx. BxWB was complementary to Torrent Suite and Ion Reporter for variant analysis and interpretation. From the test for 8 TAAD patients who was negative for FBN1 sequencing, 4 significant variants were identified: TGFB2:c.1042C>T (LP), TGFB2:c.1322G>A (VUS), FBN1:c.5788+5G>A (P), NOTCH1:c.6557G>T (VUS). **Conclusions:** Ion S5 XL showed better QC metrics and performance characteristics compared to Ion PGM Dx. Method-based validation with reference material provided useful and accurate performance evaluation. It is recommended to use more than one bioinformatics tools for complementary strategy. This panel is expected to provide efficiency and high diagnostic yield for TAAD patients.

2847W

Identification of mosaic mutations in two Chinese families with MFS/LDS.

Y.Guo, Zhu¹, H. Yang¹, Z.Wei Zhou¹, Y.Ming Luo², H.Yin Zhang¹, Z. Zhou¹. 1) State Key Laboratory of Cardiovascular Disease, Beijing Key Laboratory for Molec, Beijing, China; 2) State Key Laboratory of Cardiovascular Disease, Center of Vascular Surgery, Fuwai Hospital, National Center for Cardiovascular Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, 100037, China.

Background: Marfan syndrome (MFS) and Loeys-Dietz syndrome (LDS), are two autosomal dominant connective tissue disorders, with many overlapping phenotypic characteristics. Considering their fully penetrance at an early stage, the variant observed in a healthy parent might be downgraded to VUS (variant of uncertain significance) or likely benign/benign, as it can be considered as strong evidence for a benign interpretation. However in that case, mosaicism must be taken into account. **Methods:** A total of 563 patients with suspected MFS or early onset TAAD were referred to our center to take a gene panel testing, involving 15 genes associated with Marfan and its related aortic diseases, including *FBN1*, *TGFBR2* and so on. Then the suspected mutations were verified by Sanger sequencing. Thereinto, 2 rare mutations were identified, which were inherited from their healthy parents, but Sanger sequencing results displayed an unbalanced double peak in the mutation site, suggesting mosaicism might exist. To confirm the mosaicism, a deep sequencing (5000×) on the single amplicon was performed on the DNA sample extracted from whole blood on Life PGM™ instrument. **Results:** Two mosaic mutations were discovered and confirmed in our study. The first mutation *TGFBR2*, R356P was identified in patient AD257, who was a 8 year-old boy presented with aortic pseudo-aneurysm, tortuous bilateral carotid artery, cleft palate, hypertelorism, which strongly suggested LDS. However, his father harbored the same mutation but appeared quite normal except that cardiac ultrasonography revealed reduced left ventricular diastolic function. Finally, the overall quantity 25% of mutated alleles in the father's blood was determined through deep sequencing. The second mutation *FBN1*, C1305F was detected in patient AD700-1, who was a 7 year-old boy with typical MFS features. His father was 30 years old, tall (190cm/100Kg) and myopia, with no other abnormality. After deep sequencing, 10% mutated alleles were detected from the father's blood and confirmed the mosaicism. According to 2015 ACMG criterion, these two variants were both classified as pathogenic. **Conclusions:** Mosaicism could be a more common phenomenon in MFS and its related diseases than it has been realized so far. We should take it into account when no mutation is found in a classic MFS patient, and when a variant is also observed in the atypical/healthy parent. It might help to improve the diagnostic rate.

2848T

The importance of gene-based knowledge in variant classification: An analysis of *FBN1* missense variant classifications in ClinVar. L.M. Baudhuin, K.E. Kotzer, M.L. Kluge, S.A. Lagerstedt. Mayo Clinic, Rochester, MN.

Background: Genetic variant classification for inherited disorders can be enhanced with gene-specific knowledge. In order to better understand the classification of critical and non-critical region-specific variants by the clinical laboratory community, we analyzed *FBN1* missense variants submitted to the ClinVar database. **Methods:** ClinVar *FBN1* variant data was downloaded and relevant missense variants were binned into 11 different categories based on whether they impacted a previously defined critical residue. ClinVar submitter variant classification and evidence (when available) were analyzed utilizing ACMG 2015 criteria. **Results:** A total of 1405 unique *FBN1* variants with 1963 classifications from 18 different submitters were downloaded. There were 355 unique missense variants (with 418 classifications) that occurred at critical amino acid residues. Overall, 82.3% of critical residue missense variants were classified as likely pathogenic or pathogenic (LP/P). There was a lower percent of LP/P classification for variants that did not impact a cysteine residue [60.1% (79/130)] compared to cysteine variants [92.3% (275/287)]. There were 285 unique missense variants (with 513 classifications) that did not occur at known critical residues. Of these, LP/P and uncertain significance (VUS) variant calls that included evidence were likely overclassified in 56% (24/43) and 30.1% (63/209) of cases, respectively. **Conclusions:** *FBN1* missense variants that occur at critical residues that do not impact cysteine residues or do not occur at known critical residues have a higher likelihood of being misclassified. These results underscore the importance of relying on both gene-based knowledge and applying appropriate ACMG criteria for variant classification, as well as updating classifications as more evidence becomes available.

2849F

Molecular spectrum of PAH gene mutations in the Aegean region of Turkey: Identification of four novel mutations. F. Ozkinay^{1,4}, M. Çoker², T. Atik¹, E. Isik¹, E. Er², M. Kose³, H. Onay⁴. 1) Pediatric Genetics, Ege University, School of Medicine, Izmir, Turkey; 2) Pediatric Metabolism, Ege University, School of Medicine, Izmir, Turkey; 3) Pediatric Metabolism, Dr. Behcet Uz Training and Research Hospital, Izmir, Turkey; 4) Medical Genetics, Ege University, School of Medicine, Izmir, Turkey.

Phenylketonuria (PKU) is a common autosomal recessive disease caused by mutations in the Phenylalanine Hydroxylase (PAH) gene located at chromosome 12q22-q24.2. The incidence of phenylketonuria is respectively high in Turkey. Mutation spectrum of PAH gene differs among race and ethnicity. The aim of this study is to investigate the molecular spectrum of PAH mutations in the population from Aegean region of Turkey. In the study 124 unrelated patients with PKU and 23 unrelated carriers referred to a reference center for molecular analysis over a 2 year period (from 2016 to 2017) were included. The majority of patients were diagnosed during newborn screening. All 13 coding exons and exon-intron boundaries of the PAH gene were sequenced using next-generation sequencing. Sixty different mutations including 4 novel mutations were detected in 271 phenylketonuria alleles. IVS6+4A/T, K396R, F149S, F55LfsX6 are the novel mutations that have not been reported in the literature. The four most common mutations were IVS10-11G>A (12.4%), A403V (9.56%), A300S (9.22%) and R261Q (9.22%). The present study showed a marked genetic heterogeneity in hyperphenylalaninemia patients from Aegean Region of Turkey. Similar to the most of other Mediterranean populations and populations from other regions of Turkey, IVS10-11G>A is the most frequent mutation in this region. Results of this study will serve as valuable tools for genetic diagnosis and appropriate genetic counseling in the related area.

2850W

IGF1R variants in patients with growth impairment: Four novel variants and genotype-phenotype correlations. L. Yang, F. Luo, W. Zhou. Children's hospital of Fudan university, shanghai, China.

Objective: IGF1R gene mutations have been associated with varying degrees of intrauterine and postnatal growth retardation, as well as microcephaly. Both autosomal dominant and autosomal recessive inheritance patterns have been reported. This study aimed to analyze the IGF1R gene in children with growth impairment using whole exome sequencing (WES) and to assess the clinical features with the autosomal dominant and recessive models. **Methods:** We performed WES in 28 unrelated patients and found 4 children harboring IGF1R gene variants. We compared the clinical findings in our cases carrying IGF1R mutations to those in patients reported in the HGMD database. **Results:** We identified four novel IGF1R gene variations by WES in 3 unrelated patients, including one missense variant (c.3740T>C (p.M1247T)) (Patient 1) inherited from an affected mother, one *de novo* nonsense variant (c.2629C>T, (p.R877X)) (Patient 2), and two compound heterozygous variations (c.2305G>C (p.E769Q); c.2684G>A (p.R895Q)) (Patient 3). To date, twenty-two patients have been described as harboring pathogenic variations in IGF1R in the HGMD database. We found that patients with compound heterozygous or homozygous variations displayed more severe phenotypes that were mainly characterized by developmental and speech delays, as well as mental retardation. **Conclusion:** We identified four novel pathogenic variations in the IGF1R gene, which expanded the known mutation spectrum. Through a comparison among patients with reported IGF1R pathogenic variations, this study determined that an autosomal recessive inheritance model of the IGF1R gene results in a more severe phenotype with developmental and speech delays, as well as mental retardation.

2851T

Screening for mutations in 14 kinds of MODY genes in patients with MODY in Japanese by next generation sequencing. *N. Iwasaki*^{1,2,3,4}, *T. Yamamoto*^{2,4}, *H. Akagawa*², *M. Ogata*^{3,5}, *M. Tsuchiya*³, *K. Saito*⁴. 1) Institute of Geriatrics, Tokyo Women's Medical University, Tokyo, Japan; 2) Institute of Integrated Medical Science, Tokyo Women's Medical University, Tokyo, Japan; 3) Diabetes Center, Tokyo Women's Medical University, Tokyo, Japan; 4) Institute of Medical Genetics, Tokyo Women's Medical University, Tokyo, Japan; 5) Department of Nutrition, Shukutoku University, Chiba, Japan.

[Aim]; Maturity onset diabetes of the young (MODY) is a subtype of diabetes mellitus characterized by a single gene disorder and a part of MODY (MODY1, 2 and 3) are known to be good targets for personalised medicine. In UK, it is reported that MODY2 and MODY3 comprise 80% of whole MODY. However, the responsible genes in MODY are not fully elucidated in Japan. The aim of the study was to screen mutations in 14 kinds of MODY genes by comprehensive methods using next generation sequencing in Japanese subjects with MODY. [Methods]; The subjects were clinically diagnosed with MODY and unidentified the causative mutations to date. Inclusion criteria were those diagnosed with diabetes before the age of 30 years old, negative for GAD antibody if tested, and BMI <30. Family history was not considered. Genotyping was performed by using HaloPlex SureDesign system (Agilent Technologies, Santa Clara, CA) that covered whole coding regions in 14 MODY genes and sequenced on an Illumina MiSeq v3(600) (Illumina, San Diego, CA). Identified variants were annotated with 1000 genome, CADD, ClinVer and gnomAD browser. This study was performed under the permission of Ethical Board for the human Genome study in the University. [Results]; Forty eight patients (M/F: 18/30; age at diagnosis: 16.5±3.9 years, BMI at first visit: 21.5±2.8 m²/kg (m±SD) were analyzed. We have found a total of 26 kinds of nonsense or missense variants in 34 subjects. Those variants were found in 10 MODY genes including *HNF4A*, *GCK*, *HNF1A*, *PDX1*, *CEL*, *PAX4*, *INS*, *ABCC8*, *KCNJ11* and *RFX6* genes. Two nonsense mutations, C370X in *GCK*/MODY2, and W324X in *HNF1A*/MODY1 were identified. The following 4 variants were seemed to be likely pathogenic (R148G in *PDX1*/MODY4, R121W in *PAX4*/MODY9, G826V in *ABCC8*/MODY12 and E140K in *KCNJ11*/MODY13). R121W in *PAX4*/MODY9 was found in 5 subjects (10.1%), and remaining was one each. Interestingly, two subjects carry dual mutations; one individual carried R148G in *PDX1* and R121W in *PAX4*, and the other proband carried C370X in *GCK* and R121W in *PAX4*. [Conclusion]; Overall, 8 in 48 subjects (16.7%) were found to have pathogenic or likely-pathogenic mutations in known MODY genes, including three unreported mutations in Japanese subjects with MODY.

2852F

Personalized medicine in diabetes: Unraveling the disease casualty of *HNF1B* gene coding variants of “unknown clinical significance” in two large Norwegian diabetes registries. *B.B. Johansson*¹, *K.A. Mohamed*¹, *M. Ringdal*¹, *A. Kaci*¹, *L. Bjørkhaug*², *P.R. Njølstad*^{1,4}, *J. Molnes*^{1,3}. 1) KG Jebsen Center for Diabetes Research, Department of Clinical Science, University of Bergen, Norway; 2) Department of Biomedical Laboratory Sciences and Chemical Engineering, Western Norway University of Applied Sciences, Bergen, Norway; 3) Department of Medical Genetics, Haukeland University Hospital, Bergen, Norway; 4) Department of Pediatrics and Adolescents, Haukeland University Hospital, Bergen, Norway.

Background/aims: Genetic variants in *HNF1B* encoding the transcription factor hepatocyte nuclear factor-1 beta (HNF-1B) can cause Maturity-Onset Diabetes of the Young type 5 (MODY5; HNF1B-MODY) but is also the most common known monogenic cause of developmental kidney disease. The aim of this study was to investigate possible pathogenic effects of 17 *HNF1B* variants identified in the Norwegian Childhood Diabetes Registry (NCDR) and the Norwegian MODY Registry, using functional protein analyses, and to correlate findings with clinical characteristics of *HNF1B* variant carriers. Methods: *HNF1B* variants were classified using a five-tier score system commonly used in clinical diagnostic laboratories. To investigate the effect of HNF1B variants on normal HNF-1B transcriptional activity, we used a Dual-Luciferase assay system in transfected HeLa cells. Results: 17 *HNF1B* variants were identified in patients of the NCDR (*n*= 5093) and in probands of the Norwegian MODY Registry (*n*=1145). 7 variants were classified as likely pathogenic or pathogenic, 6 variants of unknown significance and 4 variants were interpreted as benign or likely benign. All rare variants were further investigated by functional studies. 8 patients with *HNF1B* variants presented with renal disease, and the commonest renal manifestation in these patients were renal cysts which were present in 5/16 patients. 3/16 patients presented with renal disease without diabetes. Diabetes was present in 12/16, whereof 7 probands presented with early-onset diabetes without reported kidney disease. Extra renal features seen in the patients, in addition to diabetes, were pancreas hypoplasia or atrophy, exocrine deficiency and liver dysfunction. Only 8/16 families met the minimal MODY criteria (families with at least two generations of diabetes with at least one subject diagnosed under the age of 25 years). The effect of the observed *HNF1B* variants on transcriptional activity is now being investigated in HeLa cells. Preliminary results shows reduced transactivation activity in the two variants p.Arg137_Lys161del and a positive control p.Arg177Ter, which is also supported by others. This also coincides with the available clinical data on the patients. Conclusion: Investigation of *HNF1B* variant functional effects should support precision medicine in diabetes clinics and could help to distinguish neutral variants from variants causing HNF1B-disease, in combination with family history and clinical characteristics.

2853W

Prevalence of significant genetic variants in glycogen storage disease via custom NGS panel in a single center hospital based study. A. Bisgin^{1,2}, I. Boga², H.N. Onenli Mungan³, D. Kor², S. Tug Bozdogan^{1,2}, O. Sonmezler².

1) Cukurova University, Faculty of Medicine, Medical Genetics Dept of Balcali Hospital and Clinics, Adana, Turkey; 2) Cukurova University, AGENTEM (Adana Genetic Diseases Diagnosis and Treatment Center), Adana, Turkey; 3) Cukurova University, Faculty of Medicine, Pediatric Metabolism-Endocrinology Dept. of Balcali Hospital and Clinics, Adana, Turkey.

Objective: Glycogen storage diseases (GSDs) are mostly transmitted as autosomal recessive disorders and are therefore more frequent in countries with high consanguinity rates such as in Turkey. This study reports our local experience in the genetic diagnosis of inborn errors of metabolism over the GSDs.

Methods: 576 patients recruited for this study were all clinically suspected GSDs and admitted to our genetic diseases diagnosis and treatment center (CU AGENTEM). We have customized a GSD related multi-gene panel including 19 genes (*GAA*, *SLC37A4*, *G6PC*, *GYS2*, *AGL*, *GBE1*, *PHKA2*, *PHKB*, *PHKG2*, *PHKA1*, *PYGM*, *PYGL*, *PFKM*, *PGAM2*, *LDHA*, *ENO3*, *PGM1*, *ALDOA* and *GYG1*) for next generation sequencing (MiSeq, Illumina). **Results:** A significantly high consanguinity rate was observed among parents of affected children. 181 clinically significant variants were reported in 170 (29.5%) out of 576 patients. 69 (40.6%) of them had novel variants which was confirmed by parental testing for the carrier status. The most frequent cases were *AGL* gene related GSD type III, GSD type I with mutations in *G6PC* and *SLC37A4* genes, *PYGM* gene related GSD type V and GSD type IX with mutations in *PHKA2*, *PHKG2* and *PHKB* genes. **Conclusion:** Inborn errors of metabolism are common and it is important to enhance awareness among clinicians and families. Although our results are preliminary still suggests the idea of how it is necessary to do molecular testing fast and accurately for patient management, and genetic counselling.

2854T

Common mutation and hot-spot region of *ATP7B* gene in Vietnamese children with Wilson disease. H.M.T. Nguyen¹, H.A.P. Nguyen², M.P.T. Nguyen¹, N.D. Ngo¹, T.M. Ngo¹, N.T. Ngo¹, D.M. Tran^{1,2}, H.T. Le^{1,2}. 1) Human genetics Dept, Vietnam National Children's Hospital, Hanoi, Hanoi, Viet Nam, MsC; 2) Hepatology Dept, Vietnam National Children's Hospital, Hanoi, Hanoi, Viet Nam, MD.

Introduction The WD gene—*ATP7B*, codes for a copper transporting P-type ATPase, which is expressed mainly in hepatocytes and functions in the transmembrane transport of copper. It is a rare autosomal recessive disorder of copper metabolism, which is characterized by hepatic and neurological disease. The mutation spectrum of *ATP7B* gene is dissimilar between ethnic groups and appears to be population-specific. p.R778L is very common in Asian but p.H1069Q is the most frequent mutation in European. **Aim of study** The goal of this study is to investigate characteristics of *ATP7B* mutation in children with Wilson disease. **Patient and method** Peripheral blood was collected from unrelated 78 WD patients—46 females and 32 males. Direct DNA sequencing for 21 exons and intron boundary was used to identify the mutation in *ATP7B* gene. **Result** A total of 27 different mutations (missense mutations, n= 16; frame shift, n=6; nonsense mutations, n=1; splice site mutations, n=4) were detected in this study, which accounted for 96.8%. Of the 78 Wilson patients, *ATP7B* mutations were identified in both alleles in 38 patients (homozygous; n=26, compound heterozygote; n=49); in only one allele in three patients (n=1); and none of mutation in both alleles in two patients (n=2). Among 47 genotypes, ratio of homozygote and compound heterozygote are 62.8% and 33.3%. p.S105X often found in homozygous patients and present in 37.1% of the 156 alleles studied. Following is five other mutations, including: p.I1148T (7.3%), IVS14-2A>G (6.6%), p.L1371P (6.0%), p.T850I và p.V176SfsX28 (5.3%). Most of mutations in this study occurred in exon 2 (43.0%), exon 16 (9.9%), exon 8 (8.6%), exon 14 and intron 14 (6.6%). Procedure to detect *ATP7B* gene mutation were set up as following: (1) exon 2; (2) 6 exon (8, 16, 14, 15, 18, 20); (3) 5 exon (6, 10-13); (4) the remainings in *ATP7B* gene. **Conclusion** S105X is the most prevalent of Wilson mutation in this study. Exon 2, exon 8, exon 16, exon 14 and intron 14 are considered as hotspot regions in children with Wilson disease in Vietnam. The discovery of the characteristic of *ATP7B* gene in Wilson patients has opened up a new approach in diagnosis for this disease in Vietnam. .

2855F

Genetic testing for patients with exercise intolerance, rhabdomyolysis, or a suspected metabolic myopathy. M. Kyriass, A. Gruber, F. Xu, Q. Hurshman, J. Weber. PreventionGenetics, Marshfield, WI.

The metabolic myopathies are a heterogeneous group of disorders caused by defects in muscle energy metabolism that cause inadequate production or maintenance of ATP levels (Kahler 2017. Metabolic Myopathies. In: Inherited Metabolic Diseases: A Clinical Approach). Typical clinical features include slowly progressive muscle weakness with or without hypotonia, premature fatigue, episodic aches, cramps and myalgia, exercise intolerance with or without a "second wind" phenomenon, rhabdomyolysis which may or may not be accompanied by myoglobinuria, high creatine kinase (CK) levels, and acute renal failure (Wortmann 2002. Reumatologia. 18: 90-93; Olpin et al. 2015. PubMed ID: 25878327; Kahler 2017). Early recognition of these disorders is important as proper intervention can help prevent morbidity and/or mortality caused by the more severe symptoms (generally rhabdomyolysis and renal failure) (Olpin et al. 2015. PubMed ID: 25878327). In an effort to provide a molecular diagnosis for these patients, PreventionGenetics developed a next-generation sequencing (NGS) panel containing 59 genes associated with known metabolic myopathies or disorders that include exercise intolerance and/or rhabdomyolysis as a main clinical presentation. To date, a clear positive result was found in approximately 6% of cases. The most frequently affected gene was *PYGM*, which is known to be one of the most common causes of metabolic myopathy (Olpin et al. 2015. PubMed ID: 25878327). We have also identified causative variants in genes more rarely associated with these disorders, such as *ANO5* and *HADHB*. In addition to the clearly positive cases, an additional 14% of cases were found to carry a single pathogenic or likely pathogenic variant, and nearly 70% of cases had variants of uncertain significance that could be related to the phenotype. In the majority of cases, deletion/duplication testing was not performed to look for possibly causative copy number variants (CNVs), which may have further clarified test results. In the near future, CNV analysis will be integrated with this test. Overall, these results indicate that NGS-based panel testing for metabolic myopathies, exercise intolerance and/or rhabdomyolysis can aid in the clinical diagnosis of this patient group.

2856W

Detection of GBA missense mutations and other variants using the Oxford Nanopore MinION. C. Proukakis¹, M. Leija-Salazar¹, K. Mokretar¹, S. Mullin¹, M. Toffoli¹, M. Athanasopoulou¹, A. Donald², R. Sharma³, D. Hughes⁴, A.H.V. Schapira¹, F.J. Sedlazeck⁵. 1) Institute of Neurology, University College London, London, England, United Kingdom; 2) Department of Paediatrics, Royal Manchester Children's Hospital, UK; 3) The Mark Holland Metabolic Unit, Salford Royal Foundation NHS Trust, Salford, UK; 4) Lysosomal Storage Disorders Unit, Royal Free Hospital, London, UK; 5) Human Genome Sequencing Center, Baylor College of Medicine, Houston, USA.

Mutations in *GBA* cause Gaucher disease when biallelic, and are strong risk factors for Parkinson's disease when heterozygous. *GBA* analysis is complicated by the nearby pseudogene. We designed and validated a method for sequencing *GBA* on the Oxford Nanopore MinION. We sequenced an 8.9 kb amplicon from DNA samples of 17 individuals, including patients with Parkinson's and Gaucher disease, on older and current (R9.4) flow cells. These included samples with known mutations, assessed in a blinded fashion on the R9.4 data. We used NanoOK for quality metrics, two different aligners (Graphmap and NGMLR), Nanopolish and Sniffles to call variants, and Whatshap for phasing. We detected all known mutations in these samples in a blinded fashion, including the common p.N409S (N370S) and p.L483P (L444P), and three rarer ones, at the correct zygosity, as well as intronic SNPs. In a sample with the complex RecNcil allele, we detected an additional coding mutation, and a 55-base pair deletion. We confirmed compound heterozygosity where relevant. False positives were easily identified. The Oxford Nanopore MinION can detect missense mutations and an exonic deletion in this difficult gene, with the added advantage of phasing and intronic analysis. It can be used as an efficient diagnostic tool.

2857T

Innate immune gene expression signature in rheumatoid arthritis revealed the key role of *TLR8* and *IL1RN* in discrimination of active and inactive disease. A. Petrackova¹, P. Horak², M. Radvansky², R. Fillerova¹, V. Smotkova-Kraiczova¹, M. Kudelka³, F. Mrazek¹, M. Skacelova², A. Smrzova², M. Schubertova², E. Kriegova¹. 1) Department of Immunology, Faculty of Medicine and Dentistry, Palacky University, Olomouc, Czech Republic; 2) Department of Internal Medicine III - Nephrology, Rheumatology and Endocrinology, Faculty of Medicine and Dentistry, Palacky University Olomouc and University Hospital Olomouc, Olomouc, Czech Republic; 3) Department of Computer Science, Faculty of Electrical Engineering and Computer Science, Technical University of Ostrava, Ostrava, Czech Republic.

Background: Innate immunity plays a critical role in the pathogenesis of rheumatoid arthritis (RA). Although many studies have focused on individual mediators of innate immunity, a comprehensive innate gene expression signature associated with RA and particularly with disease activity has not been established. **Methods:** The gene expression profile of 18 innate immunity genes, Toll-like receptors (*TLR*)1-10, and seven members of the interleukin (*IL*)1/*IL1R* family, together with *CXCL8/IL8*, were assessed in peripheral blood mononuclear cells (PBMC) from patients with RA (n=67) treated according to the standard protocol and age-/gender-matched healthy control subjects (n=55) using a high-throughput SmartChip Real-Time-qPCR system (WaferGen). The subgroups of patients were formed on the basis of the disease activity, with a *Disease Activity Score in 28 joints (DAS28)* of ≥ 3.2 being taken as the active disease (inactive RA, n=35; active RA, n=32). **Results:** Expression profiling revealed the deregulation of 13 innate immunity genes in patients treated for RA when compared to the healthy controls ($P_{\text{corr}} < 0.05$), with *TLR10* and *IL1RAP* being reported for the first time. In active RA, upregulated expression of *TLR2*, *TLR4*, *TLR6*, and *TLR8* and downregulation of *TLR10* ($P < 0.04$) was observed when compared to inactive RA. Applying artificial neural networks, two gene sets within an eight-activity gene expression signature (*TLR1*, *TLR2*, *TLR3*, *TLR7*, *TLR8*, *CXCL8/IL8*, *IL1RN*, *IL18R1*) were able to discriminate between patients with the active and inactive disease. Multivariate visual network analysis revealed *TLR8* and *IL1RN* as key genes whose expression characterizes active and inactive RA. **Conclusion:** The innate immune gene expression signature revealed the key role of *TLR8* and *IL1RN* in the discrimination of active and inactive RA. Our study further supports the important role of innate immunity in RA and highlights the heterogeneity of active RA. **Funding:** MZ CR VES15-28659A, IGA UP_2018_010, IGA UP_2018_016, MH CZ – DRO (FNOL, 00098892).

2858F

Comprehensive profiling of chemokine receptors in peripheral blood mononuclear cells obtained in patients with chronic lymphocytic leukemia. E. Kriegova¹, V. Smotkova Kraiczova¹, R. Fillerova¹, G. Manukyan², A. Petrackova¹, Z. Mikulkova¹, G. Gabcova¹, L. Kruzova², R. Urbanova², P. Turcsanyi², P. Ryznerova², T. Papajik². 1) Department of Immunology, Faculty of Medicine and Dentistry, Palacky University Olomouc and University Hospital Olomouc, Olomouc, Czech Republic; 2) Department of Hemato-Oncology, Faculty of Medicine and Dentistry, Palacky University and University Hospital, Olomouc, Czech Republic; 3) Laboratory of Molecular and Cellular Immunology, Institute of Molecular Biology NAS RA, Yerevan, Armenia.

Background: Chemokines and their receptors are involved in the regulation of cell recruitment, survival, proliferation, and trafficking, all these processes crucial in the pathogenesis of chronic lymphocytic leukemia (CLL). Comprehensive profiling of chemokine receptors in CLL and its subgroups according to prognostic relevance is missing. **Aim:** To characterize the chemokine expression pattern in CLL patients and subgroups according to prognosis and cytogenetic aberrations. **Patients/Methods:** We studied the gene expression pattern of 16 canonical and 4 atypical chemokine receptors in peripheral blood mononuclear cells (PBMC) of CLL patients (n=88) and healthy subjects (n=34) by using SmartChip quantitative RT-PCR (WaferGen Bio-systems). Given the dynamic variability of CD5 expression and its link to the proliferating pool of neoplastic cells, the expression of CXCR3, CXCR4, CXCR5, CXCR7, and CCR7 was confirmed by flow cytometry on CD5^{high} and CD5^{low} cell subpopulations. **Results:** Among deregulated receptors, 5 receptors (CCR7, CCR10, CXCR3, CXCR4, CXCR5) were up-regulated and 9 receptors (CCR2-CCR6, CCR8, CCR12, CXCR1, CXCR2) down-regulated in CLL; the expression of others did not differ between CLL and controls ($P > 0.05$). In patients with del(17p) associated with a poor prognosis, higher mRNA levels of CXCR6, CXCR7 and CCR10 was observed comparing to del(13q). On protein level, CD5^{high} cell subpopulation expressed higher levels of CXCR3 ($P < 0.001$), CXCR5 ($P = 0.007$), CCR10 ($P = 0.001$), whereas CD5^{low} cells expressed higher levels of CXCR4 receptor ($P < 0.001$) in CLL as a whole as well as in subgroups. Moreover, our study revealed distinct differences in CXCR3 expression on both CD5^{high} and CD5^{low} between IgHV^{mut} and IgHV^{unmut} subgroups.

Conclusion: Our results provide a complete picture of expression patterns of chemokine receptors in CLL. Further studies are needed to clarify how chemokine receptor network affects neoplastic development and progression. Grant support: MZ CR VES16-32339A, IGA_UP_2018_016, MH CZ – DRO (FNOL, 00098892).

2859W

Retrospective analysis of a patient cohort referred for carrier testing for Alpha Thalassemia. *P.H. Fernandes, A. Kruger, J.P. Tazelaar, J. Zdrodowski, C.A. Holland.* Progenity, Inc., Ann Arbor, MI.

Alpha thalassemia is an autosomal recessive disorder, affecting up to 5% of the world's population. It is most often caused by deletions involving one or both α -globin genes (*HBA1* and *HBA2*). The alpha thalassemia deletion detection test is intended for carrier testing in adults of reproductive age in accordance with recommendations by ACOG. We performed a retrospective analysis of 75,332 patients referred for alpha thalassemia carrier testing. Testing included Multiplex PCR-based analysis to identify 7 common α -globin deletions, including $-\alpha 3.7$, $-\alpha 4.2$, $--SEA$, $--FIL$, $--MED$, $-\alpha 20.5$, and $--THAI$. Of the patients tested, 5191 (6.9%) were positive for at least one α -globin deletion. Among those that were positive, 4805 (92.6%) were heterozygotes, and 373 (7.2%) were homozygotes, who all had the $-\alpha 3.7/-\alpha 3.7$ genotype. We also identified 13 (<1%) compound heterozygotes. These included 7 patients with the $-\alpha 3.7$ and $-\alpha 4.2$ deletions, and 6 patients, whose genotype resulted in defective production of 3 out of 4 α -globins (HbH disease). The highest positivity rate was seen within the African American/Black population (32.4%), followed by the populations of Others/Mixed (7.9%), Asian (7.3%), Native American (5.4%), Hispanic (4.6%), Ashkenazi/Jewish (4.6%), and Caucasian/White (1.6%). To our knowledge, this is the largest and most ethnically diverse population studied in relation to alpha thalassemia carrier status. Analysis of these common α -globin deletions was successful in identifying silent carriers of alpha thalassemia and those with alpha thalassemia trait as well as those with HbH disease in our patient cohort. In addition, positive carriers were identified at an unexpectedly high frequency in relatively understudied ethnic backgrounds, like Native American and Hispanic populations, giving further impetus to test patients of all ethnicities. This is especially important considering the increasing number of patients with unknown or mixed ancestry.

2860T

Detection of alpha-thalassemia common large deletions by next generation sequencing. *SH. Rosenthal¹, A. Gerasimova², R. Owen¹, J. Catanese¹, A. Buller-Burckle³, F. Lacbawan¹.* 1) Advanced Diagnostics – Genetics, Genomics and R&D, Quest Diagnostics, San Juan Capistrano, CA; 2) Bioinformatics R&D, Quest Diagnostics, San Juan Capistrano, CA; 3) Molecular Genetics, Quest Diagnostics, San Juan Capistrano, CA.

Alpha-thalassemia is the most common inherited disorder of hemoglobin synthesis, affecting approximately 5% of the world's population. Genomic deletions involving the alpha-globin gene cluster on chromosome 16p13.3 are the main molecular cause of the disease, accounting for approximately 90% of cases. Multiplex gap-PCR and multiplex ligation-dependent probe amplification (MLPA) are frequently used molecular methods for the detection of these deletions. However, those methods are not easily consolidated with newer analytical procedures used for carrier screening of inherited disorders. In this study, a next generation sequencing (NGS) method was developed to detect common alpha-thalassemia deletions including $-\alpha 3.7$, $-\alpha 4.2$, $-\alpha 20.5$, $--MED$, $--SEA$, $--FIL$, $--THAI$, as well as the Constant Spring point mutation. To enable efficient sequencing on the Illumina MiSeq platform, a conventional multiplex gap-PCR was performed, followed by nested PCR for each deletion. Using this method, 18 other common inherited diseases including beta thalassemia were consolidated in one assay to create an expanded NGS carrier screening panel. Validation of the consolidated panel using 159 samples showed 100% sensitivity and specificity for all variants tested. Subsequent testing of 3,752 clinical samples using the consolidated panel showed that 7.6% (286/3,752) of cases harbored one or more alpha-thalassemia gene deletions, including 237 cases of $\alpha\alpha/-\alpha 3.7$, 40 cases of $-\alpha 3.7/-\alpha 3.7$, 7 cases of $\alpha\alpha/-\alpha 4.2$, 1 case of $-\alpha 3.7/-\alpha 4.2$, and 1 case of $\alpha\alpha/--MED$. This demonstrates the implementation of an NGS method to screen common alpha-thalassemia variants as part of a larger comprehensive carrier screening panel.

2861F**Clonal cell proliferation in paroxysmal nocturnal hemoglobinuria:**

Evaluation of *PIGA* mutations and T-cell receptor clonality. M. Kim^{1,2}, J. Lee^{1,2}, J. Park^{1,2}, Y. Kim^{1,2}, K. Han^{1,2}, N.G. Chung³, B. Cho³, S.E. Lee⁴, J.W. Lee⁴.
 1) Department of Laboratory Medicine, College of Medicine, The Catholic University of Korea, Seoul, Korea, Republic of; 2) Catholic Genetic Laboratory Center, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Korea, Republic of; 3) Department of Pediatrics, College of Medicine, The Catholic University of Korea, Seoul, Korea, Republic of; 4) Division of Hematology, Department of Internal Medicine, Catholic Blood and Marrow Transplantation Center, College of Medicine, The Catholic University of Korea, Seoul, Korea, Republic of.

Background: Paroxysmal nocturnal hemoglobinuria (PNH) is an acquired pluripotent hematopoietic stem cell disorder associated with an increase in the number of glycosyl-phosphatidyl inositol (GPI)-deficient blood cells. We investigated the PNH clonal proliferation in each blood cell lineage by analyzing the *PIGA* gene mutation and T-cell receptor (TCR) clonality. **Methods:** Twenty-four PNH patients were enrolled. Flow cytometry was used to measure GPI-deficient fractions in each cell lineage. *PIGA* gene mutation was analyzed by Sanger sequencing in granulocytes and T lymphocytes. TCR clonality assay was carried out in isolated GPI-deficient T lymphocytes. **Results:** The GPI-deficient fraction was the highest in granulocytes followed by red blood cells and T lymphocytes. *PIGA* mutations were detected in both granulocytes and T lymphocytes of 19 patients (79.2%), and the mutant burden was higher in granulocytes. The GPI-deficient fractions of granulocytes and T lymphocytes were significantly correlated with *PIGA* mutant burden of the respective cell lineage. The GPI-AP deficient RBC fraction correlated with reticulocyte count. *PIGA* mutation was more frequently observed in patients with clonal rearrangements in TCR genes ($P = 0.015$). The *PIGA* mutant burden of T lymphocytes was higher in patients showing clonal *TCRB* rearrangement. **Conclusion:** This study demonstrated that PNH clone size varied according to blood cell lineage and the clonal cells may obtain proliferation potential or gain a survival advantage compared with normal cells. **Keywords:** Paroxysmal nocturnal hemoglobinuria; Clonal cell proliferation; *PIGA* mutation; T-cell receptor clonality.

2862W

Expanded genetic testing for primary immunodeficiencies: Findings from a 207-gene next-generation sequencing panel. B. Johnson, P. Nicolosi, C. Hartshorne, S. Martin, K. Ouyang, J. Holle. Invitae, San Francisco, CA.

Many primary immunodeficiencies (PIDs) share overlapping presentations, complicating clinical diagnosis. Expanded next-generation sequencing (NGS) panels are valuable in facilitating the diagnosis of patients with PIDs. We performed a retrospective analysis of the clinical utility of a 207-gene PID NGS panel used in a clinical diagnostic laboratory. From April to October 2017, 260 panels were ordered for patients with suspected or known PIDs. Seventy-four pathogenic or likely pathogenic (P/LP) variants were identified in 63 (24.2%) patients. Eight (10.8%) of the P/LP variants were copy number variations (6 deletions, 2 duplications). Fifty-two patients (20%) had 1 P/LP variant, and 11 patients (4.2%) had 2 P/LP variants. Of positive patients, 31.7% ($n = 20$) were heterozygous carriers for autosomal recessive conditions where a second variant was not identified. Twenty-two patients had P/LP heterozygous variants in genes with autosomal recessive and autosomal dominant inheritance patterns, in which the positive findings may or may not explain the patient's phenotype. For example, 13 variants in this category were heterozygous variants in *TNFRSF13B* (TACI). Genetic diagnoses were established or likely in 30% of patients with P/LP variants (7.3% of all patients). Five patients were heterozygous for a single P/LP variant and a variant of unknown significance in the same autosomal recessive gene. Four patients in whom a genetic diagnosis was determined were also heterozygous carriers for a second, unrelated condition. One patient was found to have two distinct genetic diagnoses. Variants of uncertain significance (VUS) were identified in most (94.2%) patients. The average turnaround time from test requisition to return of results was 18 days. In total, 63% percent of genetic diagnoses were for conditions that are treatable with hematopoietic cell transplantation. These results illustrate the utility of broad NGS panels for the diagnosis of patients with PIDs..

2863T

HLA typing using capture-based next-generation sequencing. S.K. Lai¹, N.C. Chen², C.S. Chen³, P.L. Chen^{1,3,4,5}. 1) Graduate Institute of Molecular Medicine, National Taiwan University College of Medicine, Taipei, Taiwan; 2) Graduate Institute of Electronics Engineering, National Taiwan University, Taipei, Taiwan; 3) Department of Medical Genetics, National Taiwan University Hospital, Taipei, Taiwan; 4) Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan; 5) Graduate Institute of Medical Genomics and Proteomics, National Taiwan University College of Medicine, Taipei, Taiwan.

Human leucocyte antigen (HLA) genes are critical in organ transplantation and immune-related diseases. Traditional HLA typing methods, either by sequence-based typing (SBT) or sequence-specific oligonucleotide (SSO), might report ambiguous typing results, which are caused by the phasing problem or the polymorphic sites outside the sequencing regions. Recent next-generation sequencing (NGS)-based methods rely on long-range PCR, which might suffer from allele drop-out. Using probe capture-based NGS may solve the above-mentioned tackles. In this study, we established a probe capture-based NGS method for HLA typing. Probe sets for Class-I HLA gene (HLA-A, B, C), Class-II HLA gene (DRB1, DPA1, DPB1, DQA1, DQB1, DRB3,4,5) and also MICA covering exon 2 to exon 4 were designed. Ten samples previously known to carry the most common HLA types in Taiwanese were used as the stage one validation for this method. The results were 100% concordant to those from the SBT method. We then proceeded to test the 85 samples from The International HLA Reference Standards from International Histocompatibility Working Group (IHWG). An average of 99% results were consistent with the results from IHWG. The remaining 1% results were inconsistent, including 2 DPA1 alleles, 9 DPB1 alleles, 4 DQA1 alleles, 4 DQB1 alleles and 1 DRB1 allele; we later proved that those results from IHWG were wrong. The ratio of two allele read depth in heterozygous loci were 40%-60% in average. Furthermore, our data showed that this method could resolve ambiguous HLA types due to phasing problems.

2864F

A Blau syndrome family: Mutated *NOD2* allele transmitted from the father with somatic and germline mosaicism. Y. Lin^{1,2}, C. Zeng¹, X. Chen¹, T. Li¹, H. Liu¹, H. Liu³, H. Hu², L. Liu¹. 1) Department of Genetics and Endocrinology, Guangzhou Women and Children's Medical Center, Guangzhou, Guangdong, China; 2) Guangzhou Institute of Pediatrics, Guangzhou Women and Children's Medical Center, Guangzhou, Guangdong, China; 3) Department of Radiology, Guangzhou Women and Children's Medical Center, Guangzhou, Guangdong, China.

Objective: Blau syndrome (BS) is a rare autosomal dominant autoinflammatory disorder caused by *NOD2* gene deficits with a representative feature of granulomatous arthritis, uveitis and dermatitis on the joints, eyes and skin, which is still not well-known. In this study, we sought to ascertain the intrinsic cause of the suspected family. **Methods:** Here we employed a whole exome sequencing (WES) in a familial aggregation of 2 polyarthritis and erythra affected individuals across 2 continuous generations, followed by a confirmatory Sanger sequencing. **Results:** WES revealed a "de novo" heterozygous pathogenic variant, c.1000C>T(p.R334W), in *NOD2* gene of the proband, which was not found in the symptomatic father. A confirmatory of Sanger sequencing showed that the father shared the identical mutation to the proband as a somatic and germline mosaicism. An urgent prenatal diagnosis was further performed for this family due to an approximately 1/5 recurrence risk recommended by the spermatoc mosaic status. **Conclusion:** This is the first case of described *NOD2* mosaicism in Chinese, the most extensive investigation of the mosaic status, and the first case to speculate the timing of mutagenesis in BS. Our study provides precise diagnosis and genetic counseling for the family, and alerts clinicians and geneticists to be cautious about the molecular findings to avoid omissions and prevent an intrafamilial recurrence of the identical genetic abnormalities.

2865W

Prenatal diagnosis of alpha thalassemia in Northern Vietnam. *N.T.T. Nhung¹, N.T.P. Mai¹, A.T. Lan¹, N.M. Tien¹, N.T.M. Huong¹, N.D. Ngoc¹, L.T. Thuy², L.T. Hai¹.* 1) National children's Hospital, Hanoi, VietNam; 2) Hanoi Hospital of Obstetrics and Gynecology, Vietnam.

Background: Alpha Thalassemia is the most common single-gene disorder in worldwide. There are two severe forms of death: HbH and Bart's Diseases. In this study, we conducted prenatal diagnosis for families with a history of HbH and HbBart's birth from January 2010 to April 2018. **Methods:** We analyzed 226 samples of amniotic fluid from pregnant women at high risk of HbH and HbBart's births. Their family members (pregnancies, their husbands and their children) were identified mutations of *hba* gene. C-ARMS PCR, Multiplex Gap PCR, sequencing and MLPA techniques were accurated for screening all mutations in *hba* gene. **Results:** 54/226 fetuses had no mutations in *hba* gene; 107/226 fetuses had heterozygous of mutations in *hba* gene, 10/226 fetuses had compound heterozygous of 02 mutations and 55/226 fetuses had homozygous α^0 -thal. **Conclusions:** DNA testing is a basis progress for the diagnosis and preventive treatment before birth. Prenatal diagnosis process is only done on the family who had known mutations. **Keywords:** Alpha Thalassemia, HbBart's, HbH, pregnancy, DNA testing, Vietnam.

2866T

The CFTR MASTR v2 assay combined with MASTR reporter analysis is a fast and precise research solution for SNV, indel, CNV and (TG)m(T)n locus detection. *A. Rothier¹, K. Bettens¹, B. Verhelst², C. Marques¹, R. De Smet², B. Slabbinck², J. Crappé², J. Del Favero¹.* 1) R&D, Agilent Technologies Belgium, Niel, Antwerp, Belgium; 2) R&D, Agilent Technologies Belgium, Ghent, Belgium.

Introduction: NGS based complete *CFTR* mutation analysis is the method of choice to diagnose Cystic Fibrosis in patients, neonates and to perform carrier screening. To further improve the speed and accuracy of mutation analysis the CFTR MASTR Dx* assay was extended to a complete workflow comprising of a sample to result solution for all coding *CFTR* regions, 7 selected deep intronic regions, the (TG)m(T)n heteropolymeric region, and promoter region up to position c.-108. Data quality control to measure, monitor and optimize the complete workflow and data analysis is integrated within the MASTR Reporter. In addition, the MASTR Reporter includes a customizable variant filter for selection of the most relevant variants for any population under study. The CFTR NGS workflow was extensively examined in-house using both blood and dry blood spot (DBS) derived DNA. **Methods:** The CFTR MASTR v2 with MASTR Reporter (v1.1.0) CFTR v2** was verified on 57 blood samples (for 552 SNVs, InDels and (TG)m(T)n reference variants) and on 195 blood samples (for 15 CNVs). The MASTR Reporter CFTR v2** achieved 100% sensitivity and specificity for SNVs, InDels and (TG)m(T)n events in blood. The (TG)m(T)n was correctly called by MASTR Reporter in all samples. For CNV, the MASTR Reporter CFTR v2** showed 100% sensitivity and 98.45% specificity. In addition, CFTR MASTR Dx with MASTR Reporter CFTR Dx* was validated on 50 DBS (64 SNVs and InDels) and on 116 blood samples (431 SNVs, InDels and (TG)m(T)n). Sensitivity, specificity and accuracy* were all 100% for both SNVs on DBS as for (TG)m(T)n repeats, SNVs and Indels in blood. Noteworthy, MASTR Reporter identified the combined *CFTR* heterozygous I507del and F508del mutations. **Conclusion:** The CFTR MASTR v2 assay combined with MASTR Reporter** results in a fast and precise research solution for identification of SNVs, InDels, CNVs and (TG)m(T)n. The development of an advanced PCR stutter correction algorithm in MASTR Reporter (v1.1.0) also represents a major advantage for quality control and accurate (TG)m(T)n calling. Since the CFTR MASTR Dx with MASTR Reporter* validation study called SNVs (on DBS), and SNVs, Indels and (TG)m(T)n (on blood) with 100% accuracy, the complete workflow allows for routine diagnostic testing for all *CFTR* mutations. *The MASTR CFTR Dx workflow is available as CE-IVD product in EU only. **For Research Use Only. Not for use in diagnostic procedures.

2867F

Differential cytokine pattern associated with a osteolysis predisposing allele TNF-238*A in patients with/without severe osteolysis around total hip arthroplasty. R. Fillerova¹, T. Dyskova¹, P. Schneiderova¹, V. Smotkova-Kraiczova¹, J. Gallo², E. Kriegova¹. 1) Department of Immunology, Faculty of Medicine and Dentistry, Palacky University and University Hospital, Olomouc, Czech Republic; 2) Department of Orthopedy, Faculty of Medicine and Dentistry, Palacky University and University Hospital, Olomouc, Czech Republic.

Introduction: Individual susceptibility to periprosthetic osteolysis (PPOL) around total hip arthroplasty (THA) is associated with genetic variations in cytokine genes. Thus, cytokine pattern may differ in THA patients with mild and severe PPOL and in patients carrying a risk allele TNF-238*A, associated with severe osteolysis. **Methods:** Peripheral blood mononuclear cells (PBMCs) obtained from 31 THA patients with severe (n=23) and mild (n=8) osteolysis carrying (n=12)/non-carrying (n=19) a TNF-238*A risk allele were stimulated by lipopolysaccharide (LPS). Conditioned media were collected and protein concentrations of cytokines IL-2, IL-5, IL-10, IFN- γ , TNF- α , VEGF, and RANKL were measured by Luminex and ELISA. **Results:** Elevated expression of IL-2, IL-5, VEGF, and RANKL ($P < 0.05$) was observed in supernatants from patients with severe comparing to mild osteolysis, irrespective of TNF-238 genotype. Patients with TNF-238 GG genotype with severe osteolysis showed up-regulated expression of all studied cytokines, except IL-10 and VEGF, comparing to those with mild osteolysis. In severe osteolysis, the carriers of rare TNF-238*A allele showed lower expression of IL-2, IFN- γ , TNF- α , and RANKL ($P < 0.05$) when comparing to non-carriers. **Conclusion:** Alteration in cytokine-expression pattern in LPS-stimulated PBMCs from THA patients with mild and severe osteolysis and differing by TNF-238*A genotype was observed. Molecular mechanisms by which TNF-238*A allele increases a risk of severe osteolysis should be further investigated. **Grant support:** MZ CR VES16-31852A, MZ ČR VES15-27726A, IGA UP_2018_016, MH CZ – DRO (FNOL, 00098892).

2868W

Detecting *FLT3-ITD* and *NPM1* mutations in adult patients with acute myeloid leukemia in the Mongolia. S. Oyunbileg^{1,2}, D. Batbayar², S. Tungalag¹, Kh. Batsukh¹, M. Sukhbaatar¹, E. Elbegjargal¹, A. Boldbaatar¹, O. Norov², S. Baasabjav². 1) Hematology and Bone Marrow Transplantation Center of Mongolia, Ulaanbaatar, Mongolia; 2) Mongolian National University of Medical science, Department of Physiology.

Objective: *FLT3* is a cytokine receptor that is expressed on leukemic blasts in most cases of acute myeloid leukemia. The *FLT3* gene encodes a tyrosine kinase receptor that plays a key role in controlling survival, proliferation and differentiation of hematopoietic cells. *FLT3* mutations have emerged as important targets for molecular therapy.[1] Although *FLT3* internal tandem duplication (ITD) mutations in acute myeloid leukemia (AML) confer an adverse prognosis, co-occurrence with a nucleophosphomin (*NPM1*) mutation partially improves response and survival outcomes. Our study aims to establish a conventional detection method for *FLT3/ITD* and *NPM1* mutations in adult patients with acute myeloid leukemia in the Mongolia. **Materials and methods:** Genomic DNA was extracted from peripheral blood of 24 patients newly diagnosed and was diagnosed with acute myeloid leukemia. Primers were designed to distinguish between wild-type *FLT3* and *FLT3/ITD* variants. Methods using a polymerase chain reaction (PCR), PCR-ABI 3100 Genetic Analyzer (ROX-500 size standard), and PCR – agarose gel electrophoresis were compared. **Result:** *FLT3/ITD* mutations were found in 12.5% (3) of patients. *FLT3/ITD* was associated with clinical characteristics including higher white blood cell count than acute myeloid leukemia patients without *FLT3/ITD* mutation. By the Kaplan- Meyer method, over 40% of the Adult patients with acute myeloid leukemia overall survival was more than 400 days. In the group of *FLT3-ITD* mutation overall survival was less than 100 days. Of the 24 diagnosed AML samples whose cells were available to us, we successfully screened 2 samples for *NPM1* mutation. **Conclusion:** The average age of AML morbidity in adult Mongolia is 35 years, and this type of cancer is relatively young. The overall survival rate is over 40% longer than 400 days, which is longer than the previous years. It is concluded that the treatment of this leukemia is improved. In our study, *FLT3/ITD* was detected 12%. But *NPM1* was detected 8% (2 samples). This type of mutation was low in Mongolia. It is statistically significant to the number of blood cells in diagnosis compared to uneven prevalence groups, the same as in other researchers.

2869T

Intragenic large deletions of *F8* gene in Hemophilia A: Lessons from DNA sequencing analysis. T. Atik¹, E. Isik¹, B. Akgun¹, K. Kavaklı², M. Sezgin-Evim³, T. Tahtakeser⁴, G. Tuysuz⁵, F. Sahin⁶, H. Onay⁷, F. Ozkinay¹. 1) Pediatric Genetics, Ege University, School of Medicine, Izmir, Turkey; 2) Pediatric Hematology, Ege University, School of Medicine, Izmir, Turkey; 3) Pediatric Hematology, Uludag University, School of Medicine, Bursa, Turkey; 4) Pediatric Hematology, Kanuni Sultan Suleyman Training and Research Hospital, Istanbul, Turkey; 5) Pediatric Hematology, Akdeniz University, School of Medicine, Antalya, Turkey; 6) Hematology, Internal Medicine, Ege University, School of Medicine, Izmir, Turkey; 7) Medical Genetics, Ege University, School of Medicine, Izmir, Turkey.

Hemophilia A (HA) is an X-linked bleeding disorder caused by mutations in the *F8* gene. The most frequent mutations in *F8* are intron 22 (Inv22) and intron 1 (Inv1) inversions, which occur in approximately 40-50% and 1,5-3% of patients, respectively. However, different kinds of mutations such as point mutations, small and large deletions may cause HA with a clinical feature changing from mild to severe. In this study, molecular analysis results of 253 male HA patients referred to our molecular genetic laboratory for *F8* molecular analysis were retrospectively evaluated. Firstly, patients were analyzed for Inv22 and Inv1. Subsequently, sequencing using next-generation sequencing (NGS) platform was performed in Inv22 and Inv1 mutations negative patients. Of 253 HA patients, 111 (43,8%) were found to have Inv22 (n=108; 42,6%) or Inv1 (n=3; 1,2%), 122 (48,2%) point mutations or small deletions or insertions. In 20 patients, no mutations were detected with inversion analysis and sequencing analysis. When NGS results of these patients were reevaluated, in 8 of them (3,2%), amplification failures were detected in one or more exonic regions (del ex1, del ex2-9, del ex7-13, del ex9, del ex11-12, del ex14, del ex15-22 and del ex26) and it was considered that large deletions involving these unamplified regions are present. In 12 patients, no diagnostic results were obtained with techniques used. MLPA analysis has been planned to confirm these deletions and to reveal other CNVs in these 8 patients and remaining 12 patients. As a conclusion, regarding mutation types, in Turkish HA patients mutation distribution is similar to other populations and careful examination of NGS results may reveal large intragenic deletions in *F8* gene. .

2870F

Hyper IgE syndrome, recurrent infection and sex reversal: A new recessive disorder diagnosed by serendipity. K. Mandal, P. Srivastava, A. Aggarwal, S. Phadke, D. Saxena. Sanjay Gandhi Post-Graduate Institute of Medical Sciences, Lucknow, Uttar Pradesh, India.

Introduction: Autosomal recessive hyper-IgE recurrent infection syndrome (OMIM 243700) is caused by homozygous or compound heterozygous mutations in the *DOCK8* gene (OMIM 611432). We are presenting a family with a new disorder: Hyper IgE syndrome, recurrent infection and sex reversal where the proband had a 470 Kb homozygous deletion on chromosome 9p24.3 involving *DOCK8*. The homozygous deletion in *DOCK8* gene was chance detection, by sub-telomeric Multiplex Ligation-dependent Probe Amplification (MLPA). **Clinical details:** The proband was a 10 month old female born to consanguineous parents. She had skin rash, recurrent respiratory infections and diarrhoea. In view of very high IgE levels (675.4 IU/ml) and eosinophilia (46%) she was diagnosed to have hyper IgE syndrome. She had normal developmental milestones till 15 months of age when she had herpes encephalitis. **Investigation results:** Clinical exome sequencing by NGS was done to look for various putative sequence variations causing hyper IgE syndrome. No significant sequence variation was detected. However, it revealed that the sample was that of a male. Ultrasound of abdomen revealed normal female internal genitalia. To obtain fast results about the sex of the child, we did MLPA using SALSA MLPA P070 subtelomeres probemix (MRC Holland). It revealed Y fragments. This probe set contains a probe at exon 24 of *DOCK8* at 9p24.3. To our surprise, there was homozygous deletion at this point. Later on both parents were found to be heterozygous for the deletion. Karyotype (GTG banding) of the proband also confirmed male sex. Array CGH, performed, using Kit 750 K (Affymetrix, Santa Clara, CA, USA) identified 470Kb homozygous deletion on chromosome 9p24.3 (arr[hg19] 9p24.3(208,454-678,837)x0). This region included 2 important genes; *DOCK8* and *KANK1*. **Discussion:** 46,XY sex reversal is reported in some of the individuals with heterozygous deletion at 9p24.3. Since, in most 9p deletion cases (OMIM 158170) sex reversal was not observed, Hoo et al. (1989) postulated that there might be a recessive gene on 9p24 important for the development of testes. The DMRT cluster, located at 9p24.3, has been shown to be the strongest candidate for sex reversal. In our patient, the homozygous deletion involving *DOCK8* and *KANK1* genes, distal to the DMRT cluster, refuted the DMRT haploinsufficiency as the cause of sex reversal and points towards a still unknown testis determining factor at 9p24.3.

2871W**Diagnosis X-linked agammaglobulinemia disease by molecular testing.**

T.M. Ngo, M.T.P. Nguyen, N.D. Ngo, A.T.V. Nguyen, H.T. Le, H.M.T. Le. Vietnam National Children's hospital Hanoi, Vietnam.

Background: X-linked agammaglobulinemia (XLA) is a primary immunodeficiency characterized by a failure to generate immunoglobulins of all isotypes due to the absence of mature B cells and plasma cells. Mutations of the Bruton's tyrosine kinase (BTK) gene are responsible for most of the X-linked agammaglobulinemia. **Materials and methods:** In this study, we investigated 4 male patients with XLA-compatible phenotype from 4 unrelated Vietnam families by sequencing and MLPA technique. **Results:** We identified 4 different mutations, one of which were a large deletion mutation from exon 2 to exon 5 of *BTK* gene. Three mutations had been reported previously including one missense mutation (c.862C>T) and two non-sense mutations (c.843G>A and c.753G>A). DNA samples from 3 of 4 of the mothers of males with disease and found mutations in *BTK* gene were positive for the mutation found in their son and one mother was not found a non-sense mutation (c.843G>A) in her son. **Conclusion:** The sequencing and MPLA technique were an important tool for early confirmed diagnosis of XLA disease and detective accurate the carrier. It supports for clinicians to provide genetic counseling and prenatal diagnosis in the future. **Keywords:** X-linked disease, XLA, *BTK* gene.

2872T**The gene expression profile in the acute leukemia with minor *BCR/ABL1* fusion.**

J. Lee^{1,2}, M. Kim^{1,2}, S. Hong², Y. Kim^{1,2}. 1) Department of Laboratory Medicine, College of Medicine, The Catholic University of Korea, Seoul, Seoul, Korea, Republic of; 2) Catholic Genetic Laboratory Center, College of Medicine, The Catholic University of Korea, Seoul, Seoul, Korea, Republic of; 3) NGS Division2, TheragenETEX Co. LTD, Suwon, Korea, Republic of.

Background: It is interesting that minor *BCR/ABL1* is accompanied in acute lymphoblastic leukemia (ALL), acute myeloid leukemia (AML) and mixed phenotype acute leukemia (MPAL). Morphology and immunophenotyping are commonly used to diagnose the disease, however, it has not been well established how the same genetic aberration cause three different disease category. In this study, we tried to find specific gene expression profile according to disease category to clarify differences. **Methods:** Twelve patients with minor *BCR/ABL1* were enrolled; ALL (N=5), AML (N=3), MPAL (N=4) who were diagnosed in Seoul St. Mary's hospital. RNA was extracted from bone marrow specimen and gene expression analysis was performed using using the TruSeq RNA sample prep Kit and Illumina HiSeq2500 sequencer (Illumina, CA). Filtered reads were then mapped to the human reference genome and fusion genes were detected using three different tools. Differentially expressed genes were analyzed using Gene Set Enrichment Analysis and Gene Ontology. **Results:** Gene expression values were compared according to disease category. All minor *BCR/ABL1* were identified and an additional fusion was found which was confirmed by Sanger sequencing. Grouping with gene expression profile excellently reflected phenotypic characters of three disease category. We developed experimental simulation model using scoring the selected specific genes showing significantly overexpressed in each disease category; 251 genes in AML, 117 genes in ALL group. **Conclusions:** RNA sequencing detected fusion genes reliably. Gene expression profiles were significantly different in AML, ALL and MPAL even though they shared same driver genetic aberration. Therefore, gene expression profile made it possible to group acute leukemia because it reflected disease specific phenotypic characters. Through the model developed in this study, a gene expression profile can be used to diagnose acute leukemia by disease specific profile as well as fusion gene detection especially in case showing ambiguous phenotype. **Key Words:** minor *BCR/ABL1* fusion, Mixed Phenotype Acute Leukemia, Gene expression, RNAseq.

2873F

Segmental maternal uniparental disomy 14 (UPD14) identified by clinical exome sequencing in a patient with parathyroid adenoma and diverticulosis. S.-H.L. Kang¹, M. Mulatinho¹, J. Kianmahd², A. Chan³, J.A. Martinez-Agosto^{2,4,5}, J. Yeh³, H. Lee¹, UCLA Clinical Genomics Center. 1)

Department of Pathology and Laboratory Medicine, University of California, Los Angeles, Los Angeles, CA; 2) Department of Pediatrics, Division of Medical Genetics, University of California, Los Angeles, Los Angeles, CA; 3) Department of Pediatrics, Division of Pediatric Gastroenterology, University of California, Los Angeles, Los Angeles, CA; 4) Department of Human Genetics, University of California, Los Angeles, Los Angeles, CA; 5) Department of Psychiatry, University of California, Los Angeles, Los Angeles, CA.

Chromosome 14 harbors imprinted genes subject to differential expression dependent on the parental origin. The *MEG3* and *MEG8* genes in the chromosome 14q32 region are maternally imprinted. Maternal uniparental disomy 14 [UPD(14)mat] results in loss of paternally expressed genes and overexpression of the maternally expressed genes in the imprinted domain. The cardinal clinical features associated with UPD(14)mat include low birth weight, hypotonia, feeding problems early in life, motor delay, short stature, early puberty, dysmorphic features such as broad forehead and wide nasal tip, and small hands and feet. Utilization of clinical testing methodologies that screen for regions of homozygosity have increased detection and diagnosis of patients with imprinting disorders. Here we report a 15 year-old male with minor dysmorphic features, short stature, precocious puberty, elevated BMI, mild learning disabilities, asthma, hypercalcemia and hyperparathyroidism secondary to parathyroid adenoma, and chronic intermittent constipation/diarrhea. His clinical history includes prematurity at 29 weeks gestation due to IUGR, history of prolonged NICU course, and hypotonia. At the age of 15, he was hospitalized due to intestinal perforation and found to have diverticulitis. Clinical exome sequencing (CES) was requested to identify variants in genes associated with intellectual disability and cancer predisposition. Trio CES analysis identified no clinically significant variants; however, a 6.5 Mb region of homozygosity (ROH) was detected on chromosome 14 involving the 14q32.2-14q32.33 region. The sequencing data from the trio was consistent with all variants in this ROH region being only inherited from the mother with no evidence of a paternal deletion. These findings were confirmed using a PCR-based clinical UPD test. Although many of the patient's clinical features are consistent with UPD(14)mat; his GI history and parathyroid adenoma are not part of the clinical phenotype for this disorder and have not yet been reported in UPD(14)mat patients. To our knowledge, the clinical presentation of diverticulosis and parathyroid adenoma observed in this patient adds to the phenotypic spectrum of UPD(14)mat. Furthermore, the *MEG3* and *MEG8* genes are likely candidates for the majority of this syndrome's clinical spectrum.

2874W

Using linked-read sequencing to resolve pathogenic mutations and overcome pseudogene interference in *PKD1* for autosomal dominant polycystic kidney disease testing. C. Kao¹, M. Gonzalez², K. Wang³, R. Pellegri¹, W.M. Zhang², J. Garifallou¹, C. Kaminski¹, J. Smiler¹, J.D. Blumenfeld^{2,3}, H. Renner², H. Hakonarson^{1,4}. 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Weill Cornell Medical College, New York, NY; 3) The Rogosin Institute, New York, NY; 4) Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

Autosomal dominant polycystic kidney disease (ADPKD) is an adult-onset disorder arising from a constitutional mutation of either the *PKD1* or *PKD2* genes. Mutations in *PKD1* account for ~85% of cases and typically are associated with a more severe disease compared to *PKD2* mutations. ADPKD is typically an adult onset disorder, although cysts in the kidney are often present from birth or childhood. These cysts originate from cells that lose the remaining functional copy of *PKD1* or *PKD2* and slowly begin to grow replacing most of the normal mass of the kidney. By age 60, about half of those with ADPKD will experience kidney failure, requiring dialysis or kidney transplantation. These cysts can also develop in liver and pancreas, and complications can also occur in heart valves, abdominal walls, and blood vessels. Genetic testing of ADPKD is challenging. While sequencing of *PKD2* is relatively straightforward, the *PKD1* locus is difficult to resolve using target capture-based next generation sequencing (NGS) approaches reliant on short reads due to the gene's large size and the presence of six *PKD1*-like pseudogenes (*PKD1P1-P6*) located on chromosome 16p13.1, ~13-16Mb centromeric to *PKD1*, that have on average ~97.7% homology. Current diagnostic methods are laborious and expensive, relying on long-range PCR followed by Sanger sequencing. We demonstrate here how linked read sequencing can be applied to resolve/map *PKD1* reads accurately and unambiguously within a standard target capture based NGS framework. The input DNA is partitioned into gel-beads in emulsion (GEMs) using the Chromium instrument (10x Genomics, Inc.), and NGS-compatible mini-libraries are created within each GEM tagged with a unique 16mer barcode. All mini-libraries are pooled together and undergo whole-exome target enrichment with SureSelect baits (Agilent, Inc.), then sequenced on HiSeq2500 (Illumina, Inc). The GEM barcodes provide positional information linking short reads to the original input DNA fragments captured by a specific GEM, thus enabling us to map short, homologous reads that belong to *PKD1* from those originating from the pseudogenes. We apply linked read whole exome NGS on a number of "truth" samples with known pathogenic mutations in *PKD1* and provide a framework for designing a more targeted panel that can be developed for use in ADPKD genetic testing.

2875T

Infant with ARC syndrome caused by novel homozygous deletion in the *VIPAS39* gene. V. Adir¹, L. Sagi-Dain¹, C. Hartman², Y. Elenberg², A. Rotschild³, J. Haddad-Halloun¹, O. Sadeh¹, A. Peleg¹. 1) The Institute of Human Genetics, Carmel Medical Center, Haifa, Israel; 2) Pediatric Gastroenterology and Nutrition Unit, Carmel Medical Center, Haifa, Israel; 3) Department of Neonatology, Carmel Medical Center, Haifa, Israel.

Arthrogryposis renal dysfunction and cholestasis syndrome (ARC) is a severe autosomal recessive disorder, associated with liver and kidney dysfunction and congenital joint contractures. In most patients the disease is associated with mutations in the *VPS33b* gene, while in the minority of cases it is caused by *VIPAS39* gene mutations. *VIPAS39* protein forms a complex with the *VPS33B* protein, which functions in regulation of cell apical-basolateral polarity during hepatic and renal development, and in the sorting of lysosomal proteins. We report a female newborn, the firstborn child to first-degree Arab Muslim cousins. The infant was referred to our clinic due to facial dysmorphism, microcephaly, cleft palate, bilateral club foot, limb shortening, hypoplastic kidneys with echogenic cortex, thin corpus callosum, anteriorly displaced anus, abnormal renal and hepatic functioning and failure to thrive. Next-generation sequencing showed no sequence signals for exons 13-17 of the *VIPAS39* gene. This result was confirmed by PCR reaction and gel electrophoresis of these exons. Further PCR reactions and Sanger sequencing revealed the deletion breakpoints, defining the mutation as *14:g.77,899,605_77,904,634del5029*. This finding enabled us to develop a PCR test that would enable easy identification of carriers. The severity of ARC syndrome highlights the importance of prenatal diagnosis and carrier screening in the extended family.

2876F

Genotype and phenotype of Vietnamese patients with 11 β -hydroxylase deficiency. M.T.P. Nguyen^{1,2}, C.D. Vu¹, T.H. Nguyen², T.M.H. Nguyen¹, T.T.N. Ngo¹, M.T. Ngo¹, D.N. Ngo², V.H. Nong², T.H. Le¹, H.H. Nguyen². 1) Human genetics department, National Children Hospital, Hanoi, Hanoi, Viet Nam; 2) Institute of Genome Research, Vietnam Academy of Science and Technology.

Congenital adrenal hyperplasia (CAH) is an autosomal recessive disease which is characterized by a deficiency of one of the enzyme activities for the synthesis of cortisol from cholesterol by the adrenal cortex. More than 90% cases of CAH are due to 21-hydroxylase deficiency (21OHD), whereas 5-8% arises from 11 β -hydroxylase deficiency. CAH due to 11 β -hydroxylase deficiency is frequently associated with hypertension. Mutations in *CYP11B1* gene are the cause of the lack of 11 β -hydroxylase activity. In this study, we applied PCR and sequencing techniques to screen mutations in *CYP11B1* gene for 15 patients with 11 β -hydroxylase deficiency. Results showed that 8/15 patients had heterozygous mutation p.R43Q (06 boys and 02 girls): 05/06 male patients had classical simple virilization CAH; 01/06 male patient had classical salt wasting CAH; 01 female patient had classical simple virilization CAH and 01 female patient had non-classical CAH. There are 2/15 male patients had homozygous mutation p.R43Q but had difference phenotypes: one patient had hypertension and classical simple virilization CAH; the other had classical salt wasting CAH. This study also found 05 novel mutations in 06 patients, which is p.R51K, p.E147D, p.N152K, IVS6+5G>T and p.Y395X. All of six patients had classical simple virilization CAH. This studies also identified the effects of novel mutations (p. R51K, p. E147D, p.N152K) on the level of expression of the 11 β -hydroxylase enzyme. All three mutations affect the expression level of the enzyme. In particular, p.R51K reduced the activity of 11 β -hydroxylase to about 29% compared to control. p.E147D mutation decreased to 48% and p.N152K reduced to 36% compared to control. The mutations p.R51K, p.E147D, p.N152K affected the structure of the *CYP11B1* enzyme by altering the hydrogen bonding with adjacent amino acids. We also predicted of mutation in IVS6+5G>T gene are likely to affect mRNA synthesis using MaxEntScan software (Yeo and Burge, 2004-<http://gens.mit.edu/burgelab/maxent/5ss>). Understanding of *CYP11B1* genotype and phenotype is the basis for genetic counseling and prenatal diagnosis in the future.

2877W

A novel next generation sequencing based genetic test platform for congenital adrenal hyperplasia due to CYP21A2 defects. Q. Lao¹, Y. Zhao², M. Mehta², B. Tran², D. Merke^{1,3}. 1) National Institutes of Health Clinical Center, Bethesda, MD; 2) Frederick National Laboratory for Cancer Research, Frederick, MD; 3) Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, MD.

Congenital Adrenal Hyperplasia (CAH) due to CYP21A2 defects is an autosomal recessive disease of steroidogenesis and affects approximately 1 in 15,000 in the severe classic form. Mandatory neonatal CAH screening relies on a sometimes ambiguous hormonal test plagued by false positives and false negatives rather than CYP21A2 genotyping, which has been historically challenging. CYP21A2 is located in a low copy repeat named RCCX module composed of "STK19-C4A-CYP21A2P-TNXA-STK19P-C4B-CYP21A2-TNXB", 4 pairs of highly homologous genes in tandem in the chromosome 6 major histocompatibility complex region. Most of the CYP21A2 mutations are due to gene conversions from CYP21A2P, which also presents a major obstacle of pseudogene interference in the mapping procedure of standard NGS approaches. Commonly existing copy number variations (1-4 copies in haploid) due to unequal crossovers throughout the RCCX module lead to further chaos in mapping, especially for the commonly found CAH alleles with 30-kb deletion or duplications. We aimed to develop a novel NGS based CAH genetic test platform that can 1) overcome the mapping interference by pseudogenes and copy number variations; 2) determine the haplotype/phase of detected mutations without the requirement of parental genotyping; 3) also cover TNXB since about 10% of the CAH alleles have contiguous deletions extending into TNXB causing Ehlers-Danlos Syndrome in autosomal dominant manner. We applied a 10X GENOMICS CHROMIUM partitioning and molecular barcoding system followed by Illumina MiSeq based standard targeted sequencing on two index gDNA samples of typical CAH genotypes: sample A has bi-modular RCCX in both alleles while sample B has one bi-modular allele and one mono-modular allele with a TNXA/TNXB chimera. Standard 10X GENOMICS software mapped and phased sample A successfully but failed on sample B, likely due to the interference by regional copy number variations. Still utilizing the linked reads generated by the 10X genomics system, we developed a novel *in silico* approach to overcome the interference by homologous pseudogene reads and regional copy number variations. Our preliminary results suggest that the incorporation of 10X genomics, Illumina NGS and our *in silico* analysis meets our goals and poses to be the most accurate and informatic high throughput genetic test platform for CAH.

2878T

Transaldolase deficiency (TALDO-D) as a mimic of gestational alloimmune liver disease (GALD). H. Yen¹, B. Mendelsohn², J. Vargas¹, J. Carmichael¹, L. Judge², N. Weiss², S. Winter², H. Gao¹. 1) Fulgent Genetics, Temple City, CA; 2) Department of Pediatrics, University of California, San Francisco, San Francisco, CA; 3) Medical Genetics and Metabolism, Valley Children's Hospital, Madera, CA.

Transaldolase deficiency (TALDO-D) is caused by mutations in the *TALDO1* gene. Transaldolase is the rate-limiting enzyme for the non-oxidative branch of the pentose phosphate pathway, a glucose metabolic pathway that generates NADPH and pentoses. TALDO-D is a rare inborn autosomal recessive disease that presents with a wide range of phenotypes, from IUGR, hepatosplenomegaly, congenital heart defect, to thrombocytopenia. The first patient with TALDO-D was described in 2001. Since then, no more than 40 cases have been reported in literature. Here we present two severe cases in a family resulting in early infant deaths. The older sister (SF) died at the age of 4 weeks and the younger brother (BB) died at the age of 6 weeks. Patient BB was initially suspected to have GALD due to severe neonatal hemochromatosis observed at autopsy in the older sister. Therefore, the mother was treated with intravenous immunoglobulin for the subsequent pregnancy, and patient BB received an exchange transfusion. Nevertheless, patient BB developed lethal liver disease. Both siblings had IUGR, neonatal hemochromatosis, cholestasis, splenomegaly, coagulopathy, thrombocytopenia, intracranial hemorrhage, and hypertrophic cardiomyopathy. Liver failure was the major diagnosis established at autopsy for both siblings. Subsequently, a whole exome trio analysis was performed on samples from BB and the parents. The analysis identified two different missense variants in the *TALDO1* gene. The first, a paternally-inherited heterozygous variant, NM_006755.1:c.574C>T (p.Arg192Cys), has previously been reported in multiple patients with TALDO-D in the homozygous state. The second, a maternally-inherited heterozygous variant, NM_006755.1:c.715C>T (p.Arg239Cys), has not been previously reported in literature. This variant is located in the highly conserved putative phosphate binding site, which is required for transaldolase activity. Upon clinical suspicion of TALDO-D in BB, a targeted mutation test was performed on the affected sister, resulting in the detection of both missense mutations. This study strongly supports the pathogenicity of the novel variant, p.Arg239Cys, in the *TALDO1* gene. TALDO-D can mimic GALD and should be considered in cases of GALD, as recurrence risk differs between the two disorders. Genetic or polyol testing should be considered in cases of suspected GALD to address the possibility of TALDO-D.

2879F

Mapping of FAH mutations in hepatorenal tyrosinemia type 1 in consanguineous families. *M. Zahoor¹, S. Ijaz², K. Ramzan³, H. Cheema³.* 1) Institute of Biochemistry & Biotechnology, Faculty of Biosciences, University of Veterinary & Animal Sciences, Lahore, Punjab, Pakistan; 2) Department of Genetics, King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia; 3) 3Department of Pediatric Gastroenterology and Hepatology, The Children's Hospital & The Institute for Child Health, Lahore, Pakistan.

Hepatorenal tyrosinemia type 1 (HT1) is a rare inborn error of tyrosine catabolism with a worldwide prevalence of 1 out of 100,000 live births affecting in infancy. HT1 is clinically characterized by hepatic and renal dysfunction resulting from the deficiency of fumarylacetoacetate hydrolase (FAH) enzyme, caused by recessive mutations in the *FAH* gene. Seven Pakistani families, having affected child with HT1, were enrolled over a period of 6 years. All regions of the *FAH* gene spanning exons and splicing sites were PCR amplified and mutation analysis was carried out by direct sequencing. Results of sequencing were confirmed by PCR-RFLP analysis. Evolutionary analysis of the *FAH* has also been performed along with mutation spectrum for treatment strategy. Four different *FAH* mutations were found to co-segregate with the disease phenotype. Two of these *FAH* mutations (c.192G>T and c.974C>T) and one novel (c.67T>C; p.S23P) were mapped in one family each while (c.1062+5G>A [IVS12+5G>A]) found in four families. Five of the affected children has been died from during course of time as they were presented late with acute form segregating (c.1062+5G>A [IVS12+5G>A] and c.192G>T) mutations. Most of the children do not survive due to delayed diagnosis followed by untimely treatment. The mutation spectrum in the disease is different on severity and prognosis of the disease. This tragic condition advocates the establishment of expanded newborn and carrier screening program for HT1 in affected families and sporadic cases specially with mutations having high severity and founder affect.

2880W

Compound heterozygous missense variants in the *ADAM17* gene cause neonatal inflammatory skin and bowel disease 1. *I. Imoto^{1,2}, M. Saito², K. Suga², T. Naruto², T. Kohmoto², M. Otsu², K. Horiuchi³, H. Nakayama⁴, S. Higashiyama⁴, R. Nakagawa², S. Tange², K. Masuda², S. Kagami².* 1) Aichi Cancer Center, Nagoya, Aichi, Japan; 2) Tokushima University, Tokushima, Japan; 3) Keio University School of Medicine, Tokyo, Japan; 4) Ehime University Graduate School of Medicine, Toon, Ehime, Japan.

A disintegrin and metalloprotease 17 (ADAM17), also called TNF α converting enzyme (TACE), is the major sheddase for various proteins including TNF α and EGFR ligands. The genetic deficiency with null mutations of ADAM17 has been linked to neonatal inflammatory skin and bowel disease 1 (NISBD1). We identified the compound heterozygous missense variants in *ADAM17* in the infant with erythroderma with exudate in whole body, recurrent skin infection and sepsis, and prolonged diarrhea through targeted exome sequencing. *ADAM17* was transcribed equally from both alleles and translated. Both missense variants locate outside of the catalytic domain. To determine the activity of two ADAM17 mutants, we constructed plasmids expressing ADAM17 with those mutations (ADAM17-Mut1 and Mut2), and transfected these mutant plasmids, plasmid expressing wild type ADAM17 (ADAM17-WT), or empty vector into ADAM17/ADAM10 double knockout mouse embryonic fibroblasts (ADAM10,17^{-/-}mEFs). ADAM17-WT, Mut1, and Mut2 proteins were expressed at similar levels on the plasma membrane of transfected cells. However, impaired basal and phorbol myristate acetate (PMA)-stimulated TACE activities were observed in both ADAM17-Mut1 and Mut2 compared with ADAM17-WT in in vitro TACE assay using cell lysates from transfected cells. In addition, PMA-inducible shedding activity for alkaline phosphatase (AP)-tagged TNF α , TGF α , or HB-EGF was strongly reduced in both ADAM17 mutants compared with ADAM17-WT in transfected cells. Because clinical features of the patient seem to be explained by the defect of ADAM17 activity, compound heterozygous missense variants detected in *ADAM17* may lead to the development of NISBD1.

2881T

Implementation of three-tier sequencing algorithm improves cystic fibrosis newborn screening. D.M. Kay¹, M. Berdella², J. Boyer³, J. DeCelle-Germana⁴, A. Dozor⁵, C.N. Fortner⁶, R. Giusti⁶, D.M. Goetz⁷, L. Guida⁸, E. Hughes¹, R. Kaslovsky⁹, C. Kier¹⁰, L. Krein¹, E. Langfelder-Schwind², A. Laterza-Ozrowski¹⁰, L. Martin¹¹, C.A. Saavedra-Matiz¹, H. Sadeghi¹², Z. Sultana³, A. Ting¹³, C. Stevens¹, M. Caggana¹, New York State Cystic Fibrosis Newborn Screening Consortium. 1) Wadsworth Center, New York State Department of Health, Albany, NY; 2) Mount Sinai Beth Israel, New York, NY; 3) Westchester Medical Center, Valhalla, NY; 4) Cohen Children's Medical Center, New Hyde Park, NY; 5) Upstate Medical University, Syracuse, NY; 6) NYU Langone, New York, NY; 7) Children's Hospital of Buffalo, Buffalo, NY; 8) Good Samaritan Hospital, West Islip, NY; 9) Albany Medical Center, Albany, NY; 10) Stony Brook University, Stony Brook, NY; 11) University of Rochester Medical Center, Rochester, NY; 12) Columbia University Medical Center, New York, NY; 13) Mount Sinai Medical Center, New York, NY.

Cystic Fibrosis (CF) newborn screening (NBS) began in NY using an IRT-DNA algorithm in 2002. Despite several improvements, a high false positive rate and low positive predictive value (PPV=3.7%) prompted implementation of a three-tier IRT-DNA-SEQ algorithm on 12/1/2017. Infants with high immunoreactive trypsinogen (IRT; top 5%) were screened for a panel of 39 variants. Those with one variant or ultra-high IRT (top 0.1%) were comprehensively sequenced using Illumina's MiSeqDx CF Clinical Sequencing Assay with supplemental testing to target specific deletions. Variants were confirmed via Sanger sequencing. Infants with two *CFTR* variants were referred to Specialty Care Centers for diagnostic testing, and the primary care provider and hospital of birth were issued reports for all screen negatives, with recommendation for genetic counseling of carriers. Benign/likely benign variants were not reported. During the first four months, 38 infants with two variants were referred. Forty-four *CFTR* alleles were known CF-causing/pathogenic, 19 were varying clinical consequence (VCC), and 13 were uncertain significance (VOUS). Sequencing identified 161 as carriers and 65 with no clinically significant variants. With the original IRT-DNA algorithm, 264 infants would have been referred. The average number of referrals decreased from 73 to 9 per month. Median age at first sweat test was 24 days (range 10–72). Nine infants had confirmed CF, one is classified as possible CF and diagnosis is pending for two (PPV=27.8%; 10/36). Twenty-six infants with two variants and intermediate/low sweat chloride levels are classified as CF-Related Metabolic Syndrome (CRMS)/CF Screen Positive, Inconclusive Diagnosis (CFSPID) and will be followed. The CRMS/CFSPID to CF ratio increased from 0.9 to 2.6, and 25/26 CRMS/CFSPID cases carried at least one VCC or VOUS. Implementation of IRT-DNA-SEQ resulted in 85.6% reduction in referrals and 7-fold increase in PPV. We project 110 infants will be referred annually, and 650 infants will be reported as negative after comprehensive genetic analysis. Additional testing increases lab workload, turnaround times and cost, but is balanced by reduction in follow-up by NBS staff and Care Centers, and an overall reduction in healthcare costs due to elimination of most false positives. Identification of asymptomatic infants with two *CFTR* variants and intermediate/negative sweat results poses new clinical and practical challenges for care teams and families.

2882F

Clinical and molecular features of X-linked hypophosphatemic rickets patients in a cohort of 32 unrelated Han families in Southern China. L. Liu, Y. Lin, M. Wu, X. Li, J. Cheng, Y. Huang, C. Zeng, X. Mao, Z. Zhou, W. Zhang, C. Li, H. Sheng, Y. Cai, Y. Shao, X. Su, Z. Liu, M. Peng, M. Jiang. Department of Genetics and Endocrinology, Guangzhou Women and Children's Medical Center, Guangzhou, Guangdong, China.

Objectives: X-linked hypophosphatemic rickets (XLHR) is the most common HR caused by deleterious mutations within *PHEX* gene with a representative feature of hypophosphatemia, skeletal deformities and growth retardation. As XLHR is not well-known and affected individuals are frequently diagnosed as vitamin D deficiency, our study aimed to provide accurate diagnosis and avoid mistreatment. Moreover, the clinical, radiological, biochemical and mutational spectrum of XLHR patients in Southern China were also investigated in this study. **Methods:** A cohort comprised 44 pediatric HR patients and 57 relatives from 40 unrelated families were recruited for genetic analysis with a combined strategy of DNA sequencing and qPCR assay. The clinical features, radiological findings, and serum biochemical parameters of probands were enrolled and further scored for phenotypic correlation analysis. **Results:** 29 diverse mutations were identified in 36 probands (81.82%) along with the *PHEX* gene, and 15 of them were novel. 2 high-frequent single nucleotide replacement mutations, c.1601C>T (p.P534L) and c.2239C>T (p.R747X), were found in 3 unrelated families. Physical examinations, radiological inspections, biochemical tests and genetic analyses revealed typical features of XLHR, and the degree of phenotype might be potentially correlated to mutation type, as the more detrimental mutation would trigger more severe skeletal impairment.

Conclusions: Our study extends the mutational spectrum of *PHEX* gene, contributes to accurate diagnosis, and provides evidence for potential hotspots and predictor of disease severity. Our study also addresses *PHEX* gene as the most common but extremely unstable causative gene for HR, which indicates the priority to survey *PHEX* gene when a patient was suspected to suffer from HR. In addition, a supplementary qPCR or MLPA assay is suggested to remedy the weakness of classical sequencing method in the field of gross insertion/deletion.

2883W

Genetic study of 20 Egyptian patients with Dyggve-Melchoir Clausen syndrome and identification of 5 novel mutations. G.A. Otaify¹, M.S. Aglan¹, N. Patel², S. Ismail³, M.S. Zaki⁴, G.M.H. Abdel-Salam¹, K. Amr⁵, G. Hosny⁶, F.S. Alkuraya^{2,5,6}, S.A. Temtamy¹.

1) Department of Clinical Genetics, Division of Human Genetics and Genome Research, Centre of Excellence for Human Genetics National Research Centre, Cairo, Egypt; 2) Department of Genetics, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia; 3) Department of Medical Molecular Genetics, Division of Human Genetics and Genome Research, Centre of Excellence for Human Genetics National Research Centre, Cairo, Egypt; 4) Department of Orthopedic Surgery, Banha University, Banha, Egypt; 5) Saudi Human Genome Project, King Abdulaziz City for Science and Technology, Riyadh, Saudi Arabia; 6) Department of Anatomy and Cell Biology, College of Medicine, Alfaisal University, Riyadh, Saudi Arabia.

Abstract Dyggve–Melchior–Clausen (DMC) syndrome is a rare, autosomal recessive disorder characterized by severe progressive spondyloepimetaphyseal dysplasia leading to disproportionate short stature with short trunk, in addition to microcephaly, facial dysmorphism, and intellectual disability. The disease is caused by mutation in the *DYM* gene which encodes Dymeclin protein. We report in this study 20 patients from 12 families with DMC of variable clinical severity representing the largest reported cohort from the same ethnic background. Sequencing of all coding regions of *DYM* gene revealed identification of 7 mutations; among them 5 novel (nonsense and splice site) mutations and 2 previously reported (nonsense and deletion) mutations. All mutations were homozygous except for one patient who had compound heterozygous mutation. Genotype/phenotype correlation was investigated and we found that patients with nonsense mutations attained more severe phenotype. This study highlights the possibility of variable evolution of the genotype spectrum of patients with identical phenotype. The study added 5 novel mutations to the 28 previously reported ones expanding the spectrum of the molecular background of these disorders. **Key words:** Dyggve–Melchior–Clausen- dysplasia- Dymeclin- novel- mutation- spondyloepimetaphyseal.

2884T

The endothelial nitric oxide synthase gene variants as a possible predisposing factor for ankylosing spondylitis in a Turkish population. K. Ozdilli¹, A.F. Nursal², S. Gursoy³, M. Pehlivan⁴, M.S. Akaltun⁵, S. Pehlivan⁶.

1) Pediatric Bone Marrow Unit, Medipol University, Medipol University Hospital, Istanbul, Turkey; 2) Department of Medical Genetics, Faculty of Medicine, Hitit University, Corum, Turkey; 3) Department of Physical Medicine and Rehabilitation, Faculty of Medicine, Gaziantep University, Gaziantep, Turkey; 4) Department of Hematology, Faculty of Medicine, Gaziantep University, Istanbul, Turkey; 5) Department of Medical Biology, Istanbul Faculty of Medicine, Istanbul University, Istanbul, Turkey.

Ankylosing spondylitis (AS) is a chronic inflammatory arthritis that mainly affects the sacroiliac joints and spinal joints. Nitric oxide (NO), a signaling molecule, is constitutively produced from L-arginine by the enzyme endothelial nitric oxide synthase (eNOS). NO is considered as a pro-inflammatory mediator that induces inflammation. Several polymorphisms of the *eNOS* gene have been identified, and their association with various diseases has been explored. *eNOS* variants might have a functional potential to affect the bioavailability of *eNOS*. *eNOS* 894G/T (a guanine/ thymine substitution at position 894 on exon 7) variant leads a change from glutamate to aspartate at position 298 (rs1799983). Other variant of the *eNOS* is intron 4 a variable number of tandem repeats (VNTR) polymorphism which causes basal NO production. It was suggested that the *eNOS* gene variants have been associated with different autoimmune diseases. The aim of this study was to investigate the influence of the *eNOS* gene (894G/T) and (intron 4 VNTR) variants on susceptibility to AS in a Turkish population. The study included 78 AS patients and 100 healthy controls. All patients fulfilled the New York criteria for AS. The *eNOS* gene variants were genotyped by Polymerase Chain Reaction (PCR) and/or PCR based-Restriction Fragment Length Polymorphism (RFLP) technique. Odds ratio (OR) with 95% confidence interval (95%CI) were calculated using the χ^2 test. We found statistically significant difference in the TT genotype distribution of the *eNOS* (894G/T) variant in patients and control groups. ***eNOS* (894G/T) TT genotype was more frequent in the patients compared to the healthy control group ($p=0.001$).** There was no significant difference in the frequency of *eNOS* (intron 4 VNTR) genotypes and alleles between patients and healthy controls. There was no significant difference between *eNOS* variants genotype distributions and clinical features including HLA-B27, Bath Ankylosing spondylitis disease activity index (BASDAI), Bath Ankylosing Spondylitis Metrology Index (BASMI), Bath Ankylosing Spondylitis Function Index (BASFI), Bath Ankylosing Spondylitis Radiology Index (BASRI), and assessing quality of life (AS-QoL). *In conclusion, we have demonstrated that the eNOS gene (894G/T) variant may represent susceptible factor for AS in Turkish population, but further studies including larger sample sizes are needed to confirm this results.*

2885F

Ring chromosomes aberrations at a pediatric Mexican hospital: Two cases with mosaicism of chromosome 13, 46XY / 46, XY, r (13) and chromosome 18, 46, XY / 46, XY, r (18). *M. Hurtado Hernandez¹, J.M. Aparicio Rodriguez², M. Barrientos Perez³.* 1) Cytogenetics.; 2) Genetics; 3) Endocrinology, Hospital Para El Niño Poblano, San Andrés Cholula Puebla, Mexico.

The autosomic alteration due to a ring formation is a rare aberration of either chromosome 13 and 18 which is in relation with phenotypic malformations, neurologic problems and genital abnormalities. Two clinical polymalformed cases with skull treboliform dismorphies with early seizures and malformed genitals with micropenis is presented from four of the total patients found in this study. Among chromosomal alterations, the ring of autosomic chromosome 13 and 18 are not frequent, the main phenotypic alterations in this study are in relation to neurological, genital and craniofacial malformations. Taking in consideration that mutations or chromosome aberrations are alterations in the chromosome number or structure. They are mainly considered due to gametogenesis inborn error (meiosis) or during the zygote first cellular divisions. All these alterations might be observed during metaphase from the cellular cycle, where DNA loses are seen due to DNA repair processes deficiency or total absence, among others. 4617 chromosomal studies were performed at Hospital Para El Niño Poblano (Pediatric Hospital) in Mexico (from 1992 to 2011) were 34.6% (1596 patients) showed different chromosomal alterations and only two patients showed ring chromosome aberrations. These chromosome changes are classified as structural alterations. Both pediatric patients with these genetic diseases are described in this study analyzing their clinical characteristics, medical or surgical treatments according to the phenotypic alterations..

2886W

A 235 Kb deletion at 17q21.33 encompassing the COL1A1 gene, and two additional secondary copy number variants in an infant with type I osteogenesis imperfecta: A rare case report. *N. Numbere¹, D.R. Weber², M.A. Iqbal¹.* 1) Department of Pathology and Laboratory Medicine, University of Rochester Medical Center, Rochester, NY; 2) Pediatric Endocrinology, Department of Medicine, University of Rochester Medical Center, Rochester, NY.

Osteogenesis imperfecta (OI) is a rare genetic disorder characterized by increased susceptibility to fractures due to increased bone fragility. It is most commonly inherited in an autosomal dominant fashion, but recessive OI has also been reported. At least nine types (types I-IX OI) have been described based on clinical, radiologic and genetic features. About 90% of cases are due to mutations in COL1A1 (17q21.33) or COL1A2 (7q21.3) that result in quantitative or qualitative defects in type 1 collagen, a key structural constituent of the extracellular matrix of bone. In contrast to mutations, deletion of the entire COL1A1 gene is a rare cause of OI (Van Dijk et al. 2010; Bardai et al. 2016). We present a case of type I OI in a Caucasian female referred at 10 months of age for investigation of multiple fractures with minimal or no known trauma, small stature and blue sclerae. The patient had a suspected left femur fracture and bowing of left femur detected at prenatal ultrasound. Array comparative genomic hybridization revealed a previously unreported 235 kilobase (kb) pathogenic deletion at 17q21.33 encompassing COL1A1, ITGA3, PDK2, SGCA3, HILS1 genes and other RefSeq genes. Additionally, a 283 kb gain at 8p21.3 from 19375578 to 19658401 base pairs encompassing the CSGALNACT1 gene and a 163 kb loss at 10q21.3 from 68395508 to 68558659 base pairs encompassing CTNNA3 gene were found. At her most recent follow up at 21 months of age, she had suffered multiple lifetime fractures including fractures of the right ulna, right fibula, and multiple ribs. Her father has 4-5 lifetime fractures, blue sclera and normal stature. High-density targeted array done at another laboratory revealed him to be a carrier of a heterozygous 14.5 kb deletion of the entire COL1A1 gene. Together with previously reported cases of deletions, it is recommended that the presence of such large deletions and secondary copy number variations be considered for identification of common clinical features which may establish as yet unknown variations in OI.

2887T**Targeted-NGS on skeletal dysplasias: Evaluating the diagnostic yield.**

K.C. Silveira, T.Y. Kanazawa, C. Silveira, C.A. Moreno, M.D.J. Lacarrubba-Flores, D.P. Cavalcanti. Skeletal Dysplasia Group, Department of Medical Genetics, University of Campinas (UNICAMP), Campinas, São Paulo, Brazil.

Skeletal dysplasias (SD) represent a diagnostic challenge due to the great clinical-etiological heterogeneity. Currently, NGS methodologies are the best tools to reach the final diagnosis. Here, we present the preliminary results of a big Brazilian cohort of patients with SD studied by targeted-NGS. Two custom NGS panel (TruSeq Custom Amplicon and Nextera - Illumina), with selected genes according to the previous radiological diagnosis, were used. The sequencing of 119 patients was performed on a MiSeq machine. The Sanger sequencing (SS) was also used for confirmation of the pathogenic variants, as well as to investigate the insufficiently covered regions. Patients were previously classified into 19 different groups according to the Nosology of the Genetic Skeletal Disorders [AJMG, 2015; 167A:2869]. However, for the analysis the patients were also sorted into two group: group 1- assured diagnosis (N=84), and group 2- probable diagnosis (N=35). The current results include 97 patients since 22 patients are still under analysis. Pathogenic variants were found in 75/97 (77.3%) of patients. However, the diagnostic yield was higher in patients of the group 1 - 88.2% (N= 67/76), or perhaps even higher, if it takes in account that eight among the negative patients are still under SS for investigation of bad coverage regions. As presumed, the positivity rate was lower in patients of the group 2 - 38.1% (N=8/21). For the remainder 14 negative patients, six were submitted to the WES. All patients classified in the following groups, TRPV4, Slender bones, Cleidocranial D., MED-PseudoAch, Perlecan, COL11, were positive. Other groups with very good diagnostic yields were as follows: ciliopathies (83.3%), OI (76.9%), and COL2A1 (70.8%). In conclusion, these results confirm the targeted-NGS as a good strategy for molecular diagnosis of SD and show that the diagnostic yield is a consequence of the previous diagnosis, being better among those with an assured radiological diagnosis. Finally, regions insufficiently covered by NGS are still a hurdle to be overcome. Financial Support: CNPq #590148/2011-7 / Faepex #402/15 / FAPESP # 98/16006-6; 2015/22145-6.

2888F

Novel *CARD14* mutation in pityriasis rubra pilaris type V. T. Miyauchi¹, T. Nomura¹, S. Suzuki¹, M. Takeda¹, K. Natsuga¹, H. Hata¹, Y. Fujita¹, K. Arita¹, W. Nishie¹, M. Akiyama², H. Shimizu¹. 1) Hokkaido University, Sapporo, Hokkaido, Japan; 2) Nagoya University, Nagoya, Aichi, Japan.

Pityriasis rubra pilaris (PRP) is a group of rare chronic inflammatory skin disorders, clinically characterized by keratotic follicular papules, well-demarcated scaly erythematous plaques interspersed with distinct islands of uninvolved skin, and palmoplantar keratoderma. Up to 6.5% of all PRP patients have a family history. The autosomal dominant subtype, which is classified as PRP type V (PRPV), is associated with gain-of-function mutations in *CARD14*, encoding caspase recruitment domain containing protein 14 (*CARD14*) that activates the noncanonical nuclear factor-kappa B (NF-κB) pathway. Because only 12 families with *CARD14* mutations have been reported in the literature, the clinical features of PRPV remain unsatisfactorily understood. In this study, we analyzed a 60-year old Japanese male with erythroderma, palmoplantar keratoderma, and ectropion all of which had appeared since his birth. Joint contractures of the fingers were also noted. Notably, he has experienced multiple skin tumors including malignant melanoma, squamous cell carcinoma, seborrheic keratosis, and pyogenic granuloma. Histopathological findings showed hyperkeratosis with parakeratosis, acanthosis, and perivascular inflammatory infiltration in the upper dermis. Based on these observations, the patient was initially diagnosed with congenital ichthyosiform erythroderma (CIE), an autosomal recessive genodermatosis. To confirm the diagnosis, we performed whole-exome and Sanger sequencing using genomic DNA extracted from his peripheral blood cells. Unexpectedly, these tests identified a novel heterozygous mutation c.356T>C (p.Met119Thr) in *CARD14*, whereas neither homozygous nor compound heterozygous pathogenic mutations were detected in any of the 12 known causative genes for CIE: *ABCA12*, *TGM1*, *ALOXE3*, *ALOX12B*, *CYP4F22*, *NIPAL4*, *PNPLA1*, *LIPN*, *CERS3*, *SDR9C7*, *ST14*, and *SULT2B1*. This *de novo* mutation was absent in available databases. Furthermore, NF-κB luciferase reporter assay revealed that mutant *CARD14* resulting from the c.356T>C allele showed 2-fold increase in NF-κB activity compared with wild-type, suggesting the pathogenicity of the mutation. These findings finally led us to diagnose the patient as having PRPV. In summary, this study further expands our understanding of clinical and genetic features of PRPV. Multiple occurrence of skin tumors in our patient may suggest potential tumorigenesis associated with the disease, although further studies are needed to assess this hypothesis.

2889W

Exome based genetic testing for skeletal conditions: One lab's experience in ending the diagnostic odyssey. J. Dong, M. Kyriass, R. Vanneste, R. Steiner, J. Weber. PreventionGenetics, Marshfield, WI.

Skeletal disorders are a clinically and genetically heterogeneous group of diseases that impair skeletal or joint function. The diagnosis of these patients is based on clinical presentation, ultrasound or X-ray findings, and genetic testing. The diagnostic process for these challenging disorders can take up to 7 years (Velasco and Morales. 2017. PubMed ID: 29158687). Genetic skeletal disorders can be divided into 436 conditions based on the 2015 Revision of the Nosology and Classification of Genetic Skeletal Disorders and more than 300 genes have been implicated in these disorders. Recently, PreventionGenetics developed a unique Skeletal Disorders and Joint Problems Sequencing Panel with Copy Number Variation (CNV) detection based on whole exome capture and sequencing. This panel includes 258 genes associated with a variety of genetic skeletal conditions mainly selected from the article "Nosology and Classification of Genetic Skeletal Disorders –2010 Revision". These include genes related to achondroplasia/thanatophoric dysplasia, osteogenesis imperfecta, short-rib thoracic dysplasia, hypophosphatemic rickets, polydactyly/brachydactyly/syndactyly, and Ellis-van Creveld syndrome. Of the cases reported to date, approximately 40% of all test orders have been prenatal requests due to abnormal ultrasound findings. We have identified at least one pathogenic or likely pathogenic variant in 39% of these samples. For examples, a pathogenic variant in the *FGFR3* gene was found in a fetus with shortened long bones with bowing, frontal bossing, very small chest circumference, suspicious for a cloverleaf skull, large head circumferences and protuberant abdomen on 21 week's ultrasound findings. In another prenatal case with short long bones, polyhydramnios, polydactyly on 31 week's ultrasound findings, a hemizygous truncating variant in the *EVC2* gene and a large heterozygous deletion encompassing two exons of the *EVC2* gene were detected, emphasizing the added value of CNV detection. In summary, exome based panel testing is an effective method to help formulate accurate diagnosis for skeletal conditions, ending the diagnostic odyssey for patients and their families. Since this test is performed using exome capture probes, reflex to whole exome testing is available for negative or indeterminate cases.

2890T

Complex mutation landscape of pseudoxanthoma elasticum revealed by multigene next-generation sequencing panel. A. Saeidian¹, E. Ryu¹, L. Youssefian², A. Touati¹, D. Li¹, H. Vahidnezhad¹, E. Duval², E. Oliphant², SF. Terry², A. Ertel¹, P. Fortina¹, J. Uitto¹, Q. Li¹. 1) Thomas Jefferson University, Philadelphia, PA; 2) PXE International, Inc., Washington, DC, USA.

Pseudoxanthoma elasticum (PXE) is characterized by calcium hydroxyapatite deposition in the skin, eyes and the cardiovascular system. The majority of cases are caused by mutations in the *ABCC6* gene in an autosomal recessive inheritance pattern. Attesting to the genetic heterogeneity are recent demonstrations of mutations in the *ENPP1* and *GGCX* genes in patients with PXE-like phenotypes. Increasing evidence suggests the presence of predisposing genetic factors that modify the PXE phenotype, and deep phenotyping has revealed inter- and intra-familial phenotypic heterogeneity among PXE patients. In order to find causative genes or potential modifiers with significant roles in the severity of the phenotype, we developed a disease-specific, gene-targeted next-generation sequencing panel with the Illumina platform. The panel contains 29 genes, including *ABCC6*, all of which are associated with disorders of calcification. Pathogenic variants were identified by variant calling using databases including dbSNP, 1000 Genomes Project and ExAC, and pathogenicity was analyzed by prediction softwares. Among 401 probands, 130 distinct variants were identified in *ABCC6* consisting of 81 missense (62%), 22 indel (17%), 15 nonsense (12%), and 12 splicing (9%). There were biallelic *ABCC6* mutations in 301 cases, 98 of which had associated potential modifier gene mutations that may be contributing to disease severity. Moreover, 36 cases had monoallelic *ABCC6* mutations without candidate for digenic inheritance and 30 cases had monoallelic *ABCC6* mutations with candidate for digenic inheritance. No variants in *ABCC6* were found in 34 cases. Therefore, our total detection rate for *ABCC6* was 91.5%. Of the 401 confirmed cases, 281 had complete Phenodex scores available characterizing phenotypic severity. Preliminary analysis has not established clear-cut genotype-phenotype correlations. In conclusion, additional *ABCC6* mutations were identified that expand the existing database with implications for genetic counseling, prenatal genetic testing and development of allele specific personalized treatments for PXE. The identification of modifier genes for PXE will help to explain clinical variability beyond allelic mutations in the known primary disease-causing gene *ABCC6* and allow new pathways in ectopic mineralization to be discovered.

2891F

Using the SNP array evaluation tool to optimize the use of resources and guide genetic testing leading to the diagnosis of Netherton Syndrome-A case report. P. Gupta, J. Valdez-Camacho, L. Fried, L. Tambini-King, A. Gomez. St Joseph's Children's Hospital, Paterson, NJ.

Introduction: Netherton syndrome is a rare autosomal recessive skin disorder characterized by congenital erythroderma, hair-shaft abnormality and atopic manifestations with high IgE levels. Some infants with Netherton syndrome develop progressive hypernatremic dehydration, failure to thrive, and enteropathy which can be fatal. The disorder is associated with mutations in the *SPINK5* gene, which encodes the serine protease inhibitor LETK1. Diagnosis would often be delayed till the appearance of the characteristic hair-shaft finding of Trichorrhexis Invaginata or bamboo hair. In the era of molecular genetics, sequencing of the *SPINK5* gene can confirm the diagnosis of Netherton syndrome prior to the appearance of characteristic clinical findings. **Case Presentation:** A full-term male infant was born to a 22 year old G2P1 mother at 37 weeks via vacuum assisted C-section for transverse lie, maternal fever and fetal tachycardia. Thick meconium stained fluid was noted. BW was 7lbs 5oz and length 19 inches. Placenta was positive for candida. Patient presented with generalized epidermal desquamation, underlying erythema, alopecia, and foul-smelling skin. Family history was significant for consanguinity. Genetics consult was requested due to above findings and failure to thrive. Chromosome analysis with reflex to SNP microarray revealed a normal male chromosome complement and multiple regions of homozygosity. The patient subsequently had UTI with *E. cloaca*, two episodes of *Staphylococcus aureus* bacteremia, *Influenza*, and *Para-Influenza* infections suggestive of a primary immunodeficiency. He also had significantly elevated IgE, IgA and eosinophils. We used The Genomic Oligoarray and SNP array evaluation tool to look for autosomal recessive genes in the regions of homozygosity using the clinical features of our patient and *SPINK5* was one of the candidate genes. Sequencing of *SPINK5* revealed a homozygous variant, resulting in premature protein termination reported as causative for Netherton syndrome. **Discussion:** Erythroderma is not an uncommon presentation of primary immunodeficiency in early infancy with a broad differential where genetic confirmation can be an expensive process. In the appropriate clinical setting the SNP array evaluation tool can help predict the candidate gene thereby supporting single gene sequencing which is less expensive than large panel testing or whole exome sequencing.

2892W

A unified PCR/CE carrier screening workflow for CFTR, SMN1, and FMR1 variants that consolidates the detection of SNVs, Indels, CNVs, and triplet repeat expansions. L. Chen, S. Gokul, E. Hallmark, B. Haynes, G. Latham. Asuragen Inc, 2150 Woodward St, Austin, TX. 78744.

Introduction: About 5% of Western populations are genetic carriers for one of three disorders: Fragile X Syndrome (FXS), Spinal Muscular Atrophy (SMA), or Cystic Fibrosis (CF). Reliable identification of the underlying pathogenic elements for these disorders require that molecular assays accommodate multiple classes of sequence variants, namely CGG repeat expansion in *FMR1* (FXS), CNVs in *SMN1* (SMA), and SNVs and Indels in CF (*CFTR*). These variants are each technically challenging to resolve, and current screening assays often rely on distinct technologies, platforms and workflows. We describe a simple and accessible multiplex PCR approach that can report genetic variants in *FMR1*, *SMN1*, and *CFTR* by capillary electrophoresis (CE) in a single workflow. **Methods:** Multiplex primers for *FMR1*, *SMN1*, and *CFTR* PCR assays were derived using a novel pipeline and designs that allowed detection of proximal variants. AmpliX® PCR/CE *FMR1** reagents and prototype *SMN1* reagents (Asuragen) were used to genotype the corresponding targets. A prototype PCR/CE assay was developed to detect and discriminate multiple actionable SNV and Indels in the *CFTR* gene. All PCR assays were designed with dye- and mobility-tags to permit co-detection of >100 variants across the three genes using a single CE injection on a 3500 Genetic Analyzer (Thermo Fisher). Analyses were performed using AmpliX® Reporter software with assay-specific plug-in modules (Asuragen). **Results:** We developed a variation of multiplex allele-specific PCR (AS-PCR) to resolve different disease-associated SNVs and Indels in *CFTR*. In preliminary studies, different classes of these variants in heterozygous samples could be readily discriminated in a single-tube PCR, including multiple mutations clustered in complex regions that can incapacitate standard AS-PCR. Further, *CFTR* amplicons could be combined with *FMR1* and *SMN1* amplicons to resolve multifarious combinations of SNVs, Indels, CNVs (0-3 *SMN1* copies), and repeat expansions (*FMR1* premutations and full mutations) in a single PCR/CE workflow. **Conclusions:** Our results demonstrate feasibility for a streamlined, single-platform, PCR-based assay that can unify the detection of *FMR1*, *SMN1*, and *CFTR* variants. This technology has the potential to address carrier screening in decentralized laboratory settings using standardized reagents, existing instrumentation, and fit-for-purpose analysis software. *Research Use Only. Not intended for diagnostic use.

2893T

Utility and diagnostic rates of genetic testing for ataxia-related disorders in patients over 50. *M. Stosser, J. Scuffins, F. Millan, A. Lindy.* GeneDx, Gaithersburg, MD.

Identifying the etiology of a heterogeneous disorder, such as ataxia, can be challenging, especially for older patients. Although many forms of adult-onset ataxia are caused by repeat expansions in a subset of genes, multiple forms have been associated with other types of pathogenic variants and exome sequencing (ES) allows for the screening of these variants. The objective of this study was to review the diagnostic outcomes of ES for patients with ataxia, focusing on patients >50 years of age at the time of testing and comparing these results with patients <21 years. In a review of 2135 patients with ataxia tested by ES, 16% (330/2135) were >50 when referred for genetic testing compared to 69% (1477/2135) that were <21. Patients >50 had a positive diagnostic rate (PDR) of 11% (35/330), 50% (166/330) had a non-diagnostic finding, and 39% (129/330) were negative. Patients <21 had a higher PDR (35%; 510/1477) and were more often submitted for trio-analysis (concurrent sequencing/analysis of parental samples; 72%; 1063/1477). In contrast, only 5% (16/330) of patients >50 were submitted as a trio. Trio testing was associated with a higher PDR in both groups of patients. For patients >50, 19% (3/16) of trio cases were positive compared to 12% (29/245) of proband-only cases. For patients <21, 36% (386/1063) of trio cases were positive compared to 25% (44/175) of proband-only cases. Patients typically presented with additional features including structural cerebellar abnormalities, neuropathy, extrapyramidal signs, seizures, and/or white matter disease. Clinical features most commonly observed in patients >50 included structural cerebellar abnormalities (37%; 123/330) and neuropathy (23%; 77/330), whereas seizures (30%; 438/1477) and structural cerebellar abnormalities (18%; 270/1477) were common comorbidities for patients <21. Patients >50 with cerebellar abnormalities had a higher PDR (16%; 20/123) than those with neuropathy (8%; 6/77). Although genetic testing for patients >50 years of age yielded a lower PDR compared to patients <21, 11-19% of older patients are expected to receive a positive result, with improved outcomes observed for trio analysis and patients with cerebellar abnormalities. The appreciable PDR for sequence analysis of ataxia-related genes indicates that sequencing is a valuable tool for diagnosis, even in older patients. A molecular diagnosis can aid in prognosis and treatment for affected individuals and risk assessment for families.

2894F

Characterization of recessive myopathy/ataxia syndrome due to *MSTO1* variants. *J. Woods, H. Lee, R. Signer, J. Wan, J. Martinez, C. Palmer, S. Nelson, Undiagnosed Diseases Network Members.* University of California Los Angeles, Los Angeles, CA.

Mitochondrial myopathy and ataxia (MIM: 617675) is a rare disease caused by pathogenic variants in the nuclear gene *MSTO1*. Previous publications reported five patients from four families with an autosomal recessive disease form and four patients from the same family with an autosomal dominant disease form. The University of California Los Angeles Undiagnosed Diseases Network (UCLA UDN) recently identified two additional patients with autosomal recessive *MSTO1* associated disease. Cerebellar hypoplasia and myopathy are phenotypically definitive in individual with autosomal recessive forms of *MSTO1* associated disease. Individuals with autosomal recessive disease were compound heterozygous for rare pathogenic variants in *MSTO1*. Individuals with autosomal dominant disease forms have a less severe phenotype with ataxia and neurocognitive impairment. The gene product of *MSTO1* has not yet been well characterized though it is a cytosolic protein with tubulin and tubulin-like domains that localizes to the outer mitochondrial membrane. *MSTO1* plays a key role in maintaining a functional mitochondrial network. Mitochondrial network functionality is contingent upon the balance between fusion events that repair damaged mitochondria and fission events that redistribute the mitochondria. Previous staining with MitoTracker™ in fibroblasts derived from patients with pathogenic variants in *MSTO1* demonstrated fragmented mitochondrial networks. Previous functional assays with mitochondrial photoactivatable-GFP also demonstrated decreased mitochondrial fusion events. These experimental findings were reproduced utilizing fibroblasts from an individual compound heterozygous for pathogenic variants in *MSTO1* identified by the UCLA UDN. Pathogenic variants in *MSTO1* may also lead to mitochondrial DNA depletion, as mitochondrial fusion events are necessary to repair damaged mitochondrial DNA. Presently the UCLA UDN is analyzing mtDNA to determine if mtDNA depletion is a characteristic of *MSTO1* associated disease. Future studies will also focus upon mechanistically defining the role that *MSTO1* plays in mitochondrial fusion by exploring interactions with other mitochondrial fusion genes.

2895W

Novel HSP candidate gene, *HSP1-V2* regulates neurite outgrowth via JNK/c-Jun activation. J. Min^{1,2}, M. Choi^{1,3}, D. Kim^{1,2}, J. Yoon^{1,2}, J. Lee^{1,2}, J. Lee^{1,2,3}, K. Chung^{1,2,3}, D. Halder¹, S. Jun^{1,3}, H. Cho^{1,2,3}, N. Kim^{1,2,3}. 1) Rare Disease Division, KRIBB, Daejeon, Republic of Korea; 2) Genome Research Center, KRIBB, Daejeon, Republic of Korea; 3) Department of Functional Genomics, Korea University of Science and Technology, Daejeon, Republic of Korea.

Hereditary spastic paraplegia (HSP) are a group of neurodegenerative disorders that are progressive spasticity and weakness of the lower limbs by damaging of the longest neuronal fibers of corticospinal tract. In this study, we demonstrated that HSP1-V2 loci plays an important role as a novel regulator in HSP. Single nucleotide variant of HSP1-V2 (c.883G>T, p.D295Y) was identified from whole exome data of 60 HSP families using bioinformatics analysis. Reduction of HSP1-V2 decreased axon and neurite outgrowth which were recovered with overexpression of HSP1-V2. Overexpression of HSP1-V2 decreased epidermal growth factor receptor (EGFR) via phosphorylation of EGFR and increased nerve growth factor receptor (NGFR) via activation of JNK/c-Jun. The activation of JNK signal cascades was mediated by direct binding between HSP1-V2 and JNK. Intriguingly, the D295Y mutation of HSP1-V2 led to decreases of neurite outgrowth by decrease of interaction between HSP1-V2 and JNK. Overall, these results suggest that HSP1-V2 is a novel causative gene of HSP and mediates between proliferation status and differentiation of neuronal cells during neural development.

2896T

Genetics of epileptic encephalopathy. M.Y. Issa¹, G.M.H Abdel-Salam¹, M.S. Abdel-Hamid², J.G. Gleeson³, M.S Zaki¹. 1) Clinical Genetics Human Genetics and Genome Research Division, National Research Centre, Cairo, Cairo-Egypt; 2) Medical Molecular Genetics Department, Human Genetics and Genome Research Division, National Research Centre, Cairo, Egypt; 3) Laboratory for Pediatric Brain Disease, Howard Hughes Medical Institute, University of California, San Diego, San Diego, CA 92093, USA.

Abstract: Early infantile epileptic encephalopathy is a heterogeneous group of a severe form of epilepsy, which was first reported by Ohtahara et al. (1976). It is characterized by frequent tonic seizures or spasms which starts in early infancy life with a specific EEG finding of suppression-burst patterns, characterized by high-voltage bursts alternating with almost flat suppression phases. The seizures are resistant to control with AEDs. Nearly 75% of "Early infantile Epileptic Encephalopathy" EIEE patients may progress to 'West syndrome,' which is characterized by tonic spasms with clustering, arrest of psychomotor development, and hypsarrhythmia on EEG (Kato et al., 2007). To date 65 causative genes are described as a causative gene for EIEE of which 28 are AR, 29 are AD and 8 causative genes are X-linked after searching OMIM <https://www.omim.org/>. Herein, we describe 46 cases from 30 non-related Egyptian families with EIEE. Whole Exome sequencing (WES) was done for patients with suspected genetic cause of epilepsy. Results: 12 different genes were revealed. The most common genes described were WWOX in 13/46 cases from 10 unrelated families, AARS in 8/46 from 4 unrelated families and PNKP 7/46 from 4 unrelated families. Other genes found are STZ2 in 5/46 patients from 3 families, SLC25A12, SLC25A22, UBA5, AP3B2, ITPA, CACNA1A, ALS35A2 and CDKL5 Conclusion: WES is a crucial tool for genetic diagnosis of EIEE to elucidate the cause and early initiation of treatment, hence proper genetic counseling. Next generation sequencing technologies have increased the speed of gene discovery tremendously. The prevalence of AR disorder is due to high consanguinity among the studied population. References: Kato, M., Saitoh, S., Kamei, A., Shiraishi, H., Ueda, Y., Akasaka, M., Tohyama, J., Akasaka, N., Hayasaka, K. A longer polyalanine expansion mutation in the ARX gene causes early infantile epileptic encephalopathy with suppression-burst pattern (Ohtahara syndrome). Am. J. Hum. Genet. 81: 361-366, 2007. Ohtahara S, Ishida T, Oka E, Yamatogi Y, Inoue H, Karita S, Ohtsuka Y (1976) (On the specific age dependent epileptic syndrome: the early-infantile epileptic encephalopathy with suppression-burst.) No to Hattatsu 8:270-279.

2897F

Hereditary neuropathies and differential diagnoses: Experiences from more than 300 cases of next-generation sequencing based genetic panel diagnosis. K. Eggermann¹, L. Soellner¹, M. Begemann¹, B. Gess², M. Haussler³, J. Weis⁴, I. Kurth¹. 1) Institute of Human Genetics, University Hospital RWTH Aachen, Aachen, Germany; 2) Department of Neurology, University Hospital, RWTH Aachen, Aachen, Germany; 3) Department of Pediatrics, University Hospital, RWTH Aachen, Aachen, Germany; 4) Institute of Neuropathology, University Hospital RWTH Aachen, Aachen, Germany.

Peripheral neuropathies or polyneuropathies (PNP) are a heterogeneous group of disorders of the peripheral motor, sensory and autonomic nerves with an estimated prevalence of 5-8% in the general population and with a higher prevalence among the elderly. PNPs can have a multitude of etiologies and present with overlapping phenotypes; hereditary forms include Charcot-Marie-Tooth disease/hereditary motor sensory neuropathy (CMT or HMSN), hereditary motor sensory and autonomic neuropathies (HSAN or HSN), hereditary motor neuropathies (HMN) and small fiber neuropathies (SFN). Mutations in more than 100 different genes have been described so far and more are yet to be discovered; hereditary neuropathies may follow all different patterns of inheritance. In some instances, mutations in neuropathy-associated genes may also be responsible for other clinical entities such as spastic paraplegia (HSP) or myopathy. Using next generation sequencing (NGS), diagnostics has become easier, more straightforward and less time-consuming in recent years. Although in the majority of cases no causal treatment is yet available, knowing the cause of the disorder provides relief for patients and it allows conclusions regarding possible risks for offspring, siblings or other family members. Learning about the underlying pathology might also provide options for treatment in the near future. Based on our experiences from more than 300 cases of genetic panel diagnostics, we suggest an algorithm for molecular diagnostic work-up and present a summary of the results. As a first step, probes are analyzed for copy number variants of PMP22, followed by GJB1 sequencing in families with a clear X-chromosomal inheritance. Among the mutation-negative probes then analyzed by NGS, approximately 15% showed a pathogenic sequence variant; in approximately the same proportion of patients one or more variants of unknown significance (class III) were detected which prompted us to ask for further clinical (and/or molecular) investigations, segregation analyses or functional analyses. Some unexpected results will be discussed.

2898W

Identification of causative variations in cases with microcephaly using cytogenetic microarray and next generation sequencing. S. Masih, P. Srivastava, K. Mandal, S.R. Phadke. Sanjay Gandhi Post Graduate Institute of Medical Sciences, Lucknow, Uttar Pradesh, India.

Microcephaly is a genetically heterogeneous neurodevelopmental disorder. It may be an isolated trait or in combination with other major birth defects or malformations. Despite of remarkable advances in genetic testing, the molecular diagnosis remains unknown in about 40% of cases of microcephaly. We present the clinical features including associated malformations and neurological abnormalities along with the results of copy number variations (CNVs) identified by microarray and sequence variations identified by whole exome sequencing. Cytogenetic microarray was done using Affymetrix 750K in 34 patients (4 consanguineous and 30 non-consanguineous) and identified pathogenic/likely pathogenic CNVs in 7 cases. Homozygosity mapping was done in 4 consanguineous patients and candidate gene was identified in one case. Exome sequencing was done in 24 patients (11 consanguineous and 13 non-consanguineous). Pathogenic/likely pathogenic sequence variations were identified in 10 patients out of which 5 were born to consanguineous parents. Homozygous variations were identified in CIT, ANKLE2, ORC1, SLC13A5 and CENPF genes. Two were reported previously while 3 are novel. The siblings with likely pathogenic variations in CENPF had severe short stature (height -3.5SD) and microcephaly (head circumference -8SD) without ophthalmological abnormalities as reported in cases of Stromme syndrome. Causative sequence variations were identified in 5 patients born to non-consanguineous parents. One patient had novel heterozygous likely pathogenic sequence variation in CTNBN1 and two other had reported pathogenic sequence variations (heterozygous in MECP2 and hemizygous deletion in PQBP1 gene). One novel likely pathogenic homozygous sequence variation was identified in a patient with similarly affected siblings (non-consanguineous) in DCPS gene. The phenotype of this case was prenatally detected cerebellar hypoplasia and severe congenital microcephaly and convulsions since neonatal period. An interesting finding in one patient showed heterozygous likely pathogenic sequence variations two genes, namely; MCPH1 and ASPM genes. The two variations were inherited from asymptomatic parents (one from each) indicating the possibility of digenic inheritance. These novel findings of double heterozygosity, likely pathogenic variations in CENPF and DCPS genes in cases with novel phenotypes need further functional studies. The detailed phenotypes and data of sequence variants will be presented. .

2899T

ARX gene polyalanine repeat expansions are not common cause of intellectual disability or epilepsy. E. Nikkola, C-Y. Lee, J.D. Vargas, A. Zhong, S. Lau, H. Gao, S.P. Strom. Fulgent Genetics, Temple City, CA, US.

Background: The pathogenic mutations in *ARX* have been shown to cause X-linked West syndrome, Ohtara syndrome, infantile spasms, intellectual disability (ID), and syndromic epilepsy in males. Affected females have also been reported, typically presenting with milder symptoms. The best described pathogenic variants in *ARX* gene are short tandem repeat expansions of polyalanine tracts in exons 2 and 4. As widely used short read NGS methods cannot reliably detect the larger repeat expansions, PCR with Sanger sequencing has been used to evaluate these tracts in our clinical lab. Our aim is to evaluate the clinical utility of this complementary testing. **Methods:** We systematically performed Sanger-based polyalanine tract testing for all clinical cases involving intellectual disability and/or epilepsy. We estimated the detection rates for both polyalanine repeat expansions in exons 2 and 4, and also for other pathogenic variants in *ARX* gene identified by NGS. **Results:** We evaluated total of 2,476 clinical cases; 1,351 were from intellectual disability panels, 593 from epilepsy panels, and 532 from others (autism, Rett Syndrome, etc.). We did not detect any pathogenic polyalanine repeat expansions. For sequence variant analysis, we identified 4 pathogenic variants and 17 variants of unknown significance. When we include both the pathogenic variants and the variants of unknown significance, the detection rates were still small; 0.4 % for ID panel, 1% for epilepsy, and 0.3% to 4.5% for other panels with smaller sample sizes. The highest detection rate was for Rett-Angelman gene panel, suggesting that when the symptoms are well refined, the genetic defect is potentially easier to identify. **Conclusions:** Based on over 2,476 cases analyzed, we conclude that *ARX* polyalanine repeat expansions are not as common cause for ID or Epilepsy as had been originally indicated. As 1433 (58%) of these cases are male, we can rule out allele drop-out as a significant source of false negatives. We recognize that our study population is highly heterogeneous and some of the individuals tested have milder symptoms, therefore assaying *ARX* gene repeat expansions may be warranted when ID presents with infantile epilepsy/movement disorder or brain and genital malformations.

2900F

SMN1 and SMN2 copy number assays universally perform on five different thermocyclers. Z. Kabir, Y. Machida, S. Paquerault, L. Jiang, J. Huuskonen. Canon US Life Sciences, Inc., Rockville, MD.

Introduction Spinal muscular atrophy (SMA), an autosomal recessive disorder, causes a spectrum of muscle weakness and atrophy that affects newborns and young adults. SMA is largely caused by *SMN1* gene deletion, the gene that codes for the survival motor neuron (SMN). Disease severity is modulated by the low-level expression of SMN from the 99% identical *SMN2* gene. Here, we determine the assay performance of the Novalele copy number assays for *SMN1* and *SMN2* on five thermocyclers capable of high-resolution melting (HRM). **Methods** Using cell-line genomic DNA and genomic DNA isolated from blood with four different, commercially available DNA extraction kits, the copy number of each uncharacterized sample was determined using the Novalele copy number assays and confirmed with a comparable method, either multiplex ligation-dependent probe amplification (MLPA) or digital PCR. Using the DNA samples, the Novalele copy number assays were individually evaluated on five different, commercially available HRM-enabled thermocyclers. Reaction chemistry and DNA amount as well as PCR and melt protocols were uniform for all of the thermocyclers. All Novalele copy number assay data were analyzed using the Novalele HRM Analyzer. Both accuracy and repeatability studies were completed. The accuracy testing compared the copy number determined to the known copy number or the result of the comparable method, as applicable. The repeatability study was conducted on each of the five thermocyclers and comprised three replicate plates containing 48 or 96 samples depending on the thermocycler plate format allowed. **Results** Both assays correctly determined the *SMN1* or *SMN2* copy number of all DNA samples as compared to the commercially available technology. The results were 100% accurate for all HRM-enabled thermocyclers and DNA extraction kits. We analyzed 1060 data points for the repeatability study. Four of the thermocyclers had 100% consistent results for all data points. One thermocycler had 95.4% (95% confidence interval: 89.4–100.0%) repeatability. Four of the thermocyclers were able to provide results within one hour; the other thermocycler provided results within 90 minutes due to slower melt conditions. **Conclusions** The Novalele copy number assays can reliably determine copy number using multiple thermocyclers and DNA extraction kits without changing PCR protocols or reaction chemistry. The Novalele copy number assays are for Research Use Only. Not for use in diagnostic procedures.

2901W

A novel partial *de novo* duplication of *JARID2* in a patient sharing phenotypic features with deletion patients. K. Kiiski¹, K. Avela², L. Viitasalo².

1) HUSLAB, Laboratory of Genetics, Helsinki University Hospital, Helsinki, Finland; 2) Department of Clinical Genetics, Helsinki University Hospital, Helsinki, Finland.

We present a novel partial *de novo* duplication of the *JARID2* gene (OMIM*601594) in a patient sharing phenotypic features with previously characterized patients with *JARID2* deletions. The duplication was detected using a 180k whole genome chromosomal microarray showing that the 309 kb duplication in chromosomal region 6p23p22.3 covers at least exons 1-6 of the *JARID2* gene. The MLPA method revealed that exons 1-7 of *JARID2* are duplicated (NM_004973.3). A FISH study was consistent with a tandem duplication but the orientation of the duplication remained unelucidated. The identified duplication is located in a known 6p22 microdeletion syndrome region, however, a microduplication syndrome has not been characterized. According to literature, partial duplications of *JARID2* are rare. In DECIPHER, there are five patients listed with a partial *JARID2* duplication, the size of which varies between 270-2,060 kb. One of the deletions is known to be *de novo* and one inherited from an unaffected parent. In all the cases the clinical significance remains unknown. To date, the identified duplication in *JARID2* is the largest partial duplication to be characterized. Deletions of *JARID2* have been identified in patients with intellectual disability, behavioral abnormalities, gait disturbances and dysmorphic features such as prominent supraorbital ridges, deep set eyes, midface hypoplasia and infraorbital dark circles. It has been suggested that *JARID2* is likely a dosage sensitive gene. Our patient is a 12-year-old boy, the second child to nonconsanguineous Finnish parents. The pregnancy and delivery were uneventful. He was born at term with birth measurements within normal range. He started walking at the age of one year. However, since the first years of life he has had problems with coordination and clumsiness. At 3.5 years of age, he was referred to a phoniatrist due to language delay and unclear speech. He attended a normal curriculum at school but needed extra support since he had problems with behavior and attention. The patient also has difficulties with social contacts and daily activities. He has markedly dark infraorbital circles and slightly prominent supraorbital ridges. Currently, RNA studies of *JARID2* are ongoing in order to identify the sequence as well as expression level of the mutated allele. We suggest that partial duplications of *JARID2* may cause a similar phenotype that is associated with deletions of *JARID2*.

2902T

Detection of copy number variations in patients with neurodevelopmental disorders. A.R. Patel^{1,2}, S.S. Chettiar², D.D. Jhala². 1) B.J. Medical College, Civil Hospital Campus, Ahmedabad, Gujarat, India; 2) GeneXplore Diagnostics and Research Center Pvt. Ltd., Ellis Bridge, Ahmedabad, Gujarat, India; 3) Department of Zoology, Biomedical technology and Human Genetics, Gujarat University, Ahmedabad, Gujarat, India.

Detection of copy number variants (CNVs) by chromosomal microarray analysis (CMA) is now well accepted as genetic diagnostic test for the complex neurodevelopmental disorders like Developmental delay (DD), Intellectual disabilities (ID) and Autism spectrum Disorders (ASD). Chromosomal microarray (CMA) offers high resolution scan of whole genome which, significantly contribute to understand the etiology of these complex neurodevelopment impairments. Considering the detection efficiency and impact on medical management of CMA, it is first-choice of test for in-depth analysis of children with these disorders. In the present study, 110 children (68 Males and 42 Females; Age ranges from: 2yrs. to 17yrs.) diagnosed with DD, ASD or ID were recruited for precise genetic analysis. The microarray platform utilized in this study was specifically designed to increase detection of CNVs in genomic regions of demonstrated relevance to DD/ID/ASD. Array CGH was performed on oligonucleotide microarray platform (G2565CA, Agilent, USA). Data of the array was analyzed using cytogenomics software (Agilent technologies). Results of CMA showed total 36 patients (32.72%) with a total 47 Copy number variants (CNVs) and from these findings 29 (61.70%) represented a known diagnostic CNVs contributing to the clinical representation and 18 represented non-diagnostic CNVs i.e. variants of unknown significance (VOUS). Nine patients (25%) with developmental delay (DD) showed 12 CNVs, Fifteen patients (42%) with Intellectual disabilities (ID) showed 21 CNVs and Twelve patients (33%) with Autism spectrum Disorders (ASD) showed 14 CNVs. Chromosomal imbalances of chromosome 2, 15 and 22 were most common. CNVs were detected more in males than females. The likelihood of an abnormal microarray result increased with the number of clinical abnormalities. CMA has demonstrated great value in clinical assessment of neurodevelopmental disorders. Present study concludes that optimized CMA platform can increase diagnostic yield. Pathogenic findings give insights into the etiology of patients, neurodevelopmental conditions, and in many cases impact medical management. The development of novel and accurate methods to interpret the potential pathogenicity of VOUS will further enable patients and clinicians to identify maximum benefits of genetic testing for clinical care.

2903F

A possible new epilepsy locus: Chromosome 4q28.1q28.3 deletion encompassing *PCDH10* segregating in a family with a history of epilepsy and intellectual disabilities. G.H. Scharer¹, S. Offord², L. Weik², E. Virlee¹, N. Zemlicka¹, P. Aggarwal¹, R. Lorier¹, U. Broeckel^{1,3}. 1) Advanced Genomics Laboratory, Medical College of Wisconsin, Milwaukee, WI, 53226, USA; 2) Genetics Center, Children's Hospital of Wisconsin, Dept of Pediatrics, Milwaukee, WI, 53226, USA; 3) Department of Pediatrics, Medical College of Wisconsin, WI, 53226, USA.

Copy number variations (CNVs) are causally linked to a significant percentage of diagnoses of monogenic epilepsy. The prevalence is even higher in patients presenting with generalized epilepsy and intellectual disability. We present the case of a 6 year old female with a history of early developmental delays (speech, gross motor), absence spells and gait ataxia, who was diagnosed with generalized epilepsy based on EEG at 4 years of age. Brain MRI did not show structural abnormalities and patient responded well to ethosuximide (20mg/kg/day). Family history was significant for one full sibling with seizures, another full sibling with psychiatric disorder and cognitive disability, as well as a maternal half-sibling with seizures. The biological mother is reported with intellectual disabilities; and additional maternal relatives are reported with unspecified seizures indicating a dominant genetic disorder. Initial genetic work-up included normal metabolic screening studies and CMA (CytoScanHD™ microarray), which revealed an approximately 9Mb interstitial deletion at chromosome 4q28.1q28.3 (chr4:127,068,548-136,111,450; hg37). The deletion interval includes 40 genes, 8 of these have entries in OMIM; however, none of them have an obvious association with epilepsy to date. Further investigation of affected family members confirmed the deletion in the mother and one of the affected siblings, additional segregation studies are ongoing; although initial observations do support the CNV as a likely cause of the shared phenotype in this family. Upon further review of the haplo-insufficient genes in the deleted region in this case we focused more attention on the *PCDH10* gene, which encodes a protein of calcium-dependent cell-cell adhesion molecules in the cadherin superfamily. Protocadherin-10 belongs to a subclass of protocadherins (PCDH 8/10/18/19) that share a highly conserved 17-amino acid cytoplasmic motif. Of those *PCDH19* is a well-known disease gene because early infantile epileptic encephalopathy-9 (EIEE9), also known as epilepsy and mental retardation restricted to females (EFMR), is caused by mutation in *PCDH19*. Deletion in *PCDH10* was previously implicated in association studies (homozygosity mapping) as possibly linked to autism spectrum disorders. Further investigation, including screening a large sequencing database of epilepsy patients for point mutations in *PCDH10*, is needed to confirm *PCDH10* as a possible new epilepsy locus.

2904W

Clinical next-generation sequencing: A retrospective study of the diagnostic rate in pediatric epilepsy patients. M. Cadieux-Dion^{1,2}, I. Thiffault^{1,2,3}, S. Smith¹, D. Lyalin¹, R. Caylor¹, E. Farrow^{1,3,4}, C. Saunders^{1,2,3}. 1) Center for Pediatric Genomic Medicine, Children's Mercy Hospital; 2) Department of Pathology and Laboratory Medicine, Children's Mercy Hospital; 3) University of Missouri-Kansas City School of Medicine; 4) Department of Pediatrics, Children's Mercy Hospital.

Epilepsy affects 1-3% of the population and is caused by abnormal electrical discharges in the brain (seizures). Currently, the International League Against Epilepsy recommends characterizing the type and the etiology of the seizures as well as the electroclinical syndrome to establish a diagnosis. Although the genetic basis of monogenic syndromic epilepsy is well known (TSC, NF, Rett syndrome), they represent a small proportion of genetic epilepsies. Considering phenotypic and locus heterogeneity, with seizures presenting alone or with other comorbidities such as intellectual disability, autism spectrum disorder, attention-deficit hyperactivity disorder, molecular diagnosis remains challenging. We conducted a retrospective review of patients with epilepsy who received symptom-driven whole exome or genome sequencing (WES/WGS) at our institution. Phenotypes of patient undergoing WES or WGS are recorded as HPO terms in an internal database, which was interrogated for the parent term "seizures." A total of 221 probands (WES n=213, WGS n=8) were identified, including 127 parent-child trios, 46 parent-proband duos, and 48 singletons. The majority of the cohort (76%, 167/221) had comorbidities, with epilepsy in addition to congenital anomalies and/or neurological features. A diagnostic genotype was identified in 19% (42/221), including dominant (32), recessive (7) and X-linked (3) variants. Among cases with a dominant inheritance, 53% (17/32) were explained by a *de novo* variant. The majority of those diagnosed had complex phenotypes (76%, 32/42) compared to seizures only. The most common diagnoses were established based on pathogenic variants in *SCN1A* (3), *CHD2* (2), *PRRT2* (2) and *SYNGAP1* (2). Although pathogenic variants in genes encoding ion channels are well known causes of epilepsy, they are underrepresented in our diagnostic cohort (16%, 7/42). We hypothesize that a significant number of variants of unknown clinical significance would be diagnostic if proven *de novo* since 42.5% of analyses lacked one or both parental samples. In addition, novel gene-disease associations for epilepsy will increase future diagnostic rates. Our data emphasize the need of gene curation, phenotyping and data sharing in pediatric disorders to increase the diagnostic rate of NGS in individuals with epilepsy. .

2905T

Utility of comprehensive next-generation sequencing panel for neuromuscular diseases in Korean patients. *D. Jang¹, J.S. Ryu².* 1) Department of Rehabilitation, Incheon St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea; 2) Department of Rehabilitation Medicine, Seoul National University College of Medicine, Seoul National University Bundang Hospital, Seongnam, Republic of Korea.

Background: Neuromuscular disease is a very broad term that encompasses many diseases that affect the neuromuscular system. Genetic and clinical heterogeneity, unspecific clinical features, unidentified genes and the implication of large genes give challenges in routine molecular diagnosis, increasing turn-around time and effort to make molecular validation of the diagnosis. Recently, multigene disease-specific panels have been emerging as a molecular diagnostic tool, which is a part of next generation sequencing (NGS) technique.

The purpose of the present study was to assess the utility of a comprehensive neuromuscular diseases NGS panel. Methods: We designed a NGS panel targeting 293 genes (version 1) and 410 genes (version 2) associated with comprehensive hereditary neuromuscular diseases. A total 91 patients were enrolled. All patients were analyzed by a targeted NGS panel of comprehensive hereditary neuromuscular diseases (Version 1 was used from June 2016 to September 2017, and version 2 was used from October 2017 to May 2018), chromosomal microarray and karyotyping. Results: A genetic diagnosis was possible in 36 of 91 patients (39.5%). Thirty-four patients were diagnosed through the comprehensive neuromuscular diseases NGS panel and two were confirmed by chromosome microarray test. Four of 55 patients who were not diagnosed were performed additionally a whole exome sequencing test and one of them was diagnosed. A diagnostic yield of version 2 was higher than version 1. Total 37 definitive causative variants and 10 possible causative variants were identified, of which 17 were novel variants. Conclusions: Comprehensive neuromuscular diseases NGS panel can improve the genetic diagnosis efficiency. Due to the rapid discovery of new disease causing genes, an update of gene panel is required.

2906F

Novel compound heterozygous *EPG5* mutations consisted with a missense mutation and a microduplication in the exon 1 region identified in a Japanese patient with Vici syndrome. *S. Shimada¹, H. Hirasawa¹, T. Takeshita¹, N. Nakatsukasa¹, S. Shimojima^{2,3}, N. Nagata¹, Y. Yamamoto^{2,3}.* 1) Tokyo Women's Medical University, Department of Pediatrics, Tokyo, Japan; 2) Tokyo Women's Medical University, Institute of Medical Genetics, Tokyo, Japan; 3) Institute of Medical Genetics, Tokyo Women's University, Tokyo, Japan.

Vici syndrome is a rare, autosomal recessive, multisystem disorder, characterized by agenesis of the corpus callosum, cataracts, psychomotor delay, cardiomyopathy, hypopigmentation, and recurrent infections. Mutations in the ectopic P-granules autophagy protein 5 homolog gene (*EPG5*), which encodes a key autophagy regulator, are responsible for this syndrome. A 3-year-old Japanese girl manifesting similar symptoms to those found in patients with Vici syndrome showed intractable diarrhea, rather than immunodeficiency. Whole exome sequencing identified only a heterozygous variant in *EPG5*, MN=020964.2(*EPG5*):c.3389A>C(p.His1130Pro), which was inherited from her mother. Sequencing analyses of the *EPG5* mRNA showed only an altered nucleotide "C" at position, c.3389, indicating decreased expression of the wild type allele. Microarray-based comparative genomic hybridization revealed a *de novo* microduplication in the exon 1 region. This was considered the cause of the decreased expression of the wild type allele. In conclusion, we successfully identified novel compound heterozygous mutations in *EPG5* in a patient who was clinically considered to have Vici syndrome.

2907W

Dynamic approach to genetic testing for childhood-onset epilepsy. J. Balciuniene¹, E.T. Dechene¹, G. Akgumus¹, E.J. Romasko¹, K. Cao¹, M. Sarmady^{1,2}, N.B. Spinner^{1,2}, L.K. Conlin^{1,2}, I. Helbig^{3,4}, A. Abou Tayoun^{1,2,5}. 1) Division of Genomic Diagnostics, Department of Pathology and Laboratory Medicine, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 3) Division of Neurology, The Children's Hospital of Philadelphia, Philadelphia, PA; 4) Department of Neurology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 5) Al Jalila Children's Specialty Hospital, Dubai, UAE.

The genetic landscape of epilepsy has been rapidly changing through continuous discovery of new disease genes. This is necessitating more dynamic genetic testing approaches to keep up with an evolving gene repertoire and achieve the highest possible clinical sensitivity. We developed an exome-sequencing based panel for idiopathic childhood epilepsy. The test focuses on an initial proband-only comprehensive analysis for sequence and copy number changes in 100 thoroughly curated epilepsy genes, with additional genetic testing options being offered afterwards. Our cohort includes 151 individuals referred for seizures, with and without additional findings. The panel analysis established molecular diagnoses in 16 probands (11%), while the majority of cases were inconclusive (N=87, 57%) or negative (N=48, 32%). Parental testing was available for 15 of the patients with inconclusive findings and led to molecular diagnoses for 7 of them (46%), increasing the overall diagnostic yield to 15.2%. Ten probands with inconclusive or negative findings were reflexed to whole exome testing, and 4 were diagnostic (40%) in genes not included in our panel. This raised the overall diagnostic yield of the study to 17.9% (27/151). For the 27 positive cases, disease causing genes/loci were: *SCNA1* (4x), *STXBP1* (2x), *del16p11.2* (2x), *IQSEC2* (2x), *ATP1A2*, *ATP1A3*, *CACNA1*, *GABRA1*, *KCNQ2*, *KCNT1*, *PRRT2*, *SCN2A*, *SCN8A*, *del of DEPDC5*, *TPP1*, *PCDH19* and *UBE3A*, which were on the curated panel, and *SMC1A*, *SETBP1*, *NR2F1* and *TRIT1* which were identified by the exome test. Of note, probands with epilepsy onset in infancy (N=44) (between 1 and 12 months of age) had the highest diagnostic rate (36%). For the remaining 124 non-diagnostic cases, analysis was expanded to 13 additional genes implicated in epilepsy since the launch of our panel in 2016. While promising findings were identified in at least 4 cases, supporting variant information is still needed to reclassify these cases. We show that an exome-based panel is an efficient testing option that enables rapid analysis of a pre-selected set of genes while retaining unprecedented flexibility in gene content. We argue epilepsy testing benefits from a dynamic strategy with parental samples readily available for follow up analysis of inconclusive findings, and whole exome trio for the remaining non-diagnostic cases. Periodic reanalysis, to capture information in newly identified genes, is expected to further increase the diagnostic yield.

2908T

Clinical exome testing as an effective diagnostic tool for epilepsy/seizure disorders. F. Xia^{1,2}, L. Meng^{1,2}, W. Bi^{1,2}, P. Liu^{1,2}, X. Wang^{1,2}, R. Xiao^{1,2}, A.A. Braxton^{1,2}, P.A. Ward^{1,2}, W. He², F. Vettrini², J. Dong², H. Cheng^{1,2}, D.M. Muzny², R.A. Gibbs², C.M. Eng^{1,2}, Y. Yang^{1,2}. 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Baylor Genetics Laboratories, Houston, TX; 3) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX.

Background: In the United States, about 3 million adults and 0.5 million children suffer active seizure disorders each year, which account for 1.2% of the population. Clinical whole exome sequencing (WES) has been proven as a powerful tool for evaluation of genetic disorders, however, the effectiveness of WES on heterogeneous neurological disorders, such as epilepsy/seizures, has not been thoroughly evaluated. **Method:** We retrospectively analyzed clinical WES cases consecutively tested in our clinical laboratory. Of those individuals, 3,026 had an indication of active epilepsy or history of seizure disorders. Single nucleotide variants and small insertions/deletions were detected by WES analysis and chromosomal microarrays, cSNP arrays and other techniques were employed to confirm large copy number variation (CNV) and other types of mutations. Variant classifications were based upon modified ACMG guidelines. Clinical correlations were determined by a 3-level review system. **Results:** The overall diagnostic rate was 34.5% (1,056/3,026) within this cohort. In addition, 17% cases carried variants of unknown significance that warrant further study. Among patients younger than 18 years-old, the yield was 35.1% (978/2,785), while 32.5% (78/241) of adults were diagnosed. There are contributing defects in more than 800 genes. Among them, *SCN1A*, *STXBP1*, *MECP2*, *KCNT1*, *CDKL5*, *DYRK1A*, *SMC1A*, *WDR45*, *KCNQ2*, *PURA*, *SCN8A*, *SYNGAP1*, *SCN2A*, *ARID1B*, *PCDH19*, and *TBC1D24* together account for ~24% of the 1,056 diagnosed cases. Ninety genes encoding membrane-bound transporters and channels comprised the largest group by contributing ~25% of the diagnosed cases. More than 500 genes only had contributing variant(s) in a single case and ~6% of diagnoses were made by identification of large CNVs or chromosomal changes. **Conclusion and discussion:** Clinical WES diagnosed an overall ~1/3 of a diversified epilepsy/seizures cohort, and suggested follow-up work for more than 1/2 of the diagnosed cases. Several genetic mechanisms underlie epilepsy pathology, including: 1) Ionic changes; 2) Alteration of neuron transmitters and their receptors; 3) Metabolic dysregulation; 4) Immunity deficiency; 5) Structural developmental defects; and 6) Tumorigenesis – and each can require distinctive therapies. Clinical WES can distinguish these causes and therefore is a cost-effective platform that should be considered for all epilepsy/seizure related disorders.

2909F

Undiagnosed disease program in Korea, one year pilot project: Lessons and learned.

W. Kim¹, S. Kim¹, B. Lim¹, J. Lee², H. Kim¹, J. Ko³, K. Kim¹, S. Choi⁴, H. Kim¹, H. Hee⁴, J. Choi⁵, A. Cho⁶, J. Moon⁷, M. Seong⁸, S. Park⁹, C. K¹⁰, Y. Lee¹⁰, Y. Kim¹¹, J. Kim¹², W. Kim¹², H. Park¹², J. Park¹³, Y. Ahn¹⁴, J. Hwang¹⁵, H. Park¹⁶, Y. Lee¹⁷, M. Choi¹⁷, J. Chae¹. 1) Department of Pediatrics, Pediatric Clinical Neuroscience Center, Seoul National University Children's Hospital, Seoul National University College of Medicine, Seoul Korea; 2) Department of Pediatrics, Department of Genome Medicine and Science, Gachon University Gil Medical Center, Incheon Korea; 3) Department of pediatrics, Division of clinical genetics, Seoul National University college of Medicine, Seoul, Korea; 4) Department of Pediatrics, Seoul National University Bundang Hospital, Gyeonggi-do, Korea; 5) Department of Pediatrics, SMG-SNU Boramae Hospital, Seoul, Korea; 6) Department of Pediatrics, Ewha Womans University College of Medicine, Seoul, Korea; 7) Department of Neurology, Laboratory for Neurotherapeutics, Biomedical Research Institute, Seoul National University Hospital, Seoul, Korea; 8) Department of laboratory medicine, Seoul National University Hospital, Seoul, Korea; 9) Department of Laboratory Medicine and Genetics, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea; 10) Department of Pediatrics, Kyungpook National University Children's Hospital, Daegu, Korea; 11) Department of Pediatrics, Chonnam National University Medical School, Gwangju, Korea; 12) Department of Pediatrics, College of Medicine, Chungbuk National University, Cheongju, Korea; 13) Department of Pediatrics, Division of Intensive Care, Seoul National University College of Medicine, Seoul, Korea; 14) Division of Rare Disease, Korea national Institute of Health, Chongcheongbuk-do, Korea; 15) Division of Metabolic Disease, Center for Biomedical Sciences, National Institute of Health, Chongcheongbuk-do, Korea; 16) Division of Cardiovascular Disease, Korea National Institute of Health, Chongcheongbuk-do, Korea; 17) Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul Korea.

Background: Korean undiagnosed rare disease program (KUDP) was launched on January 2017 as one year pilot project with global increasing interest of undiagnosed patients in rare diseases. We would summarize this project and reinforce the unmet needs in undiagnosed rare diseases. **Results:** The patient enrollment, assessment, and diagnostic process were decided by the clinical expert consortium. Patients followed the diagnostic workflow, after being categorized into four groups: I) insufficient clinical information or lack of standard diagnostic processes, II) clinical diagnosis and its targeted confirmatory test being concluded at initial expert consortium meeting, III) clinical diagnosis made at initial meeting but next generation sequencing being required due to its genetic heterogeneity, and IV) no definite clinical diagnosis with heterogeneous and nonspecific clinical presentations. Ninety-seven patients were enrolled in this project, which included 10 patients with category II, 67 with category III, and 20 with category IV. Of those 85 patients (87.6%) were children (<15 years old), 59 (60.8%) were males, and 79 (81.4%) had neurologic symptom as their main problem. For 1 year, 72 patients completed the diagnostic assessment with clinical, biochemical, pathological and molecular genetic analysis. Of those, 28 patients (28/72 38.8%) were reached the diagnosis. Thirteen patients were diagnosed based on the traditional tests such as biochemical assay, single or targeted genetic analysis, and chromosomal microarray. We performed whole exome sequencing in 53 patients, of which 15 (28.3%, 15/53) were reached the final diagnosis. One new disorder was discovered with international collaboration and 3 other candidate genes were on functional validation or waiting for further research. **Conclusions:** KUDP pilot project through efficient clinical diagnostic workflow showed fair diagnostic yield rate with further possibility of additional diagnosis and new scientific discovery in complex rare diseases. KUDP also showed the unmet need in complex and multisystem disorders and utility of emerging genomic technology and further research in rare conditions, still remained undiagnosed.

2910W

De novo missense variants in the alternative exon 5 of SCN2A are a rare cause of neurodevelopmental disorders with or without seizures.

D.N. Shinde¹, L. Rohena^{2,3}, S. Weatherspoon⁴, K.L. Helbig⁵, C. Antolik¹, D.R. Hamlin², J.M. Berg², C. Schultz¹, Z. Powis¹, S. Tang¹, K. Radtke¹. 1) Ambry Genetics, Aliso Viejo, CA; 2) Department of Pediatrics, Division of Medical Genetics, San Antonio Military Medical Center, San Antonio, TX; 3) Department of Pediatrics, Division of Medical Genetics, University of Texas Health Science Center at San Antonio, San Antonio, TX; 4) Department of Pediatric Neurology, Le Bonheur Children's Hospital, University of Tennessee Health Science Center, Memphis, TN; 5) Division of Neurology, Children's Hospital of Philadelphia, Philadelphia, PA.

SCN2A encodes Na_v1.2, the alpha subunit of the neuronal voltage-gated sodium channel that is responsible for generation and propagation of action potentials. Heterozygous mutations in SCN2A cause a spectrum of neurodevelopmental disorders including benign familial infantile seizures and early infantile epileptic encephalopathy, with or without autism spectrum disorders and intellectual disability. Na_v1.2 has two developmentally regulated isoforms (neonatal [N] and adult [A]) that use alternatively spliced coding exons 5N and 5A, respectively. During early brain development, the N isoform is more abundantly expressed than the A isoform, but its relative proportion decreases as it is replaced by the A isoform during postnatal development. N channels are less excitable than A channels, and mutations in SCN2A that increase neuronal excitability of the N channels to the level of the A channels, are thought to increase susceptibility to seizures in the neonatal period. Although most mutations in SCN2A are found in both the isoforms, to date, there is only one report of a likely pathogenic variant, c.634A>G (p.N212D), found exclusively in the N isoform in an infant with Ohtahara syndrome. Here, we report two additional unrelated affected individuals with *de novo* missense variants in exon 5N of SCN2A, c.647T>A (p.L216H) and c.668G>T (p.R223I), that were identified using trio-based whole exome sequencing. Both the affected individuals presented with global developmental delay and abnormal brain MRI findings. Additionally, infantile spasms were observed in one of the affected individuals, while the other presented with intellectual disability and autistic behaviors. Both altered amino acids are located in the voltage-sensor S4 transmembrane helix of Na_v1.2, and structural analysis indicates that the alterations impact channel function by disrupting charge gating. Functional studies to determine impact on channel activity are ongoing, and identification of additional affected individuals may help to elucidate the phenotypic spectrum of variants in exon 5N of SCN2A. Our results highlight the clinical utility of reporting variants in alternative isoforms of genes with clinically well-characterized primary isoforms.

2911T

Whole exome sequencing and rhPCR accurately resolve copy number variation in highly homologous *SMN1* and *SMN2* genes. J. Sistonen, M. Valori, H. Junnila, A. Korppoo, L. Koskinen, E. Salminen, T.P. Alastalo, S. Myllykangas, J.W. Koskenvuo, P. Salmenperä. Blueprint Genetics, Helsinki, Finland.

Spinal muscular atrophy (SMA) is a neuromuscular disorder characterized by progressive degeneration of spinal cord motor neurons, atrophy of skeletal muscles, and generalized weakness. The clinical phenotype ranges from a severe infantile form with a limited life expectancy to an adult-onset mild form of the disease. In the majority of cases, the disease is caused by a homozygous absence of the survival motor neuron 1 (*SMN1*) gene. Highly homologous *SMN2* gene that differs from *SMN1* only by a single coding nucleotide resulting in a splicing defect and low *SMN2* function can also modify the disease severity. Patients with a higher *SMN2* copy number usually manifest a milder clinical phenotype. To address the clinical importance of accurate and efficient *SMN1* and *SMN2* copy number analysis, we developed a custom bioinformatic analysis based on whole exome sequencing data combined with a novel RNase H2-dependent PCR (rhPCR) for confirmation analysis. The bioinformatic method utilizes next-generation sequencing (NGS) reads at four loci differing between the genes while the rhPCR assays target two of the loci with RNA-modified primers and activating RNase H2 enzyme improving the assay specificity and sensitivity compared to traditional quantitative PCR-based methods. We validated the methods using a set of commercially available samples and patient DNA samples extracted from blood and saliva. Both methods showed 100% sensitivity and specificity. Additionally, we applied the bioinformatic analysis to test 2196 de-identified patient samples and observed heterozygous *SMN1* deletion in 2.6% of the samples, which is in agreement with frequencies reported earlier. In conclusion, we have established an accurate and high-throughput approach to test for *SMN1* and *SMN2* copy numbers enabling diagnostics of SMA and application of novel therapeutic strategies.

2912F

Development, validation and application of next-generation sequencing assay for carrier screening of spinal muscular atrophy. X. Meng, C. Ye. WuXi NextCODE, Shanghai, China.

Spinal muscular atrophy (SMA [MIM 253300]) is caused by biallelic mutations in the survival motor neuron 1 (*SMN1*) gene, whereas *SMN2* copy number may modify disease severity. The identification of SMA carrier status in large populations is necessary for congenital defects prevention. Traditional methods such as ddPCR and MLPA are expensive and low throughput in large scale carrier screening, and incapable of detecting novel variants. Recently, NGS-based carrier screening assays show increased efficacy in CNV detection than conventional methods in a high throughput mode for a large number of genes. We developed a capture-based NGS assay and its analysis tool to detect copy numbers of *SMN1/2*. Our algorithm first quantified the normalized unique reads on exon 7. Then, the unsupervised Gaussian Mixture Model (GMM) was trained for carrier status classification. In total, 2058 samples were used to setup baselines for each copy number mode. For validation, another 1170 samples were used with MLPA/ddPCR confirmations. The sensitivities (and positive predictive values, PPVs) for *SMN1* with 0, 1 and ≥ 2 copy numbers on exon 7 were 100% (94.7%), 98.4% (72.1%) and 97.7% (99.9%), respectively. Furthermore, the sensitivities (and PPVs) for *SMN2* were 100% (97.2%), 99.5% (96.6%) and 98.3% (99.9%), respectively. The specificities for *SMN1* and *SMN2* were 97.7% and 98.3%, respectively. With demonstrated high sensitivity and specificity, the developed NGS assay was applied in front-line carrier screening. To further reduce false positive and false negative in clinical practice, NGS identified high confident positive/carrier samples and low confident results were all confirmed by MLPA. To date, 828 samples were included in the prospective study. According to NGS results: 1) 19 samples were SMA carriers (10 high confident and 9 low confident results); 2) 7 samples were low confident negatives; 3) the rest 802 samples were high confident negatives. According to confirmations by MLPA, 1) the sensitivity is 100%; 2) the PPV is 68.4% (6 low confident "carriers" were later found as negatives); 3) 7 low confident negatives were all identified as negatives. The total SMA carrier frequency as a result is 1.57% (13/828). In conclusion, we have developed, validated and applied an effective patented NGS assay for SMA carrier screening with reduced cost and improved performances. It could be easily incorporated into other extensive screening panels for better congenital defects prevention.

2913W**Clinical approach of Angelman syndrome: A study of 23 cases. I.O.**

Focsa¹, L.C. Bohiltea¹, D. Craiu^{1,2}, C. Iliescu^{1,2}, D. Birca^{1,2}, A.C. Tutulan-Cunita³, S.M. Papuc³, A. Arghir³, M. Budisteanu^{2,3,4}. 1) C.Davila University of Medicine and Pharmacy, Bucharest, Romania; 2) Prof. Dr. A. Obregia Clinical Hospital of Psychiatry, Bucharest; 3) V. Babes National Institute of Pathology, Bucharest; 4) University of Medicine Titu Maiorescu, Bucharest.

Introduction: Angelman syndrome (AS) is a genetic disorder characterized by a severe phenotype including: global developmental delay, a happy disposition and other behavior problems, movement disorder or ataxia and by a recognizable facial dysmorphism. Epileptic seizures, microcephaly and hypo-pigmentation are reported in many patients. Several genetic mechanisms have been associated with AS: deletions of 15q11-q13, paternal uniparental disomy, imprinting defects or mutations in *UBE3A* gene. **Aim:** In this paper we will present our experience regarding the diagnosis and the management of Angelman syndrome. **Material and methods:** Our study included 23 children (7 boys and 16 girls) with a putative diagnosis of AS, of ages ranging from 6 months to 12 years. The clinical diagnosis was established due to the presence of typical features included in consensus criteria (Williams 2006). In order to confirm the diagnosis, we performed genetic tests – FISH, methylation PCR, sequencing. For each patient, a personalized management plan was established. **Results:** The genetic tests certified the diagnosis in all 23 children: in 21 cases, the 15q11-15q13 deletion was identified, whereas a mutation in *UBE3A* gene was found in 2 patients. All patients showed severe intellectual disability, ataxic gait, sleep disturbances and happy disposition. The impairment of language was severe: only 2 patients were able to say a few words. Microcephaly was noted in 18 cases. 22 children had epileptic seizures. The most frequent seizures were partial seizures (12 children), followed by tonic-clonic generalized seizures, hypotonic seizures, and infantile spasms. The EEG exhibited the characteristic AS pattern in all patients. Valproate was the anticonvulsant of choice, alone, or in combination with other antiepileptic drugs. A good control of seizures was obtained in 20 children. Physical therapy and cognitive stimulation were performed in all children, leading to improvement of the motor delay, but with very little progress in cognitive functions. **Conclusions:** Angelman syndrome is a severe disorder, with complex clinical picture, imposing a multidisciplinary approach. An early diagnosis is tremendously important in order to develop the efficiency of the personalized management plan.

2914T**Behind the Seizure™: A no-cost, 125-gene epilepsy panel for pediatric seizure onset between 2 and 4 years. S. Aradhya¹, M. Bailey², J. Cohen-Pfeffer³, S. Koh³, J.J. Millichap⁴, R. Truty⁵, E.C. Wirrell⁵, N. Miller².**

1) Invitae, San Francisco, CA; 2) BioMarin Pharmaceutical Inc., Novato, CA; 3) Emory University School of Medicine, Atlanta, GA; 4) Lurie Children's Hospital, Chicago, IL; 5) Mayo Clinic, Rochester, MN.

Objectives: CLN2 disease, a form of neuronal ceroid lipofuscinosis (NCL), commonly presents ~3 years-of-age with seizures and history of language delay. Average diagnosis occurs ~5 years: two years after symptom onset and after neurodegeneration marked by loss of ambulation and speaking ability. Our objective is to understand the molecular diagnostic yield of a US-based, targeted, no-cost epilepsy gene panel program, Behind the Seizure™ (BTS), and determine if this can decrease the age of diagnosis in CLN2 disease.

Methods: BTS eligibility criteria: (I) age at test order ≥24 months to <60 months, (II) unprovoked seizure onset at ≥24 months. Physician-reported clinical history and suspicion of seizure etiology was collected. Average test turnaround time is 10-14 days (Invitae Epilepsy Panel). **Results:** Non-BTS epilepsy gene panel testing (all ages, age of seizure onset) resulted in a 17.1% molecular diagnostic yield (n=277/1620) and 0.12% for *TPP1* (CLN2 disease). For BTS testing, the yield for CLN2 disease was 2.3% (n=4/176), a 19-fold improvement. Age-at-diagnosis of CLN2 disease was earlier than average (3-4 years versus 5 years). From BTS, 3 of 4 molecular diagnoses in *TPP1*, CLN2 disease was unsuspected. Other genes with molecular diagnoses: *MECP2*, *CHD2*, *SYNGAP1*, *SCN1A*, *GPHN*, *GRIN2A*. **Conclusions:** Behind the Seizure™ facilitated CLN2 disease identification earlier than average and in cases which CLN2 was not suspected. While there is no direct comparator for the BTS group, these findings may indicate that broad epilepsy gene panel tests implemented with clinical criteria can increase diagnostic yield for CLN2 disease and simultaneously identify other genetic causes of epilepsy. .

2915F

Molecular basis of CLN2 disease: A review and classification of *TPP1* gene variants reported world wide. N. Miller¹, E. Gardner², A. Schulz³, M. Aristonera⁴, M. Bailey¹, W.W. Xin⁴, S.E. Mole². 1) BioMarin Pharmaceutical Inc, Novato, CA; 2) University College London, London, UK; 3) University Medical Center Hamburg-Eppendorf, Hamburg, Germany; 4) Massachusetts General Hospital, Boston, MA, USA.

BACKGROUND: CLN2 disease is caused by autosomal recessive inheritance of 2 pathogenic variants *in trans* in the *TPP1* gene, leading to deficient activity of the lysosomal enzyme tripeptidyl peptidase I (TPP1). A subtype of neuronal ceroid lipofuscinosis, CLN2 disease is diagnosed through clinical findings, TPP1 enzyme deficiency, and/or molecular findings in *TPP1*. CLN2 disease classically presents with seizures at 2-4 years of age and a history of early language delay, followed by rapid psychomotor decline and death in the second decade. A form of juvenile onset disease was initially described as autosomal recessive spinocerebellar ataxia 7 (SCAR7) before being attributed to TPP1 enzyme deficiency. **METHODS:** We collected individuals reported to have TPP1 enzyme deficiency from literature review, public databases and laboratory communication. Previously unrecorded individuals collected were added to the University College London *TPP1*-specific database. **RESULTS:** We present an update on the spectrum of *TPP1* gene variants associated with CLN2 disease including 178 unique variants and 405 individuals reported to have TPP1 enzyme deficiency (739 alleles). Two known pathogenic variants, c.509-1G>C and c.622C>T (p.Arg208Ter), collectively occur in 54% of affected individuals collected; at least 91 variants are private to single families. Homozygosity occurs in 43% of individuals where both alleles are known (84% of reported individuals). Atypical CLN2 disease represents 17% of individuals recorded with associated phenotype. Increased use and early adoption of diagnostic testing for a genetic basis of epilepsy is aided by description of *TPP1* variants in association with clinical and laboratory findings. NCBI ClinVar currently holds records for only 23% of variants collected from literature. **DISCUSSION:** Effective CLN2 disease management requires timely diagnosis; however, irreversible neurodegeneration occurs before diagnosis is typically reached at age 5. Timely classification and public reporting of *TPP1* variants is essential as molecular testing increases in use as a first-line diagnostic test for pediatric-onset neurological disease.

2916W

Searching for blood biomarkers to improve the diagnosis and the management of patients with epilepsy. M. Martin¹, S.H. Avansini¹, A.S. Vieira¹, R. Secolin¹, M.L. Santos¹, F.R. Torres¹, F. Rogério², A.C. Coan², M.K.M. Alvim², M.E. Morita², C.L. Yasuda², R. Barbosa², F. Cendes², I. Lopes-Cendes¹. 1) Medical Genetics Dept., University of Campinas (UNICAMP), Campinas, Brazil; 2) Neurology and Anatomical Pathology School of Medical Sciences, University of Campinas (UNICAMP), Campinas, Brazil.

Introduction and Hypothesis: It is estimated that the misdiagnosis of epilepsy occurs in 25% of patients and 1/3 of these do not respond to clinical treatment. Thus, the search for biomarkers for the diagnosis and the prediction of response to antiepileptic drugs (AEDs) is urgently needed. Potential candidates are circulating microRNAs, present and stable in body fluids and are already associated with the diagnosis and staging of various diseases. **Objective:** This study aims to determine whether robust and non-invasive molecular signatures of circulating microRNAs could help to improve diagnosis and or management of patients with epilepsy, and to identify and validate whether these molecular signatures could also be associated with response to AEDs. **Methods:** The study is divided into two phases: an initial discovery phase with 10 patients with mesial temporal lobe epilepsy (MTLE) who are responsive to AED treatment, 10 patients with MTLE who are AED resistant, 10 patients with focal cortical dysplasia (FCD), 10 patients with genetic generalized epilepsy and 10 control individuals. These will be used to quantify plasma levels of microRNAs by microRNA-sequencing. In the second phase, we enrolled an additional independent cohort of 100 patients with MTLE using the same diagnostic criteria as previously described and we will assay the most significant microRNAs found in phase 1 using RT-qPCR. We also recruited an additional 200 healthy individuals without epilepsy as a control group. **Results/Discussion:** To date, we have completed the recruitment of patients for both phases of the study and the main phenotypic characteristics of our cohort are as follows. We found that patients with MTLE who are AED responsive have fewer seizures than patients who are AED resistant (*p-value* = 0.0017). Patients with MTLE are older than patients with FCD (*p-value* < 0.05). Patients with AED resistant MTLE have seizure onset earlier (*p* < 0.05) and use a high number of AEDs than patients with AED responsive MTLE (*p* = 0.0029). **Conclusion:** We have successfully recruited and characterized clinically a cohort of patients who are being currently studied for the discovery of circulating biomarkers for epilepsy and response to treatment. When completed our study will likely to have a relevant impact in the way clinicians manage patients with different types of epilepsy. **Supported by:** FAPESP, SP, Brazil.

2917T

PIK3CA mutations in fibrolipomatous hamartomas of nerve. *P.R. Blackburn¹, D. Milosevic¹, B.M. Howe², A.L. Folpe¹, R.J. Spinner³, J.M. Carter¹.* 1) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN; 2) Department of Radiology, Mayo Clinic, Rochester, MN; 3) Department of Neurologic Surgery, Mayo Clinic, Rochester, MN.

Fibrolipomatous hamartoma (FLH)/lipomatosis of nerve is a rare developmental malformation characterized by an aberrant lipomatous proliferation within peripheral nerve fascicles. FLH most frequently involves the median nerve, and can manifest clinically as a peripheral neuropathy. However, a significant subset of FLH is associated with proliferation of bone and/or soft tissue in the affected nerve's territory, manifesting clinically as tissue overgrowth. The molecular genetic profile of FLH, including anatomically and phenotypically-distinct variants, and its association with tissue overgrowth disorders, is not well understood. Seven FLH (from 4 females and 3 males, median age at presentation: 10 years; range: 0-33 years) involving the median (N=4), sciatic, ulnar, or peroneal nerves (N=1 each) were analyzed by whole-exome sequencing. Five patients had FLH nerve territory overgrowth (N=3 median nerve, N=1 each for sciatic and ulnar nerves). *PIK3CA* missense mutations were identified in 6 of 7 cases. The two FLH without clinically apparent overgrowth (peroneal nerve FLH and one median nerve FLH) both had the c.3140A>T, p.H1047L variant (mutant allele frequency range: 6 to 17%). Four of 5 FLH with overgrowth had the more common c.3140A>G, p.H1047R variant (mutant allele frequency range: 14 to 26%). No detectable *PIK3CA* mutations were found in the sole ulnar FLH. The *PIK3CA*-related overgrowth spectrum of disorders is characterized by post-zygotic *PIK3CA* mutations with tissue mosaicism. Within this spectrum, somatic missense mutations at the H1047 residue in the kinase domain of *PIK3CA* are frequently seen in fibroadipose overgrowth, hemihyperplasia with multiple lipomatosis, and isolated macrodactyly, with or without median nerve FLH. Our data demonstrate that *PIK3CA* p.H1047 mutations are frequent events in FLH, irrespective of anatomic site or overgrowth phenotype. While the small case number limits genotypic-phenotypic correlation, both mutant allele frequency and *PIK3CA* mutational subtype may determine the extent and distribution of tissue overgrowth. Regardless, as FLH are defined by aberrant lipomatous overgrowth within nerve, these lesions likely belong to the *PIK3CA*-related overgrowth spectrum. .

2918F

Overcoming challenges in validation of potentially mosaic variants. *C.M. Grochowski¹, R.T. Gonçalves^{2,3,4}, Z.H.C. Akdemir¹, S.R. Lalani¹, J.R. Lupski^{1,5,6,7}, C.M.B. Carvalho¹.* 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Universidade Federal de Minas Gerais (UFMG); 3) Sociedade Brasileira de Genética Médica (SBGM); 4) Faculdade da Saúde e Ecologia Humana (FASEH); 5) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 6) Department of Pediatrics, Baylor College of Medicine, Houston, TX; 7) Texas Children's Hospital, Houston, TX.

Increasingly sensitive sequencing methodologies have led to the discovery of variants present in a mosaic state. The identification of mosaic variants has a profound impact not only on genetic counseling for recurrence risk when found in a parent, but also on patient care as mutation mosaicism can have an effect on disease burden and outcome. As mosaic mutations are found in potentially 1% of exomes analyzed, orthogonal validation methods are needed. We present two cases in which ddPCR was used to validate a mosaic variant discovered in WES. The first case involving an individual with Wolf-Parkinson-White Syndrome (MIM 194200) and a missense variant in the gene *RHBDF1* which presented with a low variant read to total read ratio (32/217 14.7%) leading to the suspicion of mosaicism. A TaqMan assay was designed with two different fluorescent probes that would bind with either the mutant or reference allele. The proband showed a ratio of positive droplets for the mutant allele versus wildtype allele of 13.3% (2415/18105), corroborating the low variant count originally found in the WES data. A second case involved a diagnosis of epileptic encephalopathy (MIM 607208) and a missense variant in *SCN1A* in which the proband showed a variant read to total count of 39% (47/119) and the father (who reported epilepsy as a child that resolved with age) presented with a low variant read to total ratio of 12.7% (15/118), again raising suspicion of mosaicism. A TaqMan assay was designed but the mutant and wildtype both amplified the same region due to the length of the probe (24bp). As probes >18bp are better at discriminating these changes, we sought to lower the length through the use of LNA bases, though this approach also failed. A third attempt was performed using allele specific primers and EvaGreen. Forward primers were designed with a single bp difference in the 3' end denoting the G>A change with a destabilizing mismatch placed in the penultimate nucleotide. Using a common reverse primer, the ratio of positive droplets for the mutant allele versus wildtype was 42% (1871/4378) in the proband and 17% (1083/6217) in the father, closely matching the WES call. As NGS becomes more widely utilized, the discovery of variants in a mosaic state will increase. Overcoming the challenges to validate these variants is critical in providing parents the true nature of their recurrence risk and providing patients with a deeper understanding and subsequent care for their disease.

2919W

Low-level parental mosaicism is a characteristic of holoprosencephaly pathogenic variants previously described as *de novo*. A.F. Martinez, P. Hu, P.S. Kruszka, E. Roessler, M. Muenke. National Institutes of Health, Bethesda, MD.

Holoprosencephaly (HPE) is the most common forebrain malformation of the developing embryo. We developed a novel high-coverage targeted capture panel of 150 potential HPE genes to determine the inheritance pattern of disease-causing mutations and selected *de novo* variants (DNV) for analysis of possible parental mosaicism. We identified 28 high-confidence DNVs that were confirmed by family studies, exome data and droplet digital polymerase chain reaction. Most of the DNVs were identified among four known HPE genes (20/28) and were consistently pathogenic (95%) and exclusively exonic. Randomly distributed singleton DNVs of uncertain significance constituted the remaining detections. Analysis of the sequence alignment (BAM) files revealed variable degrees of parental somato-gonadal mosaicism in 26% of cases. In addition, evidence for likely post-zygotic *de novo* events in probands were exclusively detected in 4/8 (50%) of targets (one in an HPE gene and three without previous HPE association). Our results indicate that high-coverage sequencing is essential to providing an accurate assessment of recurrence risk of HPE in mutation-positive families. Detection and confirmation of low-level mosaicism needs to be offered to all at-risk families.

2920T

Detection of spinal muscular atrophy-associated loss of *SMN1* in a clinical whole genome sequencing test. A.L. Halpern¹, A. Chawla¹, S.S. Ajay¹, R.T. Hagelstrom¹, C.M. Brown¹, A. Buchanan¹, A.J. Connell², M.E.R. Butchbach², J. McEachern¹, D.L. Perry¹, M.A. Eberle¹, R.J. Taft¹. 1) Illumina, Inc, San Diego, CA; 2) Nemours Biomedical Research, Nemours/Alfred I. duPont Hospital for Children, Wilmington, DE.

Spinal Muscular Atrophy (SMA) is an early-onset neurogenetic disorder caused by loss of motor neurons in the spinal cord and brain stem, leading to generalized muscle weakness, atrophy and death. SMA is caused by biallelic loss of function of the *SMN1* (*survival motor neuron 1*) gene. Approximately 95% of affected individuals have a complete loss of *SMN1* with the remaining 5% of affected individuals having other pathogenic *SMN1* variants on the same or, rarely, both haplotypes. The *SMN1* gene has a paralogous gene known as *SMN2*, whose copy number can modify disease severity. *SMN1* and *SMN2* are nearly identical except for a few nucleotide differences, the most notable being c.840C>T (transcript: NM_000344.3). This nucleotide transition within exon 7 causes the vast majority of *SMN2*-derived transcripts to lack exon 7 and, as a result, produce a truncated, unstable protein. Variants in *SMN1* and copy numbers of *SMN1* and *SMN2* are routinely assessed by polymerase chain reaction (PCR) coupled with restriction fragment length polymorphism analysis, and quantitative PCR, multiplex ligation-dependent probe amplification or digital droplet PCR, respectively. These techniques are often performed in addition to other molecular investigations, adding to cost and complexity of clinical care. The high homology between *SMN1* and *SMN2* creates challenges for standard variant calling pipelines, which typically produce poor quality variant calls in most of *SMN1*, including the c.840C position. To address this, we have developed a SMA caller that reports allele depths for both 'C'- and 'T'-supporting reads derived from either *SMN1* or *SMN2*. A finding of SMA-positive is reported if the read support for c.840C is consistent with no functional copies of *SMN1*. The SMA caller was tested on 40 samples with known SMA status. Sensitivity to accurately detect the c.840C>T variant in the samples was assessed to be 95.83%, with a specificity of 100%. Future aims include detection of other variants in *SMN1*, estimating paralog-specific copy number and identification of gene conversion between *SMN1* and *SMN2*.

2921F

Targeted re-sequencing system in inherited muscular diseases. A. Iida¹, M. Okubo², M. Inoue², A. Ishiyama², Y. Saito², I. Nonaka², S. Hayashi², S. Noguchi², I. Nishino^{1,2}. 1) Medical Genome Center, National Center of Neurology and Psychiatry, Kodaira, Tokyo, Japan; 2) National Institute of Neuroscience, National Center of Neurology and Psychiatry, Kodaira, Tokyo, Japan.

Inherited skeletal muscle diseases are clinically and genetically heterogeneous rare disorders characterized by muscle weakness and degeneration. However, the pathogenesis remains largely unknown. National Center of Neurology and Psychiatry (NCNP) is a referral center for neuromuscular diseases in Japan. Since 1978, we have served the diagnosis service for muscle diseases all over Japan. Now, more than 70% of biopsied samples were gathered in NCNP. We have screened approximately 18,000 biopsied samples based on clinical and pathological information. Recently, we also established the molecular diagnostic system with four targeted gene panels containing 180 muscular genes to diagnose the etiology of muscle diseases, including muscular dystrophy (MD), congenital myopathy/congenital myasthenic syndrome (CMP), myopathy with protein aggregations/rimmed vacuoles (MFM), and metabolic myopathy (MM). To elucidate the patients with muscle diseases, we performed mutation analysis in 1,567 patients by Ion PGM sequencer coupled with the targeted gene panels. The diagnostic yield of MD panel was 55%, CMP panel was 36%, MFM panel was 39% and MM panel was 56%. Total diagnostic yield was 46%. Furthermore, novel mutations in five causative genes in five muscular diseases are detected; *VCP* in multisystem proteinopathy, *DYSF* in MD, limb-girdle type 2B, *LMNA* in CMP, *POMGNT2* in dystroglycanopathy, and *MTM1* in X-linked myotubular myopathy. The mutations in inherited muscular diseases will serve as an additional resource for analyzing correlation between mutations and phenotypes and future clinical sequencing.

2922W

A diagnostic odyssey: Multiple platforms fail to detect this *MECP2* deletion. S. Ceulemans¹, M.C. Jones^{1,2}, R. Haas^{1,2}. 1) Rady Children's Hospital, San Diego, CA; 2) University of California - San Diego, San Diego, CA.

We describe a Rett syndrome case that highlights the difficulty caused by differences in laboratory assays resulting in failure to identify intragenic copy number variants (CNVs). Our patient is a four year old female who was noted to have reduced fetal movement but otherwise had an uncomplicated pregnancy and delivery. She had feeding difficulties but met milestones until age six months, when she was hospitalized for silent aspiration identified during a swallow study. Clinical exam showed gross hypotonia, obstructive sleep apnea, with mild dysmorphic features. She developed seizures around 15 months, which have proven to be intractable. The EEG showed abundant generalized slow spike and wave complexes in awake and sleep states, indicating severe epileptic encephalopathy. In the course of her thorough and complicated work-up, she underwent muscle and skin biopsies, lumbar punctures, biochemical testing, and six molecular genetic tests at five laboratories, one of which identified a deletion of 345 base-pairs within the fourth exon of *MECP2*. Table 1 outlines each result and whether the inability to detect the deletion was due to a technical limitation or a lab error. CNVs of approximately one to two exons are difficult to identify using NGS or oligo-SNP microarray, and single-exon array appears to have better yield. Laboratory differences impact the length of the diagnostic odyssey, possible treatment interventions and prognosis, and the relationship between patient and care team.

Table 1		
Genetic Test	Results	What happened?
Lab A - Oligo-SNP microarray	Duplication of PMP22, consistent with family history	Technical limitation: No probes within exon 4, CNV not big enough
Lab B - "Clinical exome" with NGS, revised to include "100% coverage of <i>MECP2</i> "	Carrier for Gaucher, 7 VUSs	Technical limitation: CNVs between 25bp-2 exons not consistently detected
Lab C - Research exome trio, NGS	Negative	Technical limitation: not designed for CNVs
Lab A - <i>MECP2</i> del/dup, qPCR	Negative	MISS, semi-quantitative approach
Lab D - Research genome trio, NGS	Suggestive of <i>MECP2</i> deletion	Needed confirmation
Lab E - <i>MECP2</i> del/dup with exon-level array, confirmed with MLPA	Partial deletion of <i>MECP2</i> exon 4	"Full coverage" is >3 probes, and 7 probes were on exon 4 alone

2923T

Bi-allelic *ATAD3A* and *ATAD3B* deletions in two siblings with multiple congenital defects including severe pontocerebellar hypoplasia, resulting in neonatal death. *F.A.T. de Vries¹, L.S. Smit², A. Go³, M.M. Veghel Van Plandsioen¹, C.F.B. Labrijn – Marks¹, M. Wilke¹, G.M.S. Mancini¹.* 1) Dept. Clinical Genetics, Erasmus Medical Centre, Rotterdam, The Netherlands; 2) Dept. Neurology, division Pediatric Neurology, and Dept. Pediatrics, division Neonatology, Erasmus Medical Centre, Rotterdam, The Netherlands; 3) Dept. Gynecology and Obstetrics, Erasmus Medical Centre, Rotterdam, The Netherlands.

In the second pregnancy of an healthy female, ultrasonography revealed polyhydramnion and talipes equinovarus. A male was born at 33 5/7 weeks gestational age and died 5 days after birth due to respiratory insufficiency. Postnatal examination showed severe pontocerebellar hypoplasia, optic atrophy, macula dystrophy, corneal oedema and congenital arthrogryposis. Routine SNP array analysis and whole exome sequencing (panel for intellectual disability, CRE-exome) did not reveal any pathogenic abnormalities. Triggered by described distinct clinical findings in patients with bi-allelic *ATAD3* gene cluster mutations^{*}, additional SNP array analysis and a region specific long range PCR[†] were performed. A ~36 kb homozygous deletion (minimal region: chr1:g.1415790_1452101 in hg19) including exon 5-16 of the *ATAD3B* gene and exon 1-2 of the *ATAD3A* gene was detected. At the time of the genetic confirmation, the woman was pregnant again and advanced sonography revealed bilateral talipes equinovarus, arthrogryposis, lissencephaly and pontocerebellar hypoplasia. Prenatal testing of the *ATAD3* gene cluster in amniotic fluid showed a similar ~36 kb homozygous deletion of exon 5-16 of *ATAD3B* and exon 1-2 of *ATAD3A*. A girl was born at 38 weeks of gestation and died within 1 hour after birth due to respiratory insufficiency. The above described techniques were used to demonstrate the carrier status of the parents and confirmed the presence of a heterozygous deletion in the *ATAD3* gene cluster in both parents. ^{*} Harel et al., 2016, *Am J Hum Genet* 99:831-845 [†] Desai et al., 2017, *Brain* 140:1595-1610.

2924F

Comprehensive genetic analyses implicate highly heterogeneous etiology of idiopathic cerebral palsy spectrum disorders. *H. Hu.* Guangzhou Women and Children's Medical Center, Guangzhou, China.

Cerebral palsy (CP) is an umbrella concept spanning a range of symptoms albeit with the common feature, i.e., an inborn disturbance of human movement and posture. CP is observed with consistent prevalence regardless of the socioeconomic advancement, and familial CP cases were reported. Seminal studies have confirmed the involvement of genetic changes, in the forms of copy number variant (CNV), insertion and deletion (Indel), and single-nucleotide variant (SNV). As a major children's medical center in China, we recruited a CP cohort of >500 individuals with multiple subtypes and severity. To elucidate the genetic causality, we initiated a pilot study of 120 idiopathic CP patients with their parents and siblings, with a comprehensive genomic assessment. We considered possible etiologic mechanisms from chromosomes to genes, via multiple high-throughput platforms. We found that up to 45% of CP cases could be due to detrimental mutations. Inherited causative mutations were comparable in prevalence with de novo ones, and post-zygotic mutations (PZM) took a nonnegligible portion. Major fractions of the causal mutations were SNVs and indels in nuclear genome, while mitochondrial genome harbored several putative mutations, and large segmental alterations accounted for 8% of CP cases. Strikingly, besides the conventional nonsynonymous and splicing-site mutations, we observed by whole-genome sequencing a host of mutations residing in the regulatory regions which might serve as culprit for disease. Our work is of critical significance to reveal the underlying reasons of CP, which is among the major causes for children referred to the neuropsychiatric departments and rehabilitation institutions.

2925W

A rapid, high-throughput multiplex PCR/CE assay that reliably quantifies *SMN1* copy number. J.N. Milligan, W. Laosinchai-Wolf, K. Jefferson, I. Beldorth, K. Nicholson, B. Printy, S. Filipovic-Sadic, J. Ashton. Asuragen, Austin, TX.

Introduction: Spinal Muscular Atrophy (SMA), an autosomal neuromuscular disease caused by loss of *SMN1* gene function, is the primary genetic cause of infant death with an incidence of ~1/10,000 and a carrier rate of ~1/50. In 95% of patients, both genomic copies of *SMN1* are absent, typically measured by detection of exon 7 which has a single coding nucleotide that distinguishes it from the highly homologous *SMN2* gene. Current ACOG carrier screening guidelines recommend offering SMA carrier screening to all women who are pregnant or are considering pregnancy. This broad recommendation requires rapid, accurate, and high-throughput *SMN1* copy number determination. Herein we report the preliminary performance of the AmpliSeq[®] PCR/CE *SMN1* Kit*, a quantitative PCR/Capillary Electrophoresis (CE) assay that quantifies *SMN1* copy number. **Methods:** The optimized and multiplexed PCR simultaneously amplified *SMN1* and an endogenous control (EC) gene and reported results within 5 hours from sample-to-answer. PCR products were separated and quantified via CE (3500 Genetic Analyzer, Thermo Fisher). The copy number of *SMN1* was calculated as the peak area ratio of target gene and EC normalized to a calibrator that was included with the kit along with a positive control. Fifteen previously characterized human cell-line samples and >230 human genomic DNA samples derived from blood or buccal samples were used to characterize the assay. All sample measurements were compared to an independent reference method. **Results:** The assay was compatible with 20-80 ng of blood or buccal genomic DNA purified using common methods. Quantitative results were automatically calculated from raw instrument data using AmpliSeq[®] PCR/CE Reporter 2.0* software, which outputted genotypes as 0, 1, 2, 3, or >3 genomic *SMN1* copies with high sensitivity and specificity. Across 236 clinical samples tested, the overall percent agreement between the prototype kit and reference method was 97% for *SMN1* copy numbers. Additionally, positive percent agreement for independently annotated single-copy *SMN1* samples was 100%, while the negative percent agreement was 99.6%. **Conclusion:** AmpliSeq[®] PCR/CE *SMN1* Kit and companion software accurately quantified *SMN1* copy numbers in <6 hours with a simple reaction setup and robust analytical performance. This rapid method can overcome operational and data analysis limitations of other methods. *This product is in development; Future availability and performance cannot be ensured.

2926T

Developing custom targeted disease research using optimized assays for next-generation sequencing: An automated system from disease to annotated variants. F. Hyland, L. Na, A. Ewing, Y. Zhu, A. Broome, M. Andersen, S. Roman. Thermo Fisher Scientific, South San Francisco, CA.

Targeted multiplex DNA enrichment panels enable sequencing and investigation of one or many genes. Identifying all important genes for a disease of interest, developing optimized and robust assays, and implementing accurate, reproducible and simple analysis pipelines can be challenging. We present methods to enable genetic disease research by automating the development and analysis of assays. Our Disease Research Database is a comprehensive knowledgebase and discovery tool for human genes and genetic disorders. The database has two key novel attributes. First, it integrates data from multiple databases into one system. Second, it provides an unbiased scoring algorithm to rank gene-disease association at any level of the disease ontology hierarchy. Optimized gene panels can be developed narrowly targeted to specific diseases, or larger gene panels can be developed for broader phenotypes. Disease categories include early onset neonatal phenotypes such as metabolic disorders, Severe Combined Immunodeficiency, heme disorders, developmental disabilities; and late onset phenotypes such as cardiovascular disorders. We developed optimized AmpliSeq assays for 5000 disease research genes; we developed a custom assay design algorithm, and all amplicons were carefully performance screened, with each gene being screened in multiple panels. We performed multiple rounds of assay designs, testing, and redesign especially for the ACMG 59 genes, and we cover a comprehensive list of newborn screening genes. From a simple web interface, scientists can select any combination of diseases, and see all the relevant genes. Any of these genes (plus additional genes), can be combined into a panel, and expected coverage for each gene can be visualized. A custom Ion AmpliSeq gene panel can be created containing any combination of these optimized assays. We created panels for neonatal screening, lung diseases, kidney diseases, rheumatoid arthritis, ACMG 59, among others. The panel size ranged from 17 to 300 genes. We tested on a variety of sample types including frozen samples, cheek swabs, and dried blood spots. We saw very consistent coverage > 97%, using 20ng of input DNA. SNP sensitivity on NA12878 was > 99%, and PPV was > 98%. Simple data analysis is available for both single sample and parent-child trios, with optimized variant calling followed by reporting of amino acid change, protein prediction, and detailed annotation of variants and genes from ClinVar, dbSNP, ExAC, OMIM.

2927F

Diagnostic yields of genetic testing for an unselected ALS cohort from a clinical lab. A.S. Lindy, A. Decker, J. Scuffins, D.A. McKnight. Neurogenetics, GeneDx, Gaithersburg, MD.

Amyotrophic Lateral Sclerosis (ALS) is a progressive neurodegenerative disease that involves both upper and lower motor neurons. ALS is clinically related to frontotemporal dementia (FTD), as pathogenic variants in the associated genes can manifest as both disorders in one individual or as separate disorders in members of a single family. To date, 100+ genes have been reported in association with ALS, but very few have been established as high yield; namely *SOD1*, *TARDBP*, *FUS*, *C9ORF72*. We evaluated a cohort of 376 individuals (average age 59 yrs; range 18-88) with a clinical suspicion of ALS, FTD, or both, to determine the diagnostic rates for *C9ORF72* repeat expansion (RE) analysis and sequencing of 24 other genes. The positive yield for *C9ORF72* RE analysis was 18% (55/313) and the average age at molecular diagnosis was 56 yrs (range 21-74). Family history was provided for 48 positive cases; 58% (28/48) had a family history of ALS/FTD. Of the 258 negative *C9ORF72* RE analyses, 35% (90/258) of individuals pursued sequencing analysis of an additional 24 genes and 3% (3/90) had a subsequent positive result. An additional 72 individuals did not have *C9ORF72* RE analysis, to our knowledge, but pursued multi-gene sequencing. Collectively, 7% (12/162) of all sequencing tests were positive and the average age at molecular diagnosis was 53 yrs. Family history was provided for all 12 individuals with a positive result: 83% (10/12) had a family history of ALS/FTD. Diagnostic results were reported in 7/24 genes sequenced: *SOD1*, *TARDBP*, *VCP*, *FUS*, *MAPT*, *OPTN*, and *SQSTM1*. By far, *C9ORF72* had the highest positive yield (18%) of all genes analyzed and a positive *C9ORF72* result was only associated with a positive family history in 58% of cases; suggesting that this analysis be considered regardless of family history. Family history appeared to be more predictive of a positive result for the other genes evaluated and was observed in 83% of positive cases. This could be due to the smaller sample size, lower overall positive rate or later age of onset associated with these genes, or preferential referral of individuals with a family history. In summary, *C9ORF72* RE analysis has a high diagnostic yield (18%) for individuals with ALS/FTD regardless of family history, but sequence analysis also has utility in the diagnosis of these individuals regardless of family history (7% positive; 17% positive without family history) and should be completed after negative *C9ORF72* RE analysis.

2928W

Precision medicine for stress disorders: Diagnostic biomarkers and repurposed drugs. A. Niculescu^{1,2,3}, H. Le-Niculescu¹, K. Roseberry¹, D. Levey¹, J. Rogers¹, K. Koasary¹, S. Prabha¹, T. Jones³, S. Judd¹, M. McCormick¹, A. Wessel¹, A. Williams³, P. Phalen¹, F. Mamdani¹, A. Sequeira⁴, S. Kurian⁵. 1) Indiana University School of Medicine, Indianapolis, IN; 2) Stark Neuroscience Research Institute, Indiana University School of Medicine, Indianapolis, IN; 3) Indianapolis VA Medical Center, Indianapolis, IN; 4) Department of Psychiatry and Human Behavior, UC Irvine, Irvine, CA; 5) Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, CA.

The biological fingerprint of environmental adversity may be key to understanding health and disease, as it encompasses the damage induced as well as the compensatory reactions of the organism. Metabolic and hormonal changes may be an informative but incomplete window into the underlying biology. We endeavored to identify objective blood gene expression biomarkers for psychological stress, a subjective sensation with biological roots. To quantify the stress perception at a particular moment in time, we used a simple visual analogue scale for life stress in psychiatric patients, a high risk group. Then, using a stepwise discovery, prioritization, validation, and testing in independent cohorts design, we were successful in identifying gene expression biomarkers that were predictive of high stress states, and of future psychiatric hospitalizations related to stress, more so when personalized by gender and diagnosis. Some of these biomarkers are increased in expression in high stress states (being putative risk genes), and others are decreased in expression (being putative protective/resilience genes). One of the top biomarkers that survived discovery, prioritization, validation and testing was FKBP5, a well-known gene involved in stress response, which serves as a de facto reassuring positive control. We also compared our biomarker findings with telomere length (TL), another well-established biological marker of psychological stress, and show that newly identified predictive biomarkers are comparable or better state or trait predictors of stress than TL or FKBP5. Over half of the top predictive biomarkers for stress also had prior evidence of involvement in suicide, and the majority of them had evidence in other psychiatric disorders, providing a molecular underpinning for the effects of stress in those disorders. Some of the biomarkers are targets of existing drugs, of potential utility in patient stratification and pharmacogenomics approaches. Moreover, the biomarkers gene expression signatures yielded new drug candidates and natural compounds upon bioinformatics drug repurposing analyses. Our work may lead to improved diagnosis and treatment for stress disorders such as PTSD, that result in decreased quality of life and adverse outcomes including addictions, violence and suicide.

2929T

Maternally inherited 8q11.23 microduplication identified in 5 affected children with ASD from two unrelated multiplex families. *M.E.S. Lewis¹, Y. Qiao¹, K. Calli¹, S. Redmond¹, S. Martell¹, C. Chijiwa¹, E. Rajcan-Separovic².* 1) Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada; 2) Pathology and Laboratory Medicine, University of British Columbia, Vancouver, British Columbia, Canada.

We have recruited >700 families with simplex and multiplex ASD and are performing chromosome microarray and whole genome sequencing studies as part of the iTARGET Autism project (<http://www.itargetautism.ca/>), open-access genomic database for autism research. In this cohort, we identified 5 affected subjects from 2 unrelated families who have ASD and a maternally inherited microduplication at 8q11.23 involving *RB1CC1*. Microduplications involving this gene have been reported in 30 cases in DECIPHER Database with variable neurodevelopmental phenotypes, as de novo or inherited (maternal or paternal). Their clinical significance, therefore remains uncertain. The mother in Family 1 has a past history of depression. The father in Family 2 has many Asperger-like features. None of the affected 5 children have outward dysmorphic features. Genetically, the 5 children with ASD were found to carry a similar maternal 8q11.23 duplication (maximum range: 53413457-53827622bp, hg19) involving 3 genes (*RB1CC1*, *ALKAL1*, and *NPBWR1*) with the first two genes shared by both families. Family 2 also contains gene *NPBWR1*. *RB1CC1* is a DNA-binding transcription factor involved in the regulation of multiple cell processes including neuronal homeostasis. Duplication of this gene has been reported to be associated with schizophrenia. Animal models have shown deletion of this gene leads to cerebellar degeneration. Somatic mutations in *RB1CC1* are more frequently observed in cancers, including breast cancers. From published papers, none of the 3 genes have been reported to be related to autism. Our two families therefore expand the phenotypic spectrum of this copy number variant which may be yet another locus for neurodevelopmental abnormalities with variable penetrance. On-going whole genome analysis may uncover additional genetic factors causing ASD.

2930F

Application of the ClinGen gene-disease clinical validity process to assess the strength of evidence for genes implicated in autism and intellectual disability. *B. Bostwick¹, C. Barry², A. Behlmann³, C. Chen¹, C. Collins⁴, A. Grant⁵, E. Riggs⁶, A. Ronza¹, R. Siegert⁷, T. Sneddon⁸, C. Thaxton⁹, M. Tokita¹⁰, J. Yin¹¹, X. Zhou¹², D. Miller¹¹, C. Schaaf¹².* *ClinGen Intellectual Disability/Autism Gene Curation Expert Panel.* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA; 2) Autism & Developmental Medicine Institute, Geisinger Health System, Lewisburg, PA, USA; 3) Department of Human Genetics, Emory University School of Medicine, Atlanta, GA, USA; 4) Perkin Elmer Genomics, Branford, CT, USA; 5) Laboratory for Molecular Medicine, Partners Healthcare Personalized Medicine, Cambridge, MA, USA; 6) Department of Clinical Genomics, Stanford School of Medicine, Stanford, CA, USA; 7) Department of Genetics, The University of North Carolina, Chapel Hill, NC, USA; 8) Institute for Genomic Medicine, Rady Children's Hospital-San Diego, San Diego, CA, USA; 9) Department of Neurology, University of California Los Angeles, Los Angeles, CA, USA; 10) The Institute for Precision Medicine, University of Pittsburgh Medical Center, Pittsburgh, PA, USA; 11) Division of Genetics & Genomics, Children's Hospital Boston & Harvard Medical School, Boston, MA, USA; 12) Institute of Human Genetics, University Hospital Cologne, Köln, Germany.

OBJECTIVE: As part of the NIH-funded Clinical Genome Resource (ClinGen) effort, the Intellectual Disability and Autism Gene Curation Expert Panel evaluates clinical, basic science, and population data to determine the strength of evidence that a gene contributes to the phenotypic spectrum of autism and intellectual disability. This collaborative workgroup of clinicians, researchers, biocurators, and molecular diagnosticians aim to annotate and share the strength of gene-disease relationships for the benefit of those in the medical genetics field and the public. **METHODS:** We used an evidence-based standardized framework to evaluate genetic and experimental evidence supporting or refuting a gene's role in a given disease. Genes implicated with ID/autism were categorized as Definitive, Strong, Moderate, or Limited using a systemized scoring rubric to assess strength of evidence based on data in published clinical reports, basic research articles, and population databases. We generated a prioritized gene list to review by selecting genes appearing in 25% or more of commercially available ID/autism sequencing or microarray panels in the Genetic Testing Registry that had not previously been evaluated by SFARI or the Geisinger Developmental Brain Disorders Database. **RESULTS:** To date, the ClinGen ID/Autism Gene Curation Expert Panel has evaluated and classified 30 gene-disease pairs. Of the 30 genes curated thus far, 22 (76%) have Definitive evidence to support an ID/autism gene-disease relationship. Surprisingly, five of the curated genes (17%) were found to have only Limited evidence for an ID/autism gene-disease relationship despite inclusion on several commercial genetic testing panels. The curation results are available on clinicalgenome.org. **CONCLUSIONS:** The ID/Autism Gene Curation Expert Panel's curated information on genes and diseases is publicly available. This information can be utilized in a variety of clinical and research applications, including building evidence-based clinical genetic testing panels, resolving discrepancies in variant interpretation, and improving patient care through an evidence-based medicine approach. Diagnostic laboratories should consider the strength of the evidence for gene-disease relationship prior to including a gene in a diagnostic panel.

2931W

Genetic test results for 523 patients with ASD/ID: The diagnostic yield of multigene analysis (Autism/ID Xpanded test) is higher than conventional first-tier tests, such as FMR1 repeat analysis and chromosomal microarray. A. Shanmugham, D. McKnight, J. Scuffins, T. Brandt. GeneDx Inc., Gaithersburg, MD.

Established guidelines for evaluating children with autism spectrum disorder (ASD), global developmental delay (GDD) and/or intellectual disability (ID) recommend FMR1 repeat analysis and chromosome microarray (CMA) as first-tier tests. However, FMR1 analysis as a first-tier test is being re-examined as studies show a low diagnostic yield for individuals without a clinical suspicion of Fragile X syndrome. We performed a retrospective study to evaluate the diagnostic yield of currently recommended first-tier tests (FMR1 and CMA) compared to a second-tier test, the Autism/ID Xpanded panel (sequencing of >2300 genes associated with ASD/ID). The patient cohort was unselected and included 523 individuals (145 females; 378 males) with ASD/GDD/ID referred to our laboratory for genetic testing. The overall positive diagnostic rate for first-tier genetic testing (FMR1 analysis and CMA) was 6.1% (32/523). Positive FMR1 full expansions were identified in only 0.6% (3/523) of cases, corresponding to 0.8% (3/378) of males. The three males were reported to have either GDD, developmental regression and ASD or ID. CMA testing identified causative copy number variants (CNVs) in 5.5% (29/523) of cases. Seventy-two percent of all individuals (375/523) proceeded to second-tier testing using the Autism/ID Xpanded panel, and almost 10% (36/375) of those had a diagnostic result. Of note, five positive CMA cases continued on with an Autism/ID Xpanded panel. In only one of these cases were other causative variants identified by sequencing. In summary, our results demonstrate that the diagnostic yield of FMR1 testing in general is low and, despite its relatively low cost, this test is not an effective first-tier test, particularly for affected females. CMA yielded a higher diagnostic rate than FMR1 analysis, similar to what is reported in the literature. The Autism/ID Xpanded panel had the highest diagnostic yield in this study. While ordering a CMA as a first-tier test is more cost effective today, technology advances improving detection of CNVs from NGS data will soon allow detection of sequence and copy number variants with a single test at reasonable cost; therefore, guidelines for evaluating children with ASD, GDD and/or ID may change in the future.

2932T

Clinical symptoms of Fragile X syndrome patients depends on mosaicism in CGG repeat length and FMR1 promoter methylation. D.V. Yudkin¹, I.V. Grishchenko², Y.V. Maksimova², A.R. Shorina². 1) SRC VB VECTOR, Novosibirsk Region, Koltsovo, Russian Federation; 2) Novosibirsk State Medical University, Novosibirsk, Russian Federation; 3) Novosibirsk City Clinical Hospital No. 1, Novosibirsk, Russian Federation.

Fragile X syndrome is a most common reason of inherited intellectual disability in human. Mechanism of the syndrome development is due to CGG repeat expansion in 5' UTR of FMR1 gene located in chromosome X. Normally the gene contains no more than 50 triplets. When repeat size became more than 200 triplets – full mutation – the syndrome develops. Repeat expansion leads to methylation and heterochromatinization of FMR1 promoter region followed by gene silencing. Absence of FMRP encoded by FMR1 is cause of abnormal neurons development and intellectual disability. CGG repeat size in patients can vary from 200 to 2000 triplets and more. Additionally patients can have somatic mosaicism and somatic repeat expansion. Woman carriers have two chromosomes X and two alleles of FMR1 gene and somatic mosaicism. Methylation of FMR1 promoter can be presented not in each cell and some patients with full mutation can save some level of FMR1 expression. All above is influence to degree of intellectual disability and clinical cases are vary. Samples of peripheral venous blood were collected in Novosibirsk City Hospital No. 1 according to international standards, which include the awareness of the subject, his or her consent to participating in the study in its entirety, and guarantees of confidentiality. All of the studies conformed to ethical standards developed in accordance with the Helsinki Declaration of the World Medical Association as amended in 2000. In addition, the studies were supervised by the Institutional Review Board. DNA was purified and analyzed by GC-rich PCR followed by fragment analysis. Study of methylation level in FMR1 5' UTR was carried out using methyl dependent restriction endonuclease followed by PCR. Clinical study was obtained from psychiatric counseling for every patient. As a result, we have investigated the severity of symptoms in dependence of repeat size, methylation level and somatic mosaicism in repeat size and methylation level. The high level of mosaicism was shown. Acuity of symptoms decreases in presence of premutant and normal alleles in mosaic form. Additionally short repeat even in full mutation leads to decreasing of chromosome fragility. This study is supported by the Russian Science Foundation (grant 18-15-00099).

2933F

Myelin redundancy: A brain biomarker for “aging” versus Alzheimer’s disease in Down syndrome. A. Van Hoek¹, L. Dai², J.R. Korenberg². 1) Neurology, University of Utah, Salt Lake City, UT; 2) Pediatrics, Medical Genetics, University of Utah, Salt Lake City, UT.

A major challenge for accelerating therapeutics of Alzheimer’s Disease (AD) and “normal” Aging is the identification of dynamic neuronal and glial biomarkers that distinguishes AD from normal Aging. Down syndrome (DS) is considered an ideal model given that the pathology of aging involves degenerative neuronal and myelin features with genetic causes like AD, which characterize 100% with DS by age 35 as forerunners of dementia affecting ~70%. However, DS-AD dementia merges with signs of accelerated aging seen in DS with three copies of the HS Chr 21 gene APP encoding the amyloid plaque protein, but without biomarkers for the age-related waning of capacities not related to AD. We examined the architecture of myelin in 22 cervical spinal cord samples obtained from the NIH NeuroBioBank (NBB) and our NIH DS study, with diagnoses of DS (41-76y), AD (54-60y) or normal control (50-74y) by genetic (confocal microscopy) and morphological (transmission electron microscopy) analyses of 11 single donor glutaraldehyde (GA) and formalin (F) fixed tissues. The results thus far indicate that DS myelin (n=6, F and GA fixed tissues) exhibited striking age-related redundancy and exaggerated ballooning aberrant sheaths connecting neighboring normal and degenerated axons and concentric sheaths separated by apparent oligodendrocytic cytoplasmic contents. In contrast, myelin in AD (n=4; 54, 60y) was ballooned, non-redundant with interlaminar dark globular electron dense and extra-axonal fibrillary material. Myelin of aged controls (52-58y) was even less ballooned and more compact. Chloroform extraction eliminated ballooned myelin laminae resulting in large empty regions in DS retaining compacted myelin and structures reflecting helical formation. These results indicate that aged and less so young DS but not AD is associated with redundant myelin sheaths that are readily extracted in contrast to normal myelin. This biomarker of normal aging in DS may suggest a linear decrease in conduction velocity with decreased effective sheath thickness and consequent cognitive loss in DS and more subtly in non-DS, non-AD aging. Although its severity may result from aberrant repair by the Chr 21 encoded gene OLIG2, the redundancy biomarker may reflect physiologic and modifiable in vivo dysfunction that constitutes a therapeutic target for non-AD aging, whether or not related to Alzheimer’s neuropathology.

2934W

Whole-genome sequencing is more effective than whole-exome sequencing for detecting structural variants. F. Mafra¹, D. Li¹, E. Bhoji¹, M. Harr¹, R. Pellegrino¹, B. Keena², E. Zackai², H. Hakonarson^{1,3}. 1) Center for Applied Genomics, The Children’s Hospital of Philadelphia, Philadelphia, PA; 2) Division of Human Genetics, The Children’s Hospital of Philadelphia, Philadelphia, PA; 3) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

Introduction: Whole exome sequencing (WES) and whole genome sequencing (WGS), are increasingly used in diagnostic and research laboratories to identify rare and common genetic variants in humans; these approaches are known as next-generation sequencing. WES focuses on the protein coding sequences, usually being the first method of choice due to the lower cost since less overall sequence is needed. Although currently more expensive, WGS is more powerful than WES for detecting potential disease-causing mutations within the exome, particularly structural variants, variants in GC-rich regions, and variants in regulatory regions. **Methods:** We compared the two NGS techniques, performing WES with the Agilent Sure Select Human All Exon kit v5 and WGS with the Illumina TruSeq DNA PCR-Free sample preparation kit on a blood sample from a 5 years old Caucasian patient with a history of epilepsy, micrognathia, microcephaly, optic nerve hypoplasia, short stature, developmental delay and febrile seizures. The findings were validated by Sanger sequencing. **Results:** No candidate genes were elucidated using WES, while the WGS revealed a *de novo* 5 kb single-exon deletion in the gene *DYRK1A* (chr21:38,853,731-38,859,397; hg19). Based on the autosomal dominant inheritance pattern associated with the disease linked to the gene, phenotypic match, published literature and population frequency databases checking (such as internal cohort, ClinGen, DECIPHER, and DGV), this variant was classified as pathogenic and was considered as the potential cause of the patient’s phenotype. The libraries showed differences in depth, read length and PCR amplification in addition to exome capture. WGS gave superior variant calls where its longer reads gave better mappability, the coverage was more uniform in total depth and in distribution of the position of a variant within a read. **Conclusions:** This case demonstrated the utility of WGS for the detection of structural variants (SV), particularly small deletions. This is due to a higher uniformity of coverage achieved with WGS and the fact that this SV was not covered by WES because of the poor probe density at the breakpoints region. The main factors underlying the heterogeneous coverage of WES are related to the hybridization/capture and PCR-amplification steps required for the preparation of sequencing libraries for WES. **Acknowledgments:** WGS funded by the Roberts Collaborative Study.

2935T

Benign-Ex & DISCRIMINATOR: Assigning pathogenicity classifications to CNVs. A. Hahn^{1,2}, M. Parida³, H. Major¹, J. Manak^{2,4}, B. Darbro^{1,2}. 1) Stead Family Department of Pediatrics, Roy J. and Lucille A. Carver College of Medicine, University of Iowa, Iowa City, IA, United States; 2) Interdisciplinary Program in Genetics, University of Iowa, Iowa City, IA, United States; 3) Biochemistry, University of Iowa, Iowa City, IA, United States; 4) Biology, University of Iowa, Iowa City, IA, United States.

Developmental disabilities affect nearly 15% of children and 10-20% of cases are attributable to copy number variants (CNVs). While over a hundred microdeletion/microduplication syndromes (MMS) have been described, several thousand CNVs have been characterized as variants of unknown significance (VUS) and genetic heterogeneity has been identified in several known MMS including DiGeorge/Velocardiofacial Syndrome, Pitt-Hopkins Syndrome, and Rhetts Syndrome. The overall goal of our research laboratory is to determine if any VUS represent novel genetic causes of known MMS. To accomplish this goal, we use a combination of computational tools to identify the underlying genetic burden imposed by CNVs. We have developed and tested two such tools (Benign-Ex & DISCRIMINATOR) on chromosomal microarray data from pediatric patients referred for clinical testing (N=3416). Developmental delay was the most prevalent phenotype followed by dysmorphic features and autism. CNV data was collected over a ten-year period and utilized three different CMA platforms: NimbleGen 385K (n=417; 12%), NimbleGen 720K (n=655; 19%), and Affymetrix CytoScanHD (n=2344; 69%). Benign-Ex utilizes DGV and ClinGen to identify regions of the human genome which are copy number variable without phenotypic consequence (benign), while DISCRIMINATOR annotates CNVs as benign, primary (a known pathogenic MMS CNV), or secondary (VUS). Here we present results from the training of Benign-Ex using data from OMIM, DECIPHER, ClinGen, and a manually curated list of all known MMS associated with developmental disabilities. These training lists were used to balance and optimize the performance of Benign-Ex to reduce identifying pathogenic regions as benign. When optimized, Benign-Ex designated more of the genome as copy number variable compared to previous analyses using DGV alone. DISCRIMINATOR utilizes the output of Benign-Ex as well as a list of known MMS to assign pathogenicity classifications to the CNV datasets (primary, secondary, or benign). In our dataset we discovered >350 primary CNVs, >15,000 secondary CNVs, and that >75% of all CNVs were classified as benign. Additionally, consistent with their prevalence rates, 22q11.2, 15q11.2, and 16p11.2 primary CNVs were the most common MMS detected by DISCRIMINATOR. Identified non-benign CNVs will be used in downstream analyses to explore the genetic heterogeneity of known MMS and identify genes and/or networks which are important to the MMS's clinical phenotype.

2936F

Evaluating dosage sensitivity of genes associated with neurodevelopmental disorders. E. Riggs¹, C. Barry¹, J.W. Ahn², T. Brandt³, O. Cano³, Y. Cao⁴, M. Goodenberger⁵, S. Kantarcic⁶, W. Khan⁷, G. Li³, C. Ngo⁸, H. Risheg⁹, C. Runke⁵, M. Su¹⁰, Y. Shen¹¹, S. Aradhya¹², E. Thorland⁵, C.L. Martin¹. 1) Autism & Developmental Medicine Institute, Geisinger, Danville, PA; 2) Genetics Laboratories, Guy's and St Thomas' NHS Foundation Trust, Great Maze Pond, London, SE1 9RT, UK; 3) GeneDx, Gaithersburg, MD; 4) Department of Pathology and Immunology Washington University in St. Louis, School of Medicine, St. Louis, MO; 5) Genomics Laboratory, Department of Lab Medicine and Pathology, Mayo Clinic, Rochester, MN; 6) Cytogenetics and Genomics, Quest Diagnostics Nichols Institute, San Juan Capistrano, CA; 7) Icahn School of Medicine at Mount Sinai, Department of Genetics and Genomic Sciences, New York, NY; 8) Department of Cytogenetics, The Children's Hospital at Westmead, Corner Hawkesbury Road Hainsworth Street, Westmead, NSW, 2145; 9) Laboratory Corporation of America® Holdings/Dynacare, Seattle, Washington; 10) Department of Pathology and Laboratory Medicine, University of Miami Miller School of Medicine, Miami, FL; 11) Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA; 12) Invitae, San Francisco, CA.

The ClinGen Dosage Sensitivity (DS) Working Group systematically evaluates genes and genomic regions for DS in an effort to facilitate the process of genomic content evaluation and promote interpretation consistency among clinical laboratories. Neurodevelopmental disorders (NDDs), such as autism and intellectual disability, are a frequent reason for referral for clinical genetic testing, including multi-gene panel testing, whole exome sequencing, and cytogenomic microarrays. Understanding which genes cause NDDs through haploinsufficiency (HI) or triplosensitivity (TS) mechanisms can help clinical laboratories interpret loss of function (LOF) and/or whole gene duplication events involving these genes. Given the relative importance of this clinical domain, we are focusing on evaluating DS of candidate NDD genes identified through various sources. Our first source is the Geisinger Developmental Brain Disorder Genes Database - an online, curated resource cataloging variants identified in published literature in the context of NDDs. As of 9/2017, 95 genes were identified in the Geisinger database as being "Tier 1," or high confidence candidate genes based on the number of de novo, pathogenic, LOF variants reported. We evaluated each of these genes for evidence of HI and/or TS. Seventy-seven genes (81.1%) were classified as having "sufficient" evidence for HI; LOF variants identified within these genes could be considered likely pathogenic (LP) or pathogenic (P) when observed in a clinical setting. Six genes (6.31%) were classified as having "emerging" evidence for HI, 3 genes (3.2%) had "little" evidence, and 1 gene (1.1%) was classified as "gene associated with autosomal recessive condition." While 1 gene (*MECP2*) had "sufficient" evidence for both HI and TS, most of the 95 genes had either "little" (7.4%) or no evidence (91.5%) for TS, indicating that there is not enough evidence to consider single, whole gene duplications of these particular genes LP or P at this time. We will continue to evaluate genes associated with NDDs in order to provide a comprehensive resource of DS information to the genomics community.

2937W

Novel first tier DNA methylation-based diagnostic platform for neurodevelopmental disorders. S. Choufani¹, K. Garg¹, C. Cytrynbaum², S. Cyrus³, B. Chung⁴, W.T. Gibson³, R. Weksberg^{1,2}. 1) Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Canada; 2) Div Clin and Metabolic Gen, Hosp Sick Children, Toronto, ON, Canada; 3) Child and Family Research Institute, Vancouver, BC, Canada; 4) Dept of Paediatrics and Adolescent Med, Li Ka Shing Faculty of Medicine, Hong Kong.

Diagnostic genetic testing has evolved dramatically in the last decade with the introduction of next generation sequencing (NGS). While significantly improving diagnostic yield, NGS have magnified some of the challenges of DNA sequence-based testing in clinical genomics, most notably the interpretation of variants of unknown significance (VUS), incomplete coverage of certain genomic regions and detection of small deletions/duplications. Our previous work showed that pathogenic mutations in epigenes (genes involved in epigenetic regulation), such as *NSD1*, *CHD7*, *KMT2D*, *KDM5C*, are associated with epigenome-specific genome-wide DNA methylation (DNAm) signatures as well as the utility of these signatures to serve as functional tools for VUS classification. In the current work, we tested the utility of these DNAm signatures for first tier diagnosis of cases with undiagnosed neurodevelopmental syndromes. We compared the profiles of 2000 methylomes (1600 from the Illumina 450k array and 400 from the Illumina EPIC array) generated in our lab from blood samples of cases with known pathogenic mutations, VUS, undiagnosed cases and controls. We tested each sample against the different epigenome-specific DNAm signatures associated with the clinical phenotypes of Sotos (*NSD1*), Weaver (PRC2 complex genes: *EZH2*, *EED*, *SUZ12*), Kabuki (*KMT2D*) and susceptibility to Autism (*CHD8*) syndromes. Using principal component analyses as well as unsupervised hierarchical clustering; genome-wide DNAm data for each case were compared to the disease-specific reference (derived from n=5-10 cases with pathogenic mutations) and a control reference (n=15). This approach allowed us to classify two undiagnosed samples with the *NSD1*-specific DNAm signature. Research-based NGS analyses on these 2 samples identified pathogenic variants in *NSD1*, thereby validating the first tier diagnosis of Sotos using the DNAm signature. A review of the phenotype of these two patients demonstrated macrocephaly, developmental delay and other characteristic features of Sotos syndrome. Similarly, we were able to predict mutations in PRC2 complex genes (n=2), *KMT2D* (n=1) and *CHD8* (n=1); these diagnoses were confirmed by NGS and post hoc clinical review of these cases. These data highlight the potential utility of gene-specific DNAm signatures not only in functionally classifying VUS as pathogenic or benign but also in categorizing individuals by syndromic diagnosis, thereby serving as a first tier diagnostic platform.

2938T

Constructing a framework for evaluating recurrent CNVs with reduced penetrance or variable expressivity: A guide for clinical interpretation. E. Andersen¹, R. Burnside^{2*}, B. Coe³, L. Conlin⁴, J. Hergig⁵, T. Higginbotham⁶, B. Hilton¹, V. Jobanputra⁷, H. Kearney⁸, K. Ouyang⁹, P. Paulraj¹, E.R. Riggs¹⁰, R. Rowsey⁶, M. Speevak⁶, E. Thorland⁸, M. Goodenberger⁶, C. Runke⁸, L. Zhang⁵, C.L. Martin¹⁰ on behalf of the Clinical Genome Resource (ClinGen) (*Equal contribution). 1) University of Utah/ARUP Laboratories, Salt Lake City, UT; 2) Laboratory Corporation of America Holdings®, Research Triangle Park, NC; 3) University of Washington, Seattle, WA; 4) Children's Hospital of Philadelphia, Philadelphia, PA; 5) Children's Mercy Hospital, Kansas City, MO; 6) Trillium Health Partners, University of Toronto, Toronto, ON; 7) New York Genome Center and Columbia University Medical Center, New York, NY; 8) Mayo Clinic, Rochester, MN; 9) Invitae, San Francisco, CA; 10) Geisinger, Autism & Developmental Medicine Institute, Danville, PA.

The ClinGen Dosage Sensitivity (DS) Curation workgroup uses an established evidenced-based review process to evaluate genes/genomic regions for haploinsufficiency and triplosensitivity, assigning a numerical score corresponding to strength of evidence associated with each dosage mechanism. While this process has been effective for evaluating individual genes in a standardized manner, the evaluation of multigenic regions, particularly recurrent copy number variants (CNVs), has been challenging. For example, regions such as 15q11.2 (BP1-BP2) or 16p12.2 exhibit incomplete penetrance and variable expressivity; and segregation data, along with statistics for population-level frequency can be challenging to interpret clinically without well-defined syndromic features. While our current framework incorporates levels of knowledge of penetrance and expressivity into gene and region scoring, more refined criteria are necessary to define "well documented" associations between phenotype and incomplete penetrance and/or variable expressivity, including strength of statistical data and segregation patterns. Moreover, literature describing recurrent CNVs often presents conflicting evidence depending on the type of study (e.g., case-control vs. cohort only). In order to effectively and consistently account for these factors, a subgroup composed of professionals with expertise on recurrent CNV interpretation is currently working to revise the criteria for dosage sensitivity scoring of recurrent CNVs. We propose incorporating additional variables into our original framework, including the number of unrelated probands, number of independent peer-reviewed studies, study type, phenotypic specificity and variability, segregation/inheritance data, population-level statistics, and conflicting data; these revisions include weighting types of evidence and study types. For example, statistically significant enrichment in cases in a case-control study will be weighted more than lack of such enrichment or case-only studies. Likewise, specific phenotypes will be weighted more than non-specific phenotypes. This new framework will provide objectivity to scoring, result in better alignment with other more quantitative ClinGen efforts with respect to variant evaluation, and aid in clinical interpretation and reporting of recurrent CNVs.

2939F

Population prevalence and inheritance pattern of recurrent CNVs associated with neurodevelopmental disorders. A study of 7986 newborns from the Norwegian Mother and Child Cohort. S. Johansson^{1,2}, K. Lavrichenko^{1,3}, D. Smajlagic^{1,2}, O. Helgeland^{1,4}, S. Berland^{1,2}, G.P. Knudsen⁴, P.M. Knappskog^{1,2}, J. Haavik⁵, M. Vaudel^{1,2}, G. Houge^{1,2}, P.R. Njølstad¹, T. Zayats⁵. 1) Department of Clinical Science, University of Bergen, Bergen, Norway; 2) Department of Medical Genetics, Haukeland University Hospital, Bergen, Norway; 3) Computational Biology Unit, Department of Informatics, University of Bergen, Bergen, Norway; 4) Department of Genetics and Bioinformatics Norwegian Institute of Public Health, Oslo, Norway; 5) Department of Biomedicine, University of Bergen, Bergen, Norway.

Introduction. Rare recurrent CNVs are common causes of neurodevelopmental disorders and risk factors for psychiatric traits. A number of regions prone to recurrent duplications and deletions have been extensively studied in various disease cohorts, however less is known about the true prevalence and inheritance pattern in the general population. Here we present an interim analysis using SNP-array-based CNV calling of the first 7986 child-mother-father trios genotyped in the Norwegian Mother and Child Cohort (MoBa). **Materials and methods.** 7986 trios of Norwegian ancestry from the MoBa cohort were genotyped using the Illumina HumanCoreExome chip and passed stringent quality control for CNV-analysis. CNVs were first called using PennCNV and PennCNVtrio. We developed a set of tools specific for trio analyses post- PennCNV trio calling, to further improve and refine the CNV calls and their inheritance pattern. All putative de novo calls were visually inspected for accuracy. **Results.** As of June 2018, a total of 7986 complete trios have been CNV genotyped and passed strict QC. We found a total of 106 children with bona fide de novo CNVs > 100 kb, corresponding to a minimal prevalence of 1.3 % large de novos per generation. As previously described, more de novo CNVs were of paternal origin (~ 60%). We next analysed a list of regions known to harbour recurrent CNVs implicated in neurodevelopmental disorders (NDD) and psychiatric traits. We found 22 de novo and 161 inherited rare deletions and duplication >100 kb intersecting the known NDD regions. For example the known recurrent chr16p11.2 microdeletion was seen in 6 newborns (4 being de novo) corresponding to a prevalence of 7.5 per 10 000 newborns with 66% being de novo. The reciprocal microduplication was found in 7 children (all inherited). Corresponding numbers for the large 2 Mb microdeletion/duplication at the chr15q13.3 locus were 4 deletions (all inherited) and 5 duplications (1 de novo), corresponding to prevalences of approximately 5-6 events per 10 000 live born children each for duplications and deletions. **Conclusion/Discussion.** Our interim analysis of an unselected and representative population of newborns and their parents provide a clearer picture of the rate of recurrent microduplication/deletions implicated in developmental delay. Further 6000 trios will be included in the final analyses providing a unique opportunity to study the phenotypic consequences using MoBa data from birth to 8 years of age.

2940W

Analysis of copy number variations in 400 patients with mental retardation or developmental delay using aCGH. B. Wu, R. Liu, W. Lu, H. Wang, W. Zhou. Children's Hospital Fudan University, Shanghai, China.

Objective: Using Array Comparative Genomic Hybridization (aCGH) to detect genomic imbalances in patients with unexplained mental retardation or developmental delay (MR/DD) and to evaluate the effectiveness of aCGH in molecular diagnosis of children with unexplained MR/DD in China. **Methods:** Children with unexplained MR/DD were enrolled in this study. DNA was extracted from whole blood of the patients by standard procedures. aCGH was performed using the Agilent 4*180 CGH-SNP chip (Boston Children's Hospital custom designed). Data analysis was performed using Agilent Cytogenomics software. Genomic positions refer to the Human Genome February 2009 assembly (GRCh37/hg19). Copy number variations (CNVs) were interpreted and classified according to phenotypic features of the patient, and records in the Database of Genomic Variants (DGV), Online Mendelian Inheritance in Man (OMIM), Clinical Genome Resource (ClinGen), Database of Genomic Variation and Phenotype in Humans using Ensembl Resources (DECIPHER), in-house databases et al. **Results:** A total of 400 patients with unexplained MR/DD with or without dysmorphisms, multiple congenital anomalies, learning difficulties and autism spectrum disorders (ASD) were recruited, in which 234 cases were male. The age ranged from 15 days to 15 years. Of all samples completed the aCGH tests, 92 cases had pathogenic CNVs, including 37 known syndromes. The size of pathogenic copy number variants ranged from 40 Kb to 38 Mb. 77 cases had detected CNVs of unknown significance, including 32 deletions and 45 duplications, accounting for 19% of the total test sample. 24 cases had detected absence of heterozygosity (AOH) greater than 9Mb, of which 4 cases had more than 3 AOH regions on the chromosome, indicating consanguinity relationship between the parents. Another 2 cases had detected 151Kb and 6.8Mb likely pathogenic microdeletions, containing autosomal recessive gene match to the phenotype of the patients in the deletion area. Sanger sequencing of the candidate gene was done after that, and another pathogenic point mutation was found in the gene. QPCR and Sanger confirmation tests using parents' samples showed that one parent carried the deletion and the other carried the pathogenic point mutation. **Conclusion:** Genome imbalance contribute up to 25% genetic etiology to patients with unexplained MR/DD. aCGH is an effective method in molecular diagnosis of children with unexplained MR/DD in China.

2941T

Mutational recurrence in developmental disorders: The use of publicly available data of large trio cohorts for efficient identification of clinically relevant pathogenic variants. F. Lecoquierre^{1,2,6}, B. Urteaga¹, A. Vitobello^{1,2}, C. Coubes³, AL. Bruehl¹, P. Garret¹, S. Nambot^{1,4}, M. Chevarin^{1,2}, S. Moutton^{1,4}, F. Tran Mau-Them^{1,2}, C. Philippe^{1,2}, A. Sorlin^{1,2,4}, Y. Duffourd^{1,2}, L. Faivre^{1,4}, C. Thauvin-Robinet^{1,2,5}, Orphanomix Consortium. 1) Inserm UMR 1231 GAD, Genetics of Developmental disorders, Université de Bourgogne-Franche Comté, FHU TRANSLAD, Dijon, France; 2) Unité Fonctionnelle «Innovation diagnostique dans les maladies rares», laboratoire de génétique chromosomique et moléculaire, Plateau Technique de Biologie, CHU Dijon Bourgogne, Dijon, France; 3) Centre de Référence Maladies Rares "Anomalies du Développement et syndromes malformatifs", Service de Génétique, CHRU Montpellier, France; 4) Centre de Référence Maladies Rares "Anomalies du Développement et syndromes malformatifs", FHU-TRANSLAD, CHU Dijon Bourgogne, France; 5) Centre de Référence Maladies Rares "Déficiences Intellectuelles de causes rares", FHU-TRANSLAD, CHU Dijon Bourgogne, France; 6) Normandie Univ, UNIROUEN, Inserm U1245 and Rouen University Hospital, Department of Genetics, F 76000, Normandy Center for Genomic and Personalized Medicine, Rouen, France.

Introduction During the past decade, extensive use of next generation sequencing in the field of developmental disorders has increased substantially our understanding of these entities. Deciphering Developmental Disorders study has reached a diagnostic yield of 23% considering only coding *de novo* mutations (DNM), but their statistical modeling predicted that up to 41.8% of cases could be explained by a pathogenic DNM detected by trio exome sequencing. These results highlight the long way to go to in the delineation of new genotype-phenotype correlations. Notably, diseases caused by specific missense variants in a modification-of-function mechanism are more challenging to model than loss-of-function variants and their identification will require innovative strategies and large scale data sharing. **Material and methods** We present an easy approach based on the publicly available denovo-db data, a large collection of *de novo* mutations from various trio cohort studies. We used denovo-db to retrospectively flag variants from our 1,200 solo exomes database, with the idea that mutational recurrence in at least two unrelated individuals with a developmental disorder (one from denovo-db and one from our local cohort), would be a powerful approach to identify strong candidate variants. **Results** This approach identified a limited set of 51 variants, from which 32 variants showed high evidence for pathogenicity, including variants already known as pathogenic in our team, but also several new diagnoses. Mutational recurrence highly relied on CpG transitions, but other mechanisms such as 'selfish spermatogonial selection' did appear clearly from our data. We identified *FEM1B* as a new gene for syndromic developmental disorder including global developmental delay, auto-aggressively, facial dysmorphic features, deafness and heart defects. We also confirmed the implication of *ACTL6B* as responsible for a neurodevelopmental disorder. **Discussion** We propose that the prospective and retrospective annotation of local variants with de-novo db could be an easy and effective strategy for detecting rare variants associated with developmental disorders. We hypothesize that this kind of approach based on mutational recurrence will be essential for the identification of several new diseases caused by specific missense variants, in the process of getting closer to the high estimated yields brought by whole exome sequencing.

2942F

Whole-exome trio sequencing in 51 patients with intellectual disability/developmental delay. Y. Dincer^{1,2}, J. Schulz¹, C.M. Wilke¹, M.Y. Cohen^{1,3}, D. Wahl^{1,4}, S. Wilson¹, S.H. Eck¹, V. Mall^{2,3}, I. Rost¹, H.G. Klein¹. 1) Center for Human Genetics and Laboratory Diagnostics, Martinsried, Germany; 2) Chair of Social Pediatrics, Technical University of Munich, Munich, Germany; 3) kbo Children's Center Munich, Munich, Germany; 4) Practice for Genetic Counseling and Psychotherapy, Augsburg, Germany.

Introduction: Due to extensive clinical and genetic heterogeneity, the causes of intellectual disability/ developmental delay (ID/DD) remain largely unknown. Whole-Exome Sequencing (WES) was shown to be an effective diagnostic strategy in patients with rare monogenic disorders. However, broad Next Generation Sequencing (NGS) approaches are still limited by time-consuming manual data interpretation. The Multiple Integration and Data Annotation Study (MIDAS) aims to accelerate NGS data analysis and to enhance the validity of the results by computer-based variant prioritization using patient's phenotype data. For this purpose, a central software system for data integration will be developed within MIDAS. **Methods:** By now, 51 patients (29 males, 22 females) with unexplained ID/DD and their unaffected parents (trios) were enrolled. All patients were evaluated by a clinical geneticist and phenotypic features were recorded in a standardized way using the Human Phenotype Ontology (HPO). Before enrollment, the patients had undergone extensive diagnostic testing (karyotyping, array-CGH, Fragile-X analysis), but none of these tests had led to a diagnosis. We performed trio WES and evaluated the identified variants using data from population (dbSNP, ExAC, gnomAD) and disease (HGMD, OMIM, ClinVar, Orphanet) databases. The remaining variants were examined by scientists and clinicians in the context of patients' clinical presentation. **Results:** By WES trio analysis, we were able to identify a causative variant in 23 of 32 yet analyzed ID/DD patients (diagnostic yield: 72%). In 19 of the 23 patients with positive WES, we identified a *de novo* variant in heterozygous state (9 loss-of-function and 10 missense variants). Additionally, we identified one homozygous missense variant each in two patients from different consanguineous families as well as one compound heterozygous variant and one homozygous nonsense variant in unrelated families. **Discussion:** WES is an effective diagnostic strategy for patients with ID/DD. Nonetheless, more than a quarter of our patients did not receive a diagnosis by WES although there is strong evidence for a genetic cause of their condition. For these patients, we will now perform a Copy Number Variation analysis based on their WES data and potentially Whole-Genome Sequencing analysis. NGS approaches, such as WES or WGS, lead to a rising amount of genetic data in diagnostics and necessitates powerful computational algorithms for improved data analysis.

2943W

Exome sequencing identified a *de novo* mutation of *PURA* gene in a patient with familial Xp22.31 microduplication. E. Rajcan-Separovic¹, Y. Qiao¹, H. Bagheri², F. Tang¹, C. Harvard³, S. Martell⁴, M.E.S. Lewis⁵, W. Robinson⁶, M. Connolly⁶, L. Arbour⁶. 1) Pathology and Laboratory Medicine, Children's and Women's Hospital of BC, Vancouver, Canada; 2) Dana-Farber Cancer Institute, Boston, USA; 3) Advanced Diagnostics Molecular Lab, University Health Network, Toronto; 4) Department of Medical Genetics, UBC, Vancouver, BC, Canada; 5) Division of Pediatric Neurology, Department of Pediatrics, UBC and BC Children's Hospital, Vancouver; 6) Department of Medical Genetics, University of Victoria, Victoria.

The clinical significance of Xp22.31 microduplication is controversial as it is reported in subjects with developmental delay (DD), their unaffected relatives and unrelated controls. We performed multifaceted studies in a family of a boy with hypotonia, dysmorphic features and DD who carried a 600 Kb Xp22.31 microduplication (7515787-8123310bp, hg19) containing two genes, *VCX* and *PNPLA4*. The duplication was transmitted from his cognitively normal maternal grandfather. We found no evidence of the duplication causing the proband's DD and congenital anomalies based on unaltered expression of *PNPLA4* in the proband and his mother in comparison to controls and preferential activation of the paternal chromosome X with Xp22.31 duplication in proband's mother. However, a *de novo*, previously reported (Tanaka et al, 2015, PMID 27148565) deleterious, missense mutation in Pur-alpha gene (*PURA*) (5q31.2), with a role in neuronal differentiation was detected in the proband by exome sequencing (NM_005859:c.563T>C; p.(Ile188Thr)). Phenotypically, the two cases shared many common features including severe developmental delay (speech and motor delay), ID, hypotonia, feeding difficulties, and facial dysmorphism such as dolicocephaly, hypertelorism and highly arched palate. We propose that the variability in the phenotype in carriers of Xp22.31 microduplication can be due to a second and more deleterious genetic mutation in more severely affected carriers. Widespread use of whole genome next generation sequencing in families with Xp22.31 CNV could help identify such cases.

2944T

Multi-step whole exome sequencing analyses in a cohort of 816 patients with malformative developmental anomalies and intellectual disability: A 6 years' summary of achieved diagnostic yield, new identified genes and undertaken strategies. F. Tran Mau-Them^{1,2,3,4,6}, A. Vitobello^{1,2,6}, A.L. Brue^{1,2}, S. Nambot⁶, F. Lecoquierre¹, S. Moutton^{2,3}, O.P.G. Orphanomix Physician's Group⁵, A.L. Mosca-Boiron¹, P. Callier¹, A. Sorlin^{1,2,3,4}, Y. Duffourd^{2,3,4,5}, C. Philippe^{1,2,4}, L. Faivre^{2,3,4}, C. Thauvin-Robinet^{1,2,3,4}. 1) Genetiques des Anomalies du Developpement, Laboratoire Génétique Moléculaire et Cytogénétique, Dijon, France; 2) Equipe EA4271 INSERM, LNC UMR1231 GAD, F-21000, Université de Bourgogne, Dijon, France; 3) Centre de Référence maladies rares « Anomalies du Développement et syndrome malformatifs » de l'Est et Centre de Génétique, Hôpital d'Enfants, CHU, Dijon, France; 4) FHU-TRANSLAD, Université de Bourgogne/CHU Dijon; France; 5) OrphanOmIX, Dijon, France; 6) These authors contributed equally to this work.

Introduction: Clinical whole exome sequencing (cWES) is thought to be the first-tier molecular test in heterogeneous rare diseases with suspected genetic etiology. Its diagnostic yield ranges from 30% to 50% depending on the applied strategy (singleton versus trio) and the investigated disorders. Laboratories performing cWES have to deploy smart strategies to identify causative variants in patients with negative exome results, taking into account economic and human resources issues. We present our last 6 years' experience in multi-step WES strategies applied to patients presenting with congenital anomalies and intellectual disability (MCA/ID) for a cost-effective reduction of the diagnostic odyssey. **Methods and results:** Initially, a group of 816 patients underwent OMIM-based singleton cWES analysis, which resulted in a positive molecular diagnosis in 239 cases (29.2%). Annual research reanalysis in 156 non-positive patients allowed additional diagnosis in 24 cases and the identification of 9 new genes involved in human disease. Knowing the burden of *de novo* variants in developmental delay disorders, we interrogated the denovo-db and DECIPHER databases seeking for evidences that could help us to reclassify the variants of unknown significance identified in our dataset. This complementary strategy permitted to identify 3 new candidate genes, which after reverse-phenotyping resulted in 3 additional diagnoses. Finally, we deployed trio cWES analysis in 63 syndromic patients with negative solo cWES, which resulted in the identification of 23 new candidate genes. International Matchmaker Exchange systems, allowed us to confirm the pathogenic role of the candidate variants in 15/63 patients (24%). Overall, these combined strategies led to the identification of 23 new genes responsible for MCA/ID. Our estimation is that trio analysis would lead a theoretical maximal diagnostic yield of 49%. **Conclusions and Perspectives:** The sequential deployment of singleton cWES, re-analysis and trio-based strategy represents an efficient strategy for diagnosing and deciphering molecular bases of MCA/ID. Annual re-analysis had a significant impact on the overall diagnostic yield, yet trio-based analysis greatly improved the selection of candidate variants while saving time. We plan to enhance such approaches through: i) introduction of a trio-like WES approach involving pooled parental DNA; ii) deployment of WGS coupled with RNA-seq to answer unexplained negative trio-like WES.

2945F

Trio Exome Sequencing of a patient with neurodevelopmental disorder and neonatal death identified a pathogenic variant in *CLPB* gene causing 3-methylglutaconic-aciduria type VII. C.M. Wilke¹, Y. Dincer^{1,2}, J. Schulz¹, M.Y. Cohen^{1,3}, S. Eilertz¹, S. Lippert¹, S.H. Eck¹, V. Malt^{1,3}, I. Rost¹, H.G. Klein¹.

1) Center for Human Genetics and Laboratory Diagnostics, Martinsried, Germany; 2) Chair of Social Pediatrics, Technical University Munich, Munich, Germany; 3) kbo Children's Center Munich, Munich, Germany.

Phenotype: Here we are presenting the case of a boy, who died on day seven of life exhibiting a variety of neurological and dysmorphological symptoms including developmental delay (DD), hypomyelination of the brain, seizures, muscular hypertonia, microcephaly, decreased body weight, short stature and craniofacial dysmorphism. Basic genetic testing (karyotyping, array-CGH, FISH) was negative. **Methods:** As part of the Multiple Integration and Data Annotation Study (MIDAS), we performed whole exome sequencing (WES) of the index patient and his healthy, unrelated parents (trio). The detected variants were evaluated using data from population (dbSNP, ExAC, gnomAD) and disease (HGMD, OMIM, ClinVar, Orphanet) databases. **Results:** By WES trio analysis, we identified a homozygous pathogenic variant in the gene *CLPB* (NM_030813.4:c.1249C/T, p.(Arg417*)), which was inherited in an autosomal-recessive pattern. This nonsense variant results in the introduction of a premature stop codon and truncation of the corresponding protein. *CLPB* is a mitochondrial protein that belongs to the ATPase superfamily. Variants in the *CLPB* gene are known to cause 3-Methylglutaconic-Aciduria Type VII, with cataracts, neurologic involvement and neutropenia (MEGCANN) syndrome, which is primarily characterized by increased levels of 3-methylglutaconic acid with neonatal or prenatal onset. The phenotype is highly variable, characterized by intellectual disability (ID), DD, progressive cerebral atrophy, seizures, movement abnormalities, microcephaly and neutropenia. As consequence of the progressive neurological symptoms, patients with severe phenotype died within the first months of life. Due to the rapid deteriorating health status of the newborn, different diagnostic tests were initiated simultaneously. WES led to the diagnosis MEGCANN, besides also metabolic testing revealed increased levels of 3-methylglutaconic acid. **Discussion:** Genetic and clinical heterogeneity in ID/DD represent a diagnostic challenge and many patients remain undiagnosed. Our data strongly support the value of WES as an effective diagnostic tool in patients with rare monogenic disorders. Over 70% of the MIDAS patient cohort received a diagnosis by WES, which is important for clarifying genetic risk factors for the affected families and for clinical management. In conclusion, the high diagnostic yield and the impact on therapeutic treatment qualifies WES as a standard set-up in diagnostics of complex genetic syndromes.

2946W

Genomic sequencing in a primarily trio-based cohort of 572 probands with developmental delay and intellectual disability. M.L. Thompson¹, K.M. Bowling¹, C.R. Finnila¹, S.M. Hiatt¹, M.D. Amaral¹, K.B. Brothers¹, K.M. East¹, W.V. Kelley¹, D.E. Gray¹, J.M.J. Lawlor¹, M.B. Neu¹, E.J. Lose², C.A. Rich², S. Simmons³, J. Krefting³, C. Holden³, R.M. Myers¹, G.S. Barsh¹, E.M. Bebin³, G.M. Cooper¹. 1) HudsonAlpha Institute for Biotechnology, Huntsville, AL; 2) University of Louisville, Louisville, KY; 3) University of Alabama at Birmingham, Birmingham, AL.

Over the past five years, the Clinical Sequencing Exploratory Research (CSER) project at the HudsonAlpha Institute for Biotechnology has focused on finding molecular diagnoses for children with unexplained developmental delay, intellectual disability, and related conditions (DD/ID). While we have previously published portions of this study, here we present an updated, comprehensive overview of data from 572 DD/ID-affected probands, 455 of which were sequenced as part of proband-parent trios. 80% of enrolled affected probands had received previous genetic testing, typically cytogenetic and/or microarray-based, and in many cases had also undergone extensive additional diagnostic testing. Overall, we generated exome data for 127 probands (365 total individuals) and genome sequencing data for 445 probands (1,115 total individuals). Pathogenic/likely pathogenic (P/LP) variants (SNVs, CNVs) were found in 29% of affected probands and variants of uncertain significance (VUS) were found in an additional 14%. Several measures have been taken to improve diagnostic yield. Systematic reanalysis of all data over the course of the study increased the P/LP rate from 23% to 29%. Furthermore, 77 genes harboring VUSs have been submitted to GeneMatcher. Thus far, 6% of these submissions led to publications describing new disease associations, including for variation in *EBF3*, *MYT1L*, *FGF12*, and *NTRK2*, and ongoing collaborations exist for an additional 18% of submitted genes. This study also assessed secondary findings within the 908 enrolled parent participants. 1.5% of parents who opted to receive secondary results harbored variation within one of the ACMG SFv2.0 genes and 1.8% harbored dominant P/LP variation in a non-ACMG gene. 70% of these participants self-reported family history or symptoms that are likely related to their genetic findings. The data underlying the results described here have been shared via ClinVar, GeneMatcher, and dbGAP. We find that WGS, especially when coupled to data sharing, is an effective first-choice diagnostic tool that advances clinical and research progress associated with pediatric neurological diseases.

2947T

A new case of “KAT6A Syndrome” in a 2-year-old Japanese female with intellectual disability and multiple minor anomalies. Y. Watanabe^{1,2}, M. Kinoshita¹, K. Fukui¹, H. Mishima³, A. Kinoshita³, K. Yoshiura³, Y. Yamashita¹.

1) Pediatrics and Child Health, Kurume University, Kurume, Japan; 2) Res Institute of Medical MS, Kurume University School of Medicine, Kurume, Japan; 3) Dept. Human Genetics, Nagasaki University, Graduate School of Biomedical Sciences, Nagasaki, Japan.

Introduction: The *KAT6A* gene encodes a lysine acetyltransferase, a member of the MYST family of histone acetyltransferases (HATs). *KAT6A* is involved in gene-specific H3 acetylation regulating developmental gene expression and in p53 acetylation and signaling, both of which have been known to control a wide range of processes including DNA repair and replication and metabolism. Since whole-exome sequencing (WES) has been introduced, mutations in *KAT6A* have been identified in individuals with intellectual disability (ID) of unknown etiology. Facial characteristics including broad nasal tip, large/low set posteriorly rotated ears, thin upper lip, down-turned corners mouth, and bitemporal narrowing are reported in the individuals, which may suggest that “KAT6A syndrome” is clinically recognizable. **Case Study:** The proband was a 2 1/2-year-old Japanese female born to a 24-year-old mother and a non-consanguineous 20 year-old father, both of whom were in good health. She was a full term (37w) and her height, weight, and OFC were 39cm (-3.6SD), 1670g (-3.2SD), and 28.2cm (-3.3SD), respectively. Epicanthal folds, hypertelorism, broad nasal tip, thin upper lip, low set/posteriorly rotated anomalous ears, and micrognathia were noted as an infant. She was kept in the NICU for 60 days for persistent pulmonary hypertension, respiratory distress, and feeding difficulty requiring a tube feeding for 30 months. A VP shunt was placed at age 4 months for hydrocephaly. Renal ultrasound study at age X months showed the left polycystic dysplastic kidney. Her developmental milestones are: head control at 3 Mo, rolls over at 10 Mo, sits without support at 12 Mo, crawls at 18Mo, cruises at 20 Mo, and walks at 29 Mo. She does not have a single word. The Trio WES (propositus and her parents) showed a de novo variant, c.3070C>T (p.R1024X) in *KAT6A* in the proband, confirming the diagnosis. **Discussion:** Since Arboleda et.al. and Tham et. al. independently reported individuals with intellectual disability with *KAT6A* mutations in 2015, there have been at least 20 cases reported to date. Three Japanese siblings (Satoh et.al.) and one patient of unknown ethnicity (Arboleda et.al.) had the identical *KAT6A* mutation found in the propositus. Neither the CNS cystic lesions nor polycystic dysplastic kidney were reported in the those individuals. Owing to lacking specific clinical and laboratory findings, WES may be the most practical tool in making a diagnosis of “KAT6A syndrome”.

2948F

Diagnostic flow-chart to improve the detection rate and to uncover the genetic heterogeneity in Pitt-Hopkins syndrome: Experience on 61 subjects. M. Zollino¹, S. Frangella¹, S. Ricciardi¹, D. Orteschi¹, C. Pantaleoni², P. Bonanni³, M. Macchiaiolo⁴, M. Murdolo¹, P. Doronzio¹, S. Lattante¹, G. Maranghi¹. 1) Institute of Genomic Medicine, Catholic University, Rome, Rome, Italy; 2) Developmental Neurology Unit, C. Besta Neurological Institute Foundation, Milan, Italy; 3) Epilepsy and Neurophysiology Unit, Scientific Institute, IRCCS Eugenio Medea, Treviso, Italy; 4) Rare Diseases and Medical Genetics, Bambino Gesù Children's Hospital Rome, Italy.

Pitt-Hopkins syndrome (PTHS, OMIM #610954) is a neurodevelopmental disorder caused by the haploinsufficiency of the *TCF4* gene. It is clinically defined by the association of severe intellectual disability, severe speech impairment, postnatal microcephaly, a typical facial gestalt, breathing abnormalities, subtle brain MRI alterations, and typical behavioral features. The PTHS phenotype overlaps with different conditions, mainly including Angelman (AS), but also Mowat-Wilson, Christianson and Rett syndromes. In the past nine years, a total of 158 subjects were analyzed both clinically and genetically because of clinical suspicion of PTHS. The diagnostic procedure included the following steps: 1) clinical evaluation of patients' phenotypes 2) direct sequencing and MLPA analysis of *TCF4*; 3) MS-MLPA for 15q11q13 (AS region) 4) NGS analysis of a panel of genes responsible for diseases included in the differential diagnosis with PTHS (*TCF4*, *CNTNAP2*, *NRXN1*, *UBE3A*, *SLC9A6*, *MECP2*, *CDKL5*, *FOXG1*, *MEF2C*, *ZEB2*, *ATRX*); 5) whole exome sequencing (in a total of 12 patients); 6) standard karyotyping or FISH analysis. The genetic diagnosis was obtained in 61 subjects. The identified variants were: *TCF4* intragenic variants (36); large 18q21 deletions (2); partial *TCF4* deletion (2); balanced de novo translocations (2); *MECP2* variants (6); *UBE3A* variants (4); *ZEB2* variants (2); *FOXG1* variants (3); *EHMT1* variant (1), and *SZT2* variant (1). We could also define a preliminary intragenic phenotypic map of the *TCF4*, allowing for the prediction of the functional consequences of variants causing the selective dysfunction of specific protein domains. Based on our results, clinical implications and diagnostic strategies will be discussed.

2949W

Dissecting TSC2-mutated renal and hepatic angiomyolipomas in an individual with ARID1B-associated intellectual disability. C. Zweier¹, A. Agaimy², C. Kraus¹, K.X. Knaup³, A.B. Ekici¹, S. Uebe¹, A. Reis¹, M. Wiesener³, B. Popp¹. 1) Institute of Human Genetics, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany; 2) Institute of Pathology, University Hospital Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany; 3) Department of Nephrology and Hypertension, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany.

Several subunits of the SWI/SNF chromatin remodeling complex are implicated in both cancer and neurodevelopmental disorders (NDD). Though there is no clinical evidence for an increased tumor risk in individuals with NDDs due to germline variants in most of these genes so far, this has been repeatedly proposed and discussed. A young woman with NDD due to a *de novo* variant in *ARID1B* now presented with a large renal (>19 cm in diameter) and multiple hepatic angiomyolipomas (AMLs) but no other signs of tuberous sclerosis. We analyzed tumor and healthy tissue samples with exome and panel sequencing. Additionally to the previously known, germline *ARID1B* variant we identified a post-zygotic truncating *TSC2* variant in both renal and hepatic AMLs but not in any of the healthy tissues. We did not detect any further, obvious tumor driver events. The identification of a passenger variant in *SIPA1L3* in both AMLs points to a common clonal origin. Metastasis of the renal AML into the liver is unlikely on the basis of discordant histopathological features. Our findings therefore point to very low-grade mosaicism for the *TSC2* variant, possibly in a yet unknown mesenchymal precursor cell that expanded clonally during tumor development. A possible contribution of the germline *ARID1B* variant to the tumorigenesis remains unclear but cannot be excluded given the absence of any other evident tumor drivers in the AMLs. This unique case highlights the blurred line between tumor genetics and post-zygotic events that can complicate exact molecular diagnoses in patients with rare manifestations. It also demonstrates the relevance of multiple disorders in a single individual, the challenges of detecting low-grade mosaicisms, and the importance of proper diagnosis for treatment and surveillance.

2950T

Diagnostic utility of genotype data from SNP arrays in neurodevelopmental disorders. J.K. Nicholl¹, C. Barnett², J. Liebell², A. Correll¹, S.L. Kang¹, Y. Hull¹, A. Attwood¹, S. Smith¹, T. Hardy¹, S. Yu¹. 1) Genetics and Molecular Pathology, SA Pathology, North Adelaide, South Australia, Australia; 2) Paediatric and Reproductive Genetics, South Australian Clinical Genetic Service, Women's and Children's Hospital, North Adelaide, South Australia, Australia.

Since 2015, 6,500 patients with neurodevelopmental disorders have been tested using one of two Illumina CytoSNP array platforms: the 300K 12 BeadChip or the Infinium -850K BeadChip. Both platforms detect not only copy number variations but also genotype data. We have reviewed the clinical utility of copy number neutral loss of heterozygosity (LOH) in this cohort. From a total of 21 cases, further molecular investigation has led to a genetic diagnosis in nine cases so far. **Regions of LOH led to the diagnosis of an autosomal recessive (AR) disorder.** 7 cases were flagged for molecular follow up. For two cases, the clinical diagnosis of an autosomal recessive disorder was established with a candidate gene located in a region of LOH. Further DNA sequencing of the candidate gene identified a homozygous pathogenic mutation: a missense mutation in the *LPL* gene in a patient with lipoprotein impairment; a mutation of the *NPHS1* gene in a patient with nephrotic syndrome. No causative mutation was identified in one case. Investigations of four cases are in progress. **Regions of LOH led to the diagnosis of various types of Uniparental isodisomy (UPD) or an AR disorder.** Two cases showed UPD involving an entire chromosome. Methylation MLPA confirmed UPD15 of paternal origin, consistent with the clinical diagnosis of Angelman syndrome in the first case. In a second patient the diagnosis of Retinitis pigmentosa was confirmed following the finding of an homozygous mutation involving the *TULP1* gene on chromosome 6. Two cases showed mosaic segmental isodisomy of chromosome 11 involving the Beckwith Wiedemann syndrome (BWS) region. Molecular analysis confirmed that the segmental UPD11 is paternal in origin, consistent with BWS. Ten cases with a single or multiple LOH involving a single chromosome were detected. In two cases, microsatellite analysis of the trio showed iso-heterodisomy of chromosome 7 consistent with Silver-Russell syndrome. Maternal hetero-isodisomy of chromosome 4 was seen in a patient with a homozygous mutation in the *IDUA* gene supporting the clinical diagnosis of Hurler syndrome. For one patient with a 37Mb LOH on chromosome 20, which was shown to be identity by descent (IBD), the *GNAS* gene was in a region of bi-parental inheritance. Three patients showed bi-parental inheritance outside of the region of LOH. Three are in progress. This review demonstrates the clinical utility of SNP arrays to detect autosomal recessive and imprinting disorders.

2951F

High resolution microarray testing in 420 individuals with developmental disorders in the south of Brazil. M. Ocampos^{1,10}, T.F. Chaves^{2,11}, I.T. Barbatto^{3,10}, L.F. Oliveira^{4,11}, G.R. De Luca^{5,12}, J.H. Barbato Filho^{6,12}, L.L.C. Pinto^{7,12}, P. Bernardi^{8,13}, A.F. Maris^{9,11}. 1) Biologist, PhD in Biotechnology and Molecular Biology; 2) Biologist and doctoral student in cell biology and development; 3) Biologist and Master of Chemical Engineering; 4) biomedic and doctoral student in cell biology and development; 5) Medical geneticist and PhD in Child and Adolescent; 6) Medical geneticist and pediatrician; 7) Medical Neuropediatric, Children's Hospital Joana de Gusmão, Florianópolis, Santa Catarina, Brazil; 8) Medical geneticist; 9) Biologist, PhD in molecular biology and genetics, university professor in the department of cell biology, embryology and genetics; 10) Neurogene Laboratory, Florianópolis, Santa Catarina, Brazil; 11) Federal University of Santa Catarina, Florianópolis, Santa Catarina, Brazil; 12) Children's Hospital Joana de Gusmão, Florianópolis, Santa Catarina, Brazil; 13) University Hospital Professor Polydoro Ernani de São Thiago, Florianópolis, Santa Catarina, Brazil.

Introduction: Currently, chromosomal microarrays are recommended as first tier in the evaluation of copy number variations (CNVs) in individuals with neurodevelopmental disorders (NDD). However, in developing countries such as Brazil, classical cytogenetic tests are still the most used in clinical practice as well as in academic research, as reflected by the scarcity of publications of microarray investigation in larger cohorts. We report here the diagnostic yield of microarrays in a cohort of 420 individuals with NDDs from the south of Brazil and investigate which phenotypes relate to a higher diagnostic answer. **Methods:** The microarrays were carried out from 2013 to 2016 through the clinical diagnostic laboratory Neurogene, Florianópolis, SC Brazil. The platforms used were the Affymetrix chip CytoScan HD (75%) and 750 K (25%). The analysis was performed on ChAS software (Affymetrix®) and with information gathered from public and private databases. The collection of clinical data of the participants was through a semi-structured questionnaire to the referring physicians. The study was approved by the ethics committee of the Joana de Gusmão Children's Hospital. **Results:** Of the 420 participants, 62% were male, 38% female, mean age of 9.5 years. A total of 96 pathogenic CNVs were found in 18% of the participants; 12% presented as the main finding variants of uncertain clinical significance and 70% showed no relevant CNVs. The main reasons for referral for the test were developmental delay (DD), intellectual disability (ID), facial dysmorphism and autism spectrum disorder (ASD). Of the participants, 67% presented DD, 41% have ID (many of which also had DD), together accounting for 80% of the participants with DD and/or ID; 52% had facial dysmorphism and 32% presented ASD. According to univariate analysis, some phenotypes in this population are predictive of higher diagnostic yield, as follows: obesity (*p-value*= 0.006, OR= 0.20), short stature (*p-value*= 0.032, OR= 0.44), abnormalities and genitourinary malformations (AGM) (*p-value*= 0.032, OR= 0.63) and ASD (*p-value*= 0.039, OR= 1.94). **Conclusion:** The diagnostic rate in this study was 18%, within the range reported in the literature (15-20%). In NDD cohorts, ASD and less frequent phenotypes such as obesity, short stature and AGM could be predictive phenotypes with higher diagnostic yield. However, for more reliable results on phenotypic data, a standardized revaluation and a larger sample are advisable.

2952W

Experience and challenges of variant classification of the DMD gene. H. Cheng^{1,2}, J. Dong², C. Gijavanekar², E. Schmitt^{1,2}, S. Peacock^{1,2}, S. Wen^{1,2}, J. Scull^{1,2}, Z. Chen^{1,2}, F. Xia^{1,2}, L. Meng^{1,2}. 1) Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030; 2) Baylor Genetics, 2450 Holcombe Blvd, Houston, TX 77021.

Background: DMD gene is involved in a wide spectrum of diseases, ranging from severe Duchenne/Becker muscular dystrophy, to X-linked dilated cardiomyopathy and intellectual impairment. Interpretation of DMD variants can be facilitated by clinical presentations, family history and functional studies. However, rules for classifying variants identified in asymptomatic females through reproductive risk screening has not been well-established. **Method:** Following ACMG variant classification guideline, we established evidence-based gene-specific guidelines for DMD variants classification. Key factors include: 1) variant type, 2) previously reported patients and their clinical presentations, and 3) frequencies in the general population. Furthermore, special rules were applied for DMD copy number variants (CNVs), including: 4) reading frames, 5) protein domain and the expected impact on DMD function, 6) similar or adjacent CNVs previously reported, and 7) exon skipping. We applied more stringent criteria to DMD exon duplications due to the uncertainty of their genomic location. **Results:** A total of 82 unique CNVs were detected in female individuals seeking DMD carrier testing, including 50 exonic deletions, 31 exonic duplications and 1 complex rearrangement. Comparing to deletions (74.0% pathogenic, 20.0% likely pathogenic, and 6.0% variant of unknown significance (VUS)), smaller portion of duplications were considered clinical significant (32.2% likely pathogenic, 67.8% VUS). A total of 116 individuals were determined to be carrier of a likely pathogenic or pathogenic DMD variant. Deletions were detected in 78 individuals (67.2%), with exon 45 to 51 being hotspots. Duplications were detected in 8 cases (6.9%), SNVs including truncating variants, splicing variants and rare missense variants were detected in 29 cases (25%); a complex arrangement was detected in 1 individuals (0.9%). **Conclusion and Discussion:** We developed gene-specific guidelines for interpreting DMD variants, and successfully applied them in clinical testing of general population carrier screening. DMD variant interpretation remains challenging for several reasons, such as unknown breakpoints, limited CNV control data, and significant frequency of private variants. Detailed and extensive collection of clinical evaluation, family segregation analysis, and functional studies in families with DMD variants are highly desirable to further address the more complicated variants.

2953T

Redefining and refining neuromuscular disease diagnostics: Implications for molecular diagnosis and treatment. A. Ankala^{1,2}, J. Alexander^{1,2}, J. Xie^{1,2}, H. Askree^{1,2}, M. Hegde^{1,3}, R. Nallamilli², A. Kesari², L. Bean^{1,2}. 1) Emory University, Atlanta, GA; 2) EGL Genetics LLC, Tucker, GA; 3) Perkin Elmer Genomics, GA.

With the advent of next generation sequencing (NGS) the traditional gene-by-gene approach for molecular diagnosis for some of the highly heterogeneous disorders like neuromuscular diseases (NMDs) has become a rarity. Today, NGS panel testing has become the first-tier testing for molecular diagnosis of NMDs; thereby eliminating the diagnostic odyssey for many individuals and families and allowing for faster and cost-effective diagnosis. While research based NGS testing has facilitated identification of new disease-causing genes, clinical diagnostic testing has uncovered new dimensions and previously unrecognized modalities of disease causation for some of the long known and well-studied genes. One classic example is that of the *DMD* gene where a complete spectrum of clinical presentation ranging from 'classic' Duchenne phenotype in a male to an early onset muscular dystrophy in a female has been observed. Having reported over 10,000 NMD sequencing cases EGL Genetics has identified atypical clinical presentations and diagnostic findings. Utilizing genotype-phenotype correlation and collective analysis of multiple unrelated case-cohorts, we have been able to identify and establish new disease causing variants including in-frame deletion variants in *CAPN3* causing a newly recognized autosomal dominant calpainopathy, mosaic loss-of-function variants in *DMD* causing an unusual milder phenotype and several 'leaky' variants predicted or shown to result in reduced protein expression and consequently a milder and later onset diseases. These data, in conjunction with other recently published reports, demonstrate the variability of clinical features that define NMDs - from congenital muscular dystrophies to early onset DMD and from adult onset LGMD and to some of the very late onset muscular dystrophies. In addition, our large-scale analysis brings into question the evidence for some previously reported genes and diseases, such as *SYNE1* and *SYNE2* that have been associated with an autosomal dominant muscular dystrophy. In this era of clinical therapeutics, where several clinical trials and therapies are emerging for genes such as *DMD*, finding the causative gene and genotype in a timely manner is often critical for enrollment of patients into relevant clinical studies. In conclusion, our data show how single genes like *DMD* and *COL6A2* can present with a spectrum of clinical phenotype and underlines why NGS based testing should be the first-tier test for NMDs.

2954F

Advancing genetic diagnosis of facioscapulohumeral muscular dystrophy via next generation genome mapping technology. Y.-W. Chen^{1,2}, H. Barseghyan¹, S. Sen Chandra¹, S. Bhattacharya¹, M. Almalvez¹, B. Dufault¹, E. Vilain^{1,2}. 1) Center for Genetic Medicine Research, Children's National Health System, Washington, DC. USA; 2) Department of Genomics and Precision Medicine, School of Medicine and Health Science, George Washington University, Washington, DC, USA.

Facioscapulohumeral muscular dystrophy (FSHD) has a prevalence of 1:7,000-20,000 in the general population and is recognized as one of the most common forms of neuromuscular disorders. FSHD is primarily (95% of cases) caused by a deletion of the D4Z4 macrosatellite repeats (each unit 3.3 kb long) on chromosome 4p35, resulting in 1-10 D4Z4 repeats (healthy range 11-100). The D4Z4 region in healthy individuals has been shown to be transcriptionally repressed. This is evident by the epigenetic marks associated with heterochromatin and DNA hypermethylation in the region. Studies have shown the region becomes transcriptionally de-repressed and DNA demethylated when the number of D4Z4 repeats is shortened to 10 units or less. Current standard genetic testing for FSHD is to detect contracted D4Z4 array using Pulsed-Field Gel Electrophoresis (PFGE) in combination with Southern Blotting, which is time consuming, labor intensive and not precise. Here, we used a novel approach, optical mapping, to identify the copy number of D4Z4 repeats as well as the methylation state of the D4Z4 array region. Optical mapping captures a pattern of fluorescent labels in native-state, megabase size DNA molecules in nanochannel arrays for *de novo* genome assembly and SV calling. This technology offers substantial advantages over the current clinical diagnostic practice as it has been successfully used for identification of SVs in Duchenne muscular dystrophy. We incorporated dual high molecular weight DNA labeling for genome assembly and epigenetic methylome profiling for FSHD diagnosis. First, long DNA molecules were nicked/labeled with BssSI endonuclease and red fluorescent nucleotides, respectively. Second, the same DNA molecules were treated with M. TaqI methyltransferase that attached green fluorescent cofactor onto non-methylated CpGs. Imaging and analysis of the sample library revealed the methylation pattern as well as the number of D4Z4 repeat units. We successfully identified the molecular diagnosis in our sample cohort and show that optical mapping enhances the clinical diagnostic workflow for FSHD patients.

2955W

Detection of mitochondrial heteroplasmic variants in a single clinical whole genome sequencing assay. A. Chawla¹, E. Dolzhenko¹, B. Lajoie¹, R.T. Hagelstrom¹, V. Rajan¹, A. Warren¹, F. Mullen¹, K. Bluske¹, A. Islip¹, J. Le¹, S.S. Ajay¹, A. Pizzino², O. Sherbini², A. Vanderver², D.L. Perry¹, J.W. Belmont¹, D.R. Bentley¹, R.J. Taft¹. 1) Illumina, Inc, San Diego, CA; 2) Children's Hospital of Philadelphia, PA.

Mitochondrial variants are routinely assessed by Sanger sequencing, quantitative PCR and long-range PCR followed by massively parallel sequencing (LR-PCR/MPS). These techniques are often performed in addition to other molecular investigations, adding complexity to test cadence and clinical care. The TruGenome Undiagnosed Disease Test offered by the Illumina Clinical Services Laboratory (ICSL) is intended to identify the underlying cause of a genetic condition through a single clinical whole genome sequencing (cWGS) assay. The current test definition targets single nucleotide variants (SNVs), small insertions and deletions, and copy number variants (CNVs) greater than 10 kb in the nuclear genome. We have extended the test definition to include detection of SNVs across the mitochondrial genome at a complete range (0-100%) of heteroplasmic frequencies. To validate mitochondrial variants, we first used DNA from four Coriell cell lines. Libraries were prepared using the Illumina TruSeq PCR-free kit and sequenced on the Illumina HiSeq X instrument with an average autosomal coverage of 30x and mitochondrial coverage of ~9000x. We employed two approaches to confirm the accuracy of alignment, variant detection and heteroplasmy estimation: 1) an orthogonal confirmation performed by an external laboratory using the LR-PCR/MPS assay and 2) a serial dilution series performed within ICSL. With the first approach, cWGS and LR-PCR/MPS detected a total of 104 mitochondrial SNVs in the four Coriell cell lines including the four known pathogenic variants. All SNVs had a maximum variant read fraction (VRF) difference of 5% between the two assays. In the mixture experiments, we performed serial dilutions using two of the Coriell cell lines to generate heteroplasmic variants with a wide range of VRFs. All VRFs across all variants and all dilutions were found to lie within 5% of their expected value. The first clinically significant mitochondrial variant detected by our cWGS pipeline, m.3243A>G, is a known pathogenic variant, detected in a 19-year-old female showing symptoms of mitochondrial leukoencephalopathy. The proband was estimated to carry the variant at 44% heteroplasmy, whereas the unaffected mother carried the same variant at only 16% heteroplasmy, compared to the population mean VRF of approximately 0.01%.

2956T

Clinical utility of Dual Genome NGS panel for mitochondrial disorder: Towards comprehensive understanding about disease etiology. H. Dai¹, Y.M. Feng², Y.J. Jiang², E. Gorman², H. Yu¹, L.S. Toon², S. Chen², J.L. Zhang^{1,2}, X. Wang^{1,2}, Z. Chen^{1,2}, H. Mei^{1,2}, C. Eng^{1,2}, L.J. Wong^{1,2}, Y. Wang^{1,2}. 1) Baylor College of Medicine, Houston, TX 77030, United States; 2) Baylor Genetics, Houston, TX 77021, United States.

Dual genome NGS panel targeting both mitochondrial and nuclear genes simultaneously is a critical approach for full molecular diagnosis of mitochondrial disorders due to the wide variability of clinical presentations linked with genetic heterogeneity throughout both the nuclear and mitochondrial genomes. Since we launched the first dual genome NGS panel in 2012, we have accumulated over 1,500 cases both diagnosed and unsolved; in the process, we have witnessed variant re-classifications, panel expansion, sensitivity improvement for detecting variants in the mitochondrial genome, and many unexpected findings requiring in-depth analysis. Through our experiences, we have found three points of great interest: Firstly, the number of solved cases seems to be evenly matched between mitochondrial and nuclear etiology, suggesting that disregarding nuclear genes when diagnosing suspected mitochondrial diseases risks missing some 50% of diagnoses. Secondly, in addition to the genes previously reported to be highly prevalent in mitochondrial disorders, we also report a high diagnostic rate from defects in the FBXL4 gene, which is an important finding as it is a relatively new gene. Finally, the vast improvements in NGS sensitivity to mutation load in mitochondrial genes makes diagnosis of mitochondrial diseases involving heteroplasmy possible by NGS panel analysis. In conclusion, the dual-genome NGS panel approach provides a comprehensive tool for diagnosis of mitochondrial diseases.

2957F

Analysis of mtDNA mutations in Serbian patients with Leber's hereditary optic neuropathy. P.G.A. Dawod^{1,2}, B. Rovacanin², M. Brankovic^{2,3}, A. Marjanovic^{2,3}, M. Jankovic^{2,3}, I. Novakovic^{2,3}, F.I. Abdel Motaleb¹, J. Janic^{2,4}, V. Kostic^{2,3}. 1) Medical Biochemistry and Molecular Biology, Faculty of Medicine, Ain Shams University, Cairo, Egypt; 2) Faculty of Medicine, University of Belgrade; 3) Neurology Clinic, Clinical Center of Serbia; 4) Child and Adolescent Neurology and Psychiatry Clinic, Belgrade, Serbia.

Introduction: Leber hereditary optic neuropathy (LHON) is a mitochondrial inherited degeneration of retinal ganglion cells (RGCs) leading to severe visual impairment or even blindness. Besides primary mtDNA mutations number of secondary changes affect phenotype. Aim of this study was to investigate the presence of mutations in mtDNA among Serbian patients affected with LHON. **Material and Methods:** Individuals included in this study were recruited from the Child and Adolescent Neurology and Psychiatry Clinic and from Neurology Clinic CCS, Belgrade, Serbia. All examined individuals for 17 unrelated families had characteristic clinical presentation suggesting the presence of LHON. Multiple segmental PCR amplicons of mtDNA have been sequenced by Sanger's method and the obtained results were compared with the referent mtDNA sequence. **Results:** Substitutions mt.3460 G>A in ND1, mt.11778 G>A in ND4, and mt.14484 T>C in ND6, were found as primary mutations in 6/17, 10/17 and 1/17 families, respectively. The most frequent primary mutations mt.3460 G>A and 11778 G>A were associated with secondary mutations (4216C/ND1-10398G/ND3-13708A/ND5-15257A/CYB-MT-15812A/CYB-MT). These changes have been well established in European haplogroup J that increase penetrance of LHON. Many other mutations were found in various parts of mitochondrial genome. **Conclusion:** This study of spectrum of mutations in mtDNA in Serbian LHON patients assists understanding of their role in etiopathogenesis and progression of disease.

2958W

Plasma derived cell-free mitochondrial DNA (cf-mtDNA): A novel and non-invasive method to identify the mitochondrial DNA haplogroup. C. Newell¹, S. Hume², S.C. Greenway^{2,5}, L. Podemski², J. Shearer³, A. Khan^{1,5}. 1) Department of Medical Genetics, Cumming School of Medicine, University of Calgary, Alberta, Canada; 2) Department of Medical Genetics, University of Alberta, Alberta, Canada; 3) Department of Cardiac Sciences, Libin Cardiovascular Institute of Alberta, University of Calgary, Calgary, Alberta, Canada; 4) Faculty of Kinesiology, University of Calgary, Alberta, Canada; 5) Department of Pediatrics, Alberta Children's Hospital Research Institute, Cumming School of Medicine, University of Calgary, Alberta, Canada.

Background: Mitochondrial disease comprises a heterogeneous group of diseases affecting 1:5000 people and can be caused by gene mutations in either nuclear or mitochondrial DNA (mtDNA). Less-invasive methods to obtain mtDNA have involved collection from peripheral blood leukocytes, urine sediment and buccal swab. However, these tissues contain low levels of mitochondria and may not adequately represent commonly affected tissues such as the brain and skeletal muscle. Recently, research suggests that cellular apoptosis and cellular turnover release small 150-200 base pair fragments of DNA from hematopoietic and non-hematopoietic cell types, which can be extracted from plasma as cell-free DNA (cfDNA). Entire mtDNA has not been successfully identified from the cell free fraction previously. As we suspect that the circular nature of mtDNA is protective, our objective was to develop a methodology to identify entire cf-mtDNA from plasma. **Methods:** Blood samples from patients with mitochondrial disease diagnosed from skeletal muscle biopsy (n=7) and healthy controls (n=7) were collected and their cfDNA isolated. To demonstrate the presence of the mitochondrial genome within these samples, we amplified the isolated DNA using custom PCR primers specific to overlapping fragments of mtDNA. cfDNA samples were then sequenced using the Illumina MiSeq sequencing platform. **Results:** We confirmed the presence of mtDNA, demonstrating that the full mitochondrial genome is in fact present within the cell-free plasma fraction. Furthermore, sequencing of the cfDNA identified a mitochondrial haplogroup match with the tissue specimen for all patients. **Conclusion:** We report a successful method showing positive identification of mtDNA by haplogroup analysis from cell-free plasma. This technique has clinical application to diagnose mitochondrial disease, evaluate mtDNA haplogroup variants and response to future transplantation or mitochondrial therapies.

2959T

The complex genetic architecture of mitochondrial disease: Exploration of 1500 cases by whole exome sequencing. S.L. Stenton^{1,2}, B. Alhaddad^{1,2}, C. Chang^{1,2}, T. Haack^{1,2}, S. Wortmann^{1,2,3}, J. Mayr³, B. Büchner⁴, P. Freisinger⁵, C. Makowski⁶, D. Rokicki⁶, R. Taylor⁷, F. Fang⁸, K. Murayama⁹, Y. Okasaki¹⁰, D. Ghezzi¹¹, C. Lamperti¹¹, A. Rötig¹², T. Strom^{1,2}, T. Klopstock⁴, T. Meitinger^{1,2}, H. Prokisch^{1,2}. 1) Institute of Human Genetics, Technische Universität München, Munich, Germany; 2) Helmholtz Zentrum München, Neuherberg, Germany; 3) Paracelsus Medical University, Salzburg, Austria; 4) LMU, München, Germany; 5) Klinikum Reutlingen, Reutlingen, Germany; 6) The Children's Memorial Health Institute, Warsaw, Poland; 7) Newcastle University, Newcastle, UK; 8) Beijing Children's Hospital, Beijing, China; 9) Chiba Children's Hospital, Chiba, Japan; 10) Juntendo University Graduate School of Medicine, Tokyo, Japan; 11) Neurological Institute 'Carlo Besta', Milan, Italy; 12) INSERM, Université Paris Descartes, Paris, France.

Mitochondrial diseases pose a diagnostic challenge due to clinical and genetic heterogeneity, propelling unbiased whole exome sequencing (WES) into the early diagnostic setting. Over 300 disease-associated genes implicated in mitochondrial energy metabolism are currently recognised, and this number continues to grow. Through global collaboration, initiated by the European network for mitochondrial diseases, GENOMIT, we assimilate data from patients investigated by WES under clinical and/or biochemical suspicion of mitochondrial disease. Systematic analysis of WES data from 1500 paediatric and adult patients established a genetic diagnosis in 735 (49%). Furthermore, this endeavour, and subsequent follow-up studies, resulted in molecular diagnosis of over 1000 patients and included discovery of 40 novel disease genes. Pathogenic variant(s) in 459 (31%) patients harboured within 138 mitochondrial disease-associated genes, with defects in *ACAD9*, *ECHS1*, *FBXL4* and *MT-ATP6* occurring most frequently. To date, over 40 cofactor metabolism defects have been discovered, leading to impaired mitochondrial energy metabolism amenable to specific cofactor supplementation. And indeed, 92 of the solved exome cases fall into this category. However, effective defect-targeted treatments for the majority of mitochondrial diseases are missing. Thus, genotyping is fast becoming the prerequisite for efficacious treatment and clinical trial inclusion. Remarkably, 33 patients with molecularly-verified mitochondrial disease demonstrated normal respiratory chain complex activity, highlighting difficulty in disease definition. This is further reflected by 276 patients harbouring disease-causing variant(s) in non-primarily mitochondrial disease-associated genes (179 genes including *TANGO2*, *CAD*, *IARS*, *SLC52A2*) and 64 patients with unsuspected but genetically-proven mitochondrial disease diagnosed in the same time-period. This study exemplifies the value of international collaboration in the rare disease setting. By coupling exomes to clinical, biochemical and experimental data, we further elucidate the heterogeneous genetic architecture of mitochondrial disease. Prompting careful consideration of diagnostic criteria, by emphasising discrepancies between clinical, biochemical and genetic means of diagnosis and demonstrating the need to consider this entire body of evidence in aggregate. GENOMIT is hence establishing a valuable global registry for both clinical and genomic data.

2960F

Long-read nanopore sequencing of human mitochondrial DNA. A. Trimouille^{1,2}, C. Mouden³, C. Boury⁴, J. Van Gils¹, A. Courrèges¹, P. Fergelot^{1,2,5}, M.L. Marin-Négrier^{1,6}. 1) CHU Bordeaux, Service de Génétique Médicale, F-33000 Bordeaux, France; 2) Univ. Bordeaux, INSERM 1211, F-33000 Bordeaux, France; 3) INRA, UMR Biodiversity of Genes and Ecosystems, Transcriptomics and Genomics Platform, Cestas, France; 4) CHU Bordeaux, Service de Pathologie, F-33000 Bordeaux, France; 5) Plateforme Génome Transcriptome, Centre de Génomique Fonctionnelle de Bordeaux, Université de Bordeaux, Bordeaux, France; 6) Univ. Bordeaux, UMR 5293, F-33000 Bordeaux, France; CNRS, Institut des Maladies Neurodégénératives, UMR 5293, F-33000 Bordeaux, France.

Mitochondrial cytopathies can be caused by mitochondrial DNA (mtDNA) anomalies such as point variants or large-scale single deletions. In addition to these causative anomalies, multiple deletions of mtDNA can arise due to pathogenic variants in nuclear genes involved in the maintenance of mtDNA. These 3 types of anomalies (point variants, single and multiple deletions) are investigated for diagnostic purpose thanks to amplification of mtDNA by 2 long-range PCR, and sequencing by NGS short reads technology. However, the use of short reads does not allow to easily and accurately identify the deletion breakpoints at the nucleotide level. In order to determine if long read sequencing technologies overcome this limitation, we perform mtDNA sequencing of controls and patients with single or multiple deletions on Oxford Nanopore MinION. This technology allowed an accurate identification at the nucleotide level of the breakpoints of the deletions, even in the case of complex multiple deletions. These breakpoints have been confirmed by Sanger sequencing. This determination is useful for family testing in the case of single deletion. Furthermore, studies could also take advantage of this method to decipher the mechanisms of occurrence and to track the accumulation of mtDNA multiple deletions in mitochondrial or degenerative diseases, but also during normal aging. In addition, despite a higher per base error rate in comparison with other NGS technologies, the accuracy of Oxford Nanopore MinION sequencing is sufficient to correctly determine the mitochondrial haplogroup of each patient. Due to the portability of MinION, it raises the possibility of in-the-field sequencing experiments for population genetics studies.

2961W

Targeted next-generation sequencing facilitates genetic diagnosis and provides novel pathogenetic insights into deafness with enlarged vestibular aqueduct. C-C. Wu^{1,2}, Y-H. Lin^{1,3}, Y-H. Lin¹, C-S. Chen², T-C. Liu¹, C-J. Hsu¹, P-L. Chen^{2,3}. 1) Department of Otolaryngology, National Taiwan University Hospital, 7 Chung-Shan South Road, Taipei, Taiwan; 2) Department of Medical Genetics, National Taiwan University Hospital, 7 Chung-Shan South Road, Taipei, Taiwan; 3) Graduate Institute of Medical Genomics and Proteomics, National Taiwan University College of Medicine, 1 Jen-Ai Road Sec.1, Taipei, Taiwan.

Enlarged vestibular aqueduct (EVA) is an inner ear malformation associated with sensorineural hearing impairment. The majority of EVA are associated with Pendred syndrome (PS) and non-syndromic DFNB4, two autosomal recessive disorders caused by mutations in *SLC26A4*. However, a significant percentage of EVA patients cannot have confirmed diagnosis by screening common *SLC26A4* mutations, constituting an enigma in genetic diagnosis. To enable comprehensive genetic examination and explore the etiologies of EVA, we designed a next-generation sequencing (NGS) panel targeting the entire length of three PS/DFNB4 genes (*SLC26A4*, *FOXI1* and *KCNJ10*) and exons of ten other genes related to EVA, and performed genetic testing in 50 EVA families without confirmative results on screening for *SLC26A4* hotspots (c.919-2A>G and p.H723R). We identified bi-allelic *SLC26A4* mutations in 34 families and *EYA1* mutations in two families, yielding a diagnostic rate of 72% (36/50). In addition, we identified two variants in *KCNJ10* and *FOXI1*, but our findings did not support previous hypothesis that mutations in these two genes are probable contributors to EVA through recessive inheritance or digenic inheritance with *SLC26A4*. Of note, a large *SLC26A4* deletion was confirmed in one step using our panel. Our results demonstrate the utility of NGS-based panel in addressing EVA families by identifying various types of gene mutations with satisfactory diagnostic yields, and provide novel insights into the pathogenesis of EVA.

2962T

Detection of stereocilin (*STRC*) gene deletions by multiplex digital droplet PCR. R.L. Margraf¹, R. Mao^{1,2}. 1) ARUP Institute for Clinical and Experimental Pathology, Salt Lake City, UT; 2) Department of Pathology, University of Utah, Salt Lake City, UT.

Stereocilin gene (*STRC* [MIM 606440]) mutations are a common cause of moderate autosomal recessive hearing loss (MIM 603720). 11% of all individuals with isolated non-syndromic hearing loss have a large *STRC* deletion, which can encompass the entire gene. The *STRC* gene is 98% homologous with its pseudogene (*pSTRC*), which includes a 100% homologous section from exons 1 to 15. *STRC* deletions cannot be accurately detected using NGS, microarray, or MLPA due to the *pSTRC* pseudogene homology. To specifically detect large *STRC* deletions, digital droplet PCR (ddPCR) assays were developed on the Bio-Rad QX200™ Droplet Digital™ PCR system. Four target assays were designed to three locations within *STRC* (Exon 19, near Exon 23, and Intron 25) and one area within the nearby *CATSPER2* gene (MIM 607249). These assays were designed to avoid common polymorphisms, and to be specific to *STRC* by covering bases that are divergent between *STRC* and *pSTRC*. Also, a stable two copy gene was used as a reference assay for calculating *STRC* copy number. Wild type (WT) sample with two *STRC* copies was tested to choose the pre-ddPCR DNA fragmentation method, and had copy number values ranging from 1.2-1.9 for sonication versus 1.9-2.1 for enzyme digestion. The restriction enzyme digestion was chosen over sonication due more precise copy number values. To test multiplexing, two target assays were combined in various ratios, along with the reference, in one reaction. Multiplexing results were very sensitive to the mixing ratios of the two target assays, since droplet data points can start to overlap at some ratios. At the chosen optimal mixing ratio, the multiplexed assays had a WT copy number average of 1.94 +/- 0.05 copies and heterozygous for deletion samples had 0.99 +/- 0.03 copies. WT and heterozygous samples correctly genotyped at a range of DNA input (7.5ng - 75ng/reaction). Samples with known and suspected homozygous *STRC* deletions amplified for the reference assay, but did not amplify for the target assays even at 2-4x higher than the recommended 37.5ng DNA input, indicating the ddPCR assays were specific to *STRC* versus *pSTRC*. The ddPCR assays were successfully multiplexed in sets of two target assays with one reference assay, and generated the expected copy number for the known genotyped samples. The addition of this *STRC* ddPCR assay will increase the clinical sensitivity of the NGS hearing loss panel test by allowing for the detection of large deleterious *STRC* deletions.

2963F

Identification of novel CNV in *PTPRQ* associated with autosomal recessive hearing loss by target exome sequencing. Z. Wei¹, F. Chang¹, Y. Pan¹, J. Li¹, X. Tian¹, F. Mu¹, S. Xu¹, Y. Li¹, P. Fang¹, Y. Yuan². 1) WuXi NextCode Genomics (Shanghai) Co., Ltd; 2) Department of Otolaryngology, Head and Neck Surgery, PLA General Hospital, Beijing, P. R. China.

Next generation sequencing is widely used in the diagnosis of inherited disease. However, the detection of copy number variants (CNV) is not implemented in most next-generation sequencing pipelines. To detect single nucleotide variants (SNV), indels and CNV simultaneously, we designed a target exome sequencing based-panel, which contains 139 genes related with hereditary hearing loss. Using this panel, three variants in *PTPRQ* in a non-syndromic sensorineural hearing loss patient are identified, including one novel large deletion. *PTPRQ* is recognized as a receptor-like phosphatase to be active against phosphatidylinositol phosphates, which is involved in the cell regulation, proliferation. Pathogenic variants in *PTPRQ* will cause progressive moderate-to-profound sensorineural hearing impairment and vestibular dysfunction (DFNB84), however, no CNV of this gene has been reported before. We here first identified a paternally inherited large deletion spanning exon 42 of *PTPRQ* (NM_001145026.1), which was called by HMM pipeline of NextGNEe and then validated by long range PCR. Sequencing of PCR product further confirmed the breakpoint of deletion (c.6453+131_6602+90delinsTTTGT). This indel is located in the phosphatase domain and predicted to cause a single exon loss of *PTPRQ*. Loss of function is considered as a known mechanism of pathogenicity for *PTPRQ*, and null variant has been reported previously in the same region, thus the CNV is regarded as a pathogenic variant. We also identified another two maternally inherited SNVs in this patient including one missense (c. 1124C>T) and one splicing variant (c.6603-3T>G). Both of them are very rare in normal population without healthy homozygote reported before. The missense is predicted to have no impact on gene product, whereas the splicing variant is predicted to affect mRNA splicing. However, more evidences are required to evaluate the pathogenicity of these two SNVs. In conclusion, CNV is an important cause of *PTPRQ* related hearing loss, which must be included in the genetic testing; compared with other methods, such as array-CGH, target exome sequencing is a cost effective, time saving and high resolution approach for CNV detection in clinical diagnosis.

2964W

Homozygous non-coding sequence change of the *OCA2* causing oculocutaneous albinism in a child. A. Maher^{1,2}, A. Tsai^{1,2}. 1) University of Colorado Denver, Aurora, CO; 2) Children's hospital Colorado.

Oculocutaneous albinism type 2 (*OCA2*) is characterized by hypopigmentation of the skin and hair and the characteristic ocular changes found in all types of albinism, including nystagmus; reduced iris pigment with iris translucency; reduced retinal pigment with visualization of the choroidal blood vessels on ophthalmoscopic examination; foveal hypoplasia associated with reduction in visual acuity; and misrouting of the optic nerve fiber radiations at the chiasm, associated with strabismus, reduced stereoscopic vision, and altered visual evoked potentials (VEP). Most non-Africans with *OCA2* are compound heterozygotes. Approximately one third of the reported non-African individuals have only one pathogenic variant detectable by sequence analysis of the coding region, the intron-exon boundaries, and several hundred bases of the 5' promoter region and 3' untranslated region of *OCA2* [Oetting & King 1999, King et al 2001a, King et al 2001b, Gargiulo et al 2011]. To date, no pathogenic variant has been reported beyond 20 bp into the intron region of an exon/intron boundary. Approximately 20% of all individuals who have complete sequencing of the entire coding sequences and the flanking intron sequences have shown only one pathogenic variant. Cryptic splice sites or variants in regulatory regions could affect transcriptional regulation [King et al 2001a]. We herein report a case of a 6 month old with no family history of albinism who presented with the characteristic symptoms of *OCA2* including hypopigmentation of the skin and the hair. Eye examination revealed transillumination defects 360 bilaterally, with foveal hypoplasia and blonde fundus bilaterally. With normal extraocular muscles movement without strabismus or other refractive errors. In addition to nystagmus secondary to foveal hypoplasia. NGS Pigmentation panel, testing for 30 genes including the *OCA1*, *OCA2*, *HPS* genes, *PAX3* and *SOX10* didn't reveal any variants in the genes of the pigmentation panel. By using TaqMan qPCR that identified a homozygous non coding sequence change of the *OCA2* gene that was identified by the lab as likely pathogenic. Both parents had lighter pigmentation and sun sensitivity. Additional study on the mechanism of this non-coding region change is underway.

2965T

WGS reveals hidden novel variants in ocular disease genes, *MIP* and *NHS*. A.S.L. Ma^{1,2}, A.E. Minoche³, J. Grigg^{1,4}, M. Flaherty⁵, G. Ho^{2,6}, D.J. Amor⁷, A. Cheng^{1,2}, B. Bennetts^{2,6}, M.J. Cowley^{3,8}, M.E. Dinger^{3,8,9}, R.V. Jamieson^{1,2,4}. 1) Eye Genetics Research Unit, Sydney Children's Hospital Network, Children's Medical Research Institute, University of Sydney, Sydney, Australia; 2) Discipline of Genetic Medicine, CHW Clinical School, Sydney Medical School, University of Sydney, Sydney, Australia; 3) Kinghorn Centre for Clinical Genomics, Garvan Institute for Medical Research, Sydney, NSW, Australia; 4) Save Sight Institute, University of Sydney, Sydney, Australia; 5) Department of Ophthalmology, The Children's Hospital at Westmead, Sydney; 6) Sydney Genome Diagnostics, Sydney Children's Hospital Network, Westmead, Sydney, Australia; 7) Murdoch Children's Research Institute and University of Melbourne Department of Paediatrics, Royal Children's Hospital, Melbourne, Australia; 8) St Vincent's Clinical School, UNSW Sydney, Sydney, Australia; 9) Genome.One, Sydney, Australia.

Introduction Congenital cataracts demonstrate marked clinical and genetic heterogeneity, often with additional ocular and extraocular features. We have previously obtained a genetic diagnosis in a high proportion of patients utilising an NGS-based panel or whole exome sequencing (WES) approach and CGH microarray, finding answers in over 70% (33/46) of congenital cataract patients. **Methods** In this study we applied whole genome sequencing (WGS) to assist with 13 congenital cataract cases unsolved after CGH microarray, and panel or WES. WGS utilized the Illumina HiSeq X Ten and a bioinformatics structural variant analysis pipeline, ClinSV, which allows confident detection of rare structural- and copy number variants, using evidence from split-reads, discordant pairs and depth-of-coverage. **Results** A causative variant was found in 2/13 patients, who harboured deletions too small to be seen on conventional microarray or capture-based sequencing. A novel 3.2kb deletion in the *MIP* gene was found in a patient with congenital cataracts and microphthalmia. A 4kb deletion in the X-linked syndromal gene *NHS* was found in a male with congenital cataracts and coloboma. Subsequent study with high-resolution chromosomal microarray was able to confirm the presence of both these deletions. **Discussion** This study demonstrates an increased yield of genetic diagnoses in a cohort of patients. While the yield of CGH microarray combined with panel/exome based testing is significant, WGS analysis revealed additional diagnoses of structural variants, providing new diagnostic and management information.

2966F

NGS surprise: Pathogenic haplotypes. I.T. Perva^{1,2}, A. Chirita-Emandi^{1,2}, M. Puiu^{1,2}. 1) Medical Genetics, Center of Genomic Medicine, Victor Babes University of Medicine and Pharmacy, Timisoara, Romania; 2) Medical Genetics Compartment, Louis Turcanu Emergency Hospital for Children, Timisoara, Romania.

Aim. To present a rare next generation sequencing interpretation approach and raise awareness to evaluate also the haplotype level of a VCF document. **Materials and methods.** Ophthalmologic evaluation of a 7 years old boy revealed severe bilateral hypermetropia (+7 diopter) and discromia. Maternal uncle presents the same eye phenotype. NGS analysis WES was performed for index and using the online databases the variants were called (Genome Aggregation Database (gnomAD), Exome Sequencing Project (ESP), 1000Genome project (1000G), Online Mendelian Inheritance in Man (OMIM) and CentoMD). **Results.** The following relevant variants were detected, in the third exon of the OPN1LW gene: c.457A>C p.(Met153Leu); c.511_513delinsATT p.(Val171Ile); c.532A>G p.(Ile178Val). Individually these have no clinical relevance, but combined, the three define a toxic opsin haplotype and subsequently determine the X linked Bornholm eye disease (OMIM®: 300843). The same haplotype was found in the maternal uncle, by WES analysis. **Discussions.** The defining positions are 153, 171, 174, 178 and 180, and the haplotype name is given by the amino acid letter code corresponding to these coordinates. The LIAVA variant combination was previously reported with third exon skipping and lack of opsin mRNA, thus interpreted as pathogenic significance. **Conclusion.** Current NGS interpretation pipelines do not, generally, include haplotype variant interpretation, doing so various diseases can be overlooked in diagnosis. By the presentation of this case, we hope that at least in the face of unsolved cases, some specialists, will evaluate a combined variant option. Funding: Development of Existing Infrastructure and Creation of New Infrastructure (Laboratories, Research Centers) POSCCE-A2-O2.2.1-2013-1, in the Center of Genomic Medicine of the University of Medicine and Pharmacy 'Victor Babes' Timisoara.

2967W**Clinical and molecular characterization of hereditary retinal degeneration in Portugal.** V. Peter¹, L. Paris², M. Quinodoz¹, K. Cisarova¹, A. Veiga², S. Meira², O. Afonso², AB. Sousa³, R. Rodrigues³, LC. Santos², C. Rivolta^{1,4}.

1) Department of Computational Biology, Unit of Medical Genetics, University of Lausanne, Lausanne, Switzerland; 2) Instituto de Oftalmologia Dr. Gama Pinto, Lisbon, Portugal; 3) (1)Medical Genetics of the Paediatrics' Department, Hospital de Santa Maria/ Centro Hospitalar Lisboa Norte – Lisbon – Portugal; 4) Department of Genetics and Genome Biology, University of Leicester, Leicester, United Kingdom.

Hereditary retinal degenerations (HRDs) represent a class of diseases leading to the progressive degeneration of the retina and, in the end, to loss of vision and blindness. HRDs are genetically highly heterogeneous and missing heritability accounts for ~40% of all clinically-diagnosed cases. Most importantly, the mutational spectrum of HRDs can vary widely depending on the population studied, and be highly dependent on the degree of genetic homogeneity of the cohort analyzed. Since there are currently no comprehensive data concerning either the spectrum of HRD mutations or the presence of population-specific disease genes in the Portuguese, we performed a cross-sectional study combining clinical and molecular assessment of 83 families from this country, affected by a variety of HRD diagnostic classes. SNP-chip genotyping of selected subjects from 57 families enabled the detection of at least one known HRD gene as most probable cause of the disease in 28 cases (49%). In 20 cases (35%) we identified regions of homozygosity that may harbor new genetic associations and are currently performing WES analyses to detect the actual pathological mutations. We also found that the degree of autozygosity in 57 unrelated subjects from our cohort (median=88 Mb) is significantly higher than in 361 Portuguese controls (median=9 Mb; p-value = 1.9×10^{-17}), which is itself higher than in 3,797 subjects from 6 European countries (Switzerland, Italy, France, Spain, Germany, Belgium; median=4 Mb; p-value = 4.15×10^{-9}). Moreover, we identified a haplotype of 9.5 Mb shared by 4 patients and harboring a deletion in the *EYS* gene, representing 10% of cases with autosomal recessive retinitis pigmentosa (n=39) in our cohort. To our knowledge, this is the first study assessing the cross-sectional genotype-phenotype landscape of HRDs in the Portuguese population. Our preliminary data reveal an unexpected level of cryptic consanguinity as one of the main contributors to the molecular pathology of these diseases, as well as the presence of at least one prevalent and population-specific mutational event for this part of Europe.

2968T**Elevated incidence of microphthalmia in Chuuk, FSM – a mystery**

solved? A.S. Schneider¹, A.A. Yomai², A.V. Levin³, D. McNear², S. Kopinsky¹, J. Capasso³, J. Chaing⁷, M. Hegde⁶, S. Shankar², T.M. Glaser². 1) Genetics, Einstein Medical Center Philadelphia, Philadelphia, PA; 2) University of California, Davis, Davis, CA; 3) Wills Eye Hospital, Philadelphia, PA; 4) Department of Health Services, Chuuk, FSM; 5) Independent Educational Consultant, Cambridge, MN; 6) Perkin Elmer Genetics, Pittsburgh, PA; 7) Molecular Vision Laboratory, Hillsboro, OR.

Microphthalmia, anophthalmia and coloboma (A/M) are rare ocular birth defects with a complex genetic basis and environmental risk factors. The incidence of A/M is elevated 4-5 fold in Chuuk state, Federated States of Micronesia, affecting 14 per 10,000 births (Yomai *et al.* 2010) compared to the worldwide average of 3 per 10,000 (Shah *et al.* 2011). This remote archipelago in the western Pacific contains 42 inhabited islands, which were peopled 1000-3000 years ago. The population is dispersed, but most residents live in a ring of barrier islands surrounding the Chuuk lagoon. The elevated incidence of A/M and population structure in Chuuk are consistent with a genetic founder effect. Moreover, Vitamin A deficiency (VAD), a known maternal risk factor for A/M disease, is endemic in Chuuk, suggesting a potential gene x environment interaction. To investigate the clinical features and etiology of A/M disease in this population, we held clinics in Chuuk during April 2017. We performed eye and medical exams, and obtained family history data and genomic DNA from 25 probands with suspected A/M and their parents from 7 different islands. Five probands from one island were found to have LCA (Leber's congenital amaurosis) with enophthalmos (recessed eye globes) and a homozygous inactivating mutation in *CRB1* (c.3134delT). The remaining 20 probands (8M, 12F) were determined to have colobomatous microphthalmia. Some exhibited poor growth and microcephaly, but were not dysmorphic, and most have normal intellect. The majority of mothers with affected children reported night-blindness during pregnancy, consistent with VAD, and we found that serum retinol and RBP levels were moderately or severely reduced in 78% (14 of 18 mothers tested). The segregation ratio for the A/M trait in 22 affected sibships was 21% (24/117), consistent with autosomal recessive or digenic inheritance. To identify gene mutations, we performed high resolution SNP analysis on all samples and whole genome sequencing (WGS) on 5 probands. Although several pathogenic sequence variants were discovered, no solitary region of homozygosity >5Mb was identified that is shared among all probands. These data exclude a single recessive founder mutation and suggest a polygenic basis for A/M disease in Chuuk, with maternal VAD as a likely contributing factor.

2969F

Cytogenomic findings in Brazilian patients with OAVS. P.A.S. Abu Hana^{1,2}, H. Sakata¹, C.P. Oliveira³, R.S. Bonadio¹, I. Ferrari¹, H.P.N. Safatle¹, M.S. Cordoba¹, M.T.A.S. Rosa³, C. Rosenberg⁴, E.L. Freitas⁴, R.E. Pogue⁵, A.C. Acevedo-Poppe¹, A. Pic-Taylor¹, S.F. Oliveira¹, J.F. Mazzeu¹. 1) University of Brasilia, Brasilia, Brazil; 2) Health Sciences University of Alagoas, Maceio, Alagoas, Brazil; 3) Health Department of the Federal District, Brasilia, Brazil; 4) Department of Genetics and Evolutionary Biology, Institute of Biosciences, University of Sao Paulo, SP, Brazil; 5) Catholic University of Brasilia, Brazil.

Oculoauriculovertebral spectrum (OAVS) (OMIM 164210) is a developmental defect involving the first and second branchial arch derivatives and intervening first pharyngeal pouch and branchial cleft, in particular the ear, mouth, mandible, eye and cervical spine. We present a genomic screening for copy number variations (CNVs) in OAVS Brazilian patients, referred to the Clinical Genetics Service – University of Brasilia, aiming to identify genomic regions and genes potentially involved in this disease pathogenesis. For this purpose, genomic DNA from venous blood samples from these patients was extracted, followed by Chromosomal microarray analysis (CMA) with Affymetrix or Agilent platforms. The study included 24 isolated and six familial cases of OAVS. CMA detected copy-number variants in 12/30 (40%) patients including four deletions (33.3%), five duplications (41.6%), two segments of absence of heterozygosity (AOH) (16.6%) and one complex rearrangement involving a deleted and a duplicated region, both “*de novo*”, in the 2p23.3-p25.3 region. Three alterations were considered pathogenic including a partial deletion at 22q11.2 (including *TBX1*), a microdeletion at 12q24.21, involving *TBX3* and a microduplication at 20p12.3-p12.2 with partial duplication of *PLCB4*. Nine CNVs were classified as VOUS. These results confirm the heterogeneity of the syndrome and show that chromosome imbalances are an important cause of OAVS.

2970W

Genomics and clinical association study in duplication of 5p. L.A.R. Bicu-do¹, M.V.P.F. Santos¹, R.M.O.F. Curado¹, S.J. Sestari¹, G.F. Vasconcelos¹, S.S. Coelho², N.A. Bergamo¹, B.F. Gamba¹. 1) Federal University of Goias, Goiania, GO; 2) University of Sacred Heart, Bauru, SP.

Duplication of 5p (5p trisomy) is a rare genetic condition associated, or no, with complex rearrangements involving other chromosomes. Phenotype expression in patients with trisomy 5p is highly variable and the clinical condition as well as its severity depend on the size and location of the duplication. In the present work we report a chromosome 5p duplication case as well as an association between chromosomal 5p breakpoints and clinical annotation based on 16 cases from literature. Chromosome 5p region, ensemble about 46 Mb of DNA which involves 241 genes (110 OMIM ® genes). Wide phenotype variability, totaling 88 shared clinical findings in all patients were annotated. The data collected were added in Gephi software to establish relationship between patients, 5p breakpoint and clinical finding. We call this analysis as intuitive analysis. According to the wide phenotypic variability and different breakpoints in the 5p chromosome region found in the present group, our intuitive analysis showed a remarkable genetic-clinical interface in four groups of individuals. Clinical findings such as facial dimorphism, hypertelorism, low ears, broad nasal bridge, microretrognathia, global developmental delay, hypotonia, language delay, ocular anomalies, respiratory difficulties, dolichocephalic, macrocephaly, arachnodactyly, crooked feet were the most shared clinical signs among all patients. This result evidences the existing phenotypic variability when we studied several cases with a chromosomal region alteration. As we observed wide phenotypic variability, broad genetic heterogeneity was also observed, but some duplicate genes are potentially responsible for the clinical findings in trisomy 5p, like *CLPTM1L*, *SLC1A3*, *NIPBL* and *GDNF* genes. These genes are present in duplicated region identified in the patient and may be responsible for the observed phenotype. We conclude that genotype-phenotype correlation it's still difficult challenge to geneticists, but the use of computational tools can be useful in this type of strategy and potentiate the analyzes. Our results point to four groups of patients that share a larger number of data (genetic and clinical), which may direct future analysis and allow better understanding of the 5p chromosomal region.

2971T

Evaluation of copy number variation (CNV) in deceased patients with previously unknown molecular etiology. C. Chung¹, K. Stumm¹, D. Watson¹, M. Harri¹, M. Khan¹, J. Glessner¹, H. Hakonarson^{1,2,3}, A. Santan^{1,4}. 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Department of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA; 3) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 4) Department of Pathology, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

Background. Recently, posthumous genetic testing has begun to shed light on the cases of patients who had unexplained deaths. The aim of this study is to analyze copy number variation (CNV) of individuals in our pediatric cohort with unknown molecular etiology and to analyze the potential relationship between the genetic testing results and the patients' phenotype and cause of death. **Methods.** Patients for this study were previously enrolled in a research biorepository at the Center for Applied Genomics at Children's Hospital of Philadelphia. Subjects were selected by a search of their medical records with ICD-9/10 codes for sudden infant death syndrome (SIDS; ICD-9: 798.0) and unknown cause of mortality (ICD-10: R99), excluding those with cancer diagnoses (ICD-10: C00-D09). Fifty-four deceased patients were identified for this study (61.1% male, 38.9% female). Thirty-one patients had ante-mortem whole genome microarrays and microarray studies were subsequently performed on an additional twenty-three cases. A custom analysis pipeline was used to call, annotate and analyze the CNV data. Results from the arrays were then compared to the clinical features of each patient in order to analyze a possible association. **Results.** The average age of the patients' death was 6.2 years (ranged from 1 hour of life to 23 years); their phenotypes included multiple congenital anomalies (68.5%), neurodevelopmental disorders (42.6%), craniofacial disorders (11.1%) and skeletal dysplasia (1.8%); some presented overlapping features. Autopsy was performed in 35.2%. Family history that increased suspicion for a genetic etiology was reported in 14.8%. Pathogenic or VUS CNVs were identified in 26% of the 54 patients. Clinical correlation was performed by a clinician and in total, five CNVs were determined to be potentially related to the cause of death. **Conclusion.** Genome analysis can help to elucidate causes of death that originally had an unknown etiology and to provide important information to the family members, such as a potential diagnosis and recurrence risk. This data further supports the utility of genome analysis in deceased patients with molecularly undiagnosed genetic conditions, therefore comprehensive molecular testing should be recommended on all patients with SIDS or unexplained cause of mortality.

2972F

9q34.11 deletion, including *DNM1* and *SPTAN1* but lacking *STXBP1*, causes a distinctive phenotype with intellectual disability, speech delay, and dysmorphic facial features. K. Kurosawa¹, N. Harada¹, T. Saito¹, Y. Enomoto¹, Y. Tsurusaki¹, H. Murakami¹, Y. Kuroda¹, M. Masuno². 1) Division of Medical Genetics, Kanagawa Children's Medical Center, Yokohama, Japan; 2) Genetic Counseling Program, Kawasaki University of Medical Welfare, Kurashiki, Japan.

9q34.11 deletion syndrome is a newly recognized syndrome characterized by epileptic encephalopathy, hereditary hemorrhagic telangiectasia, and variable degree of intellectual disability (Campbell et al., 2012). These phenotypes are caused by haploinsufficiency of the combinatory of the dosage sensitive genes including *ENG*, *STXBP1*, *SPTAN1*, and *TOR1A*. Large deletion encompassing the 9q34.11 provides a typical phenotype including severe intellectual disability, early infantile spasm. However, genotype-phenotype correlation of each gene in the region is poorly understood. Here we report a patient with severe intellectual disability, dysmorphic facial features, but no seizure episode by the age of 8 years. Microarray revealed 2.0 Mb deletion including the 10 OMIM genes from *DNM1* to *TOR1A*. Among the deleted genes, only the *DNM1* and *SPTAN1* show high score (1.0) of loss of function intolerance (pLI). For *SPTAN1* mutation, hemizyosity has been suggested to result in hypomyelination, although dominant negative effect causes early onset infantile spasm. For *DNM1* mutation, effect of hemizyosity is unknown, although dominant negative effect causes epileptic encephalopathy. Further, mutation in the PH domain of *DNM1* has been suggested to cause non-epileptic phenotype with dysmorphic facial features and developmental delay. These results indicate that the deletion of 9q34.11, including *DNM1* and *SPTAN1* but lacking *STXBP1*, causes a distinctive phenotype including dysmorphic facial appearance and ID.

2973W

Detection of pericentric inversion of Y chromosome in an infant with Pierre Robin syndrome: A case report. R. Pattarkine, C. Kumar. Westchester medical center, valhalla, NY.

Sox 9 (SRY box-9) is a protein coding gene located on chromosome 17; It acts during chondrocyte differentiation and regulates transcription of the anti-Muellerian hormone (AMH) gene. Deficiencies lead to the skeletal malformation syndrome Campomelic dysplasia, Pierre robin syndrome and frequently with sex reversal. Here we present a case of a male baby born to a mother with phenylketonuria at 38 weeks of gestation with multiple dysmorphic features and Pierre robin syndrome. Upon further evaluation multiple abnormalities like anemia, low fibrinogen, elevated PT/PTT, LDH, liver enzymes and bilirubin were identified. Because of the phenotypic abnormalities, the chromosomal microarray and the cytogenetic investigations were indicated. The microarray analysis performed on peripheral blood did not identify any clinically relevant copy number changes or regions with absence of heterogeneity. Chromosome analysis performed on GTG banded metaphases prepared from cultured lymphocytes shows baby's karyotype as 46,X,inv(Y)(p11.2q?11.23~q12). The Fluorescence in situ hybridization (FISH) test was performed to characterize the inversion breakpoints of inv(Y) chromosome. FISH analysis using Y-chromosome specific probes (alpha satellite (DYZ3)/satellite III (DYZ1) and LSI SRY probes) revealed that one copy of sex determining region (SRY) probe was in the expected position, one copy of centromeric probe in the metacentric position whereas two copies of the Y-heterochromatin probe, one at the expected Yq region and the other at the short arm of the Y chromosome confirming the pericentric inversion of Y chromosome with breakpoints on the short arm at p11.2 and on the long arm at the proximal heterochromatin region of q12. Paternal blood chromosome analysis could not be performed to confirm if this rearrangement is familial or de novo in origin. The pericentric inversion of Y chromosome is a heritable normal variant and may be associated with any phenotypic manifestations or fertility problems. In some rare cases, reduced fertility or phenotypic abnormalities related to sex determination are associated with deletion or disruption of genes located in the inverted area. The FISH results confirmed that the baby had a pericentric inversion characterized as 46,X,inv(Y)(p11.2q12). Finally, our question, whether specific inversion breakpoints are indicative for impaired fertility, remains open as no clear genotype-phenotype correlation has been shown up till now.

2974T

Cerebral white matter abnormalities in 6p25 deletion syndrome. I.M.

Bader¹, I. Reindl², G. Sander¹, G. Kronberger¹, C. Rauscher², J. Koch², J. Atzwanger², O. Rittinger², E. Haschke-Becher², W. Sperl². 1) Division for Clinical Genetics, Paracelsus Medical University Salzburg, Austria; 2) Department of Pediatrics; Paracelsus Medical University, Salzburg, Austria; 3) University Institute of Medical and Chemical Laboratory Diagnostics, Salzburg, Austria.

We report the phenotype and genotype of a 4 years and 11 months old boy who was transferred to hospital after an accident. CT-scan and cMRI showed a cranial fracture as a result of the accident and two incidental findings: 1) large bifronto-temporal arachnoidal cysts and 2) multifocal white-matter lesions involving the periventricular, deep, and subcortical cerebral white matter. The clinical phenotype of the boy included mild developmental delay, strabismus, slightly dysmorphic features and mild hearing impairment. Molecular karyotype from a blood sample detected a ~ 3,9 Mb heterozygous microdeletion in 6p25.3 - 6p25.2 deleting 34 genes, 18 of which were OMIM annotated including the transcription factor *FOXC1*. Microdeletions in 6p25 are thought to be associated with a distinctive clinical phenotype – the 6p subtelomeric deletion syndrome – that includes sensorineural hearing loss, anterior chamber eye defects, cardiac defects, developmental delay and other developmental and behavioral abnormalities, hypotonia, hip dysplasia, cerebellar abnormalities and a characteristic facial appearance. Few reports exist that describe multifocal T2-weighted and FLAIR abnormalities involving the periventricular, deep, and subcortical cerebral white matter associated with the 6p25 deletion in young children and in adults. From our case and from the review of the literature we conclude that a microdeletion in 6p25 should be considered as the cause for unclear multifocal cerebral white matter abnormalities in patients that in addition have dysmorphic features and multiple congenital anomalies even when intellectual abilities are within the normal range (Vernon et al. 2013).

2975F

Consecutive medical exome analysis at a tertiary center: Diagnostic and health-economic outcomes. *R. Kosaki¹, M. Kubota², T. Uehara³, H. Suzuki³, T. Takenouchi³, K. Kosaki³.* 1) Divisions of Medical Genetics ,National Center for Child Health and Development, Tokyo, Japan; 2) Divisions of Neuology ,National Center for Child Health and Development, Tokyo, Japan; 3) Center for Medical Genetics, and Department of Pediatrics, Keio University School of Medicine, Tokyo, Japan.

Background: The utility of whole exome analysis has been extensively demonstrated in research settings, but its clinical utility as a first-tier genetic test has not been well documented from diagnostic and health economic standpoints in real-life clinical settings. Both pediatricians and health policy makers are in need of such data. **Methods:** We performed medical exome analyses focusing on a clinically interpretable portion of the genome (4813 genes) as a first-tier genetic test for 360 consecutive patients visiting a genetics clinic at a tertiary children's hospital in Tokyo, Japan, over a 3-year period. Bioinformatics analyses were conducted using standard software including BWA, GATK, and SnpEff. **Results:** A molecular diagnosis was made in 171 patients (47.5%) involving a total of 107 causative genes. Among these 107 causative genes, 57 genes (53%) were classified as genes with potential organ-specific interventions and management strategies. Hence, clinically relevant results were obtained in approximately a quarter of the patients (92 out of 360). The estimated cost for a singleton medical exome analysis per patient was comparable to the average cost of brain magnetic resonance imaging (MRI) scans and G-band chromosome analyses spent prior to participation in this study, even though these analyses were non-contributory to the diagnosis.

2976W

Genetic testing in pediatric dentistry: A new way of thinking about oral health care. *E. Severin¹, R. Baltag².* 1) Carol Davila University of Medicine and Pharmacy, Bucharest, Romania; 2) Profiladent, Bucharest, Romania.

Background. By definition, genetic testing is a type of medical test that identifies changes in genes, chromosomes, or proteins. The results of a genetic test can confirm or rule out a suspected genetic condition or help determine a person's chance of developing or passing on a genetic disorder. Understanding variability between individuals allows for more targeted or personalized health-care based on genetic differences. Literature review provided many examples of gene mutations that cause isolated inherited dental anomalies, mutations with pleiotropic effects and of genetic testing for oral cancer susceptibility and presymptomatic diagnosis of periodontal disease. The great majority of dental health specialists recognize the importance and potential of genetics and genomics in the future, but little has been done to incorporate genetic testing extensively in clinical practice. **Objective.** In this paper, we discuss the current status of genetic testing properly applied in pediatric patients with oral health problems. We highlight successful strategies for gene identification into our dental clinical practice. **Methods and Patients.** Children with syndromic oro-dental anomalies were investigated. Clinical and radiographical assessments were performed, followed by detailed investigation into the personal and family medical history of the patients with cherubism, cleidocranial dysplasia, osteogenesis imperfecta and Fabry disease. Molecular tests were performed as well. **Results.** The research focused on four rare monogenic disorders caused by mutations in SH3BP2, RUNX2, COL1A2, and GLA genes. The selected monogenic diseases are known to involve oral and dental anomalies in permanent dentition. Different severity of oro-dental phenotypes were found. Micrognathia was associated with dental anomalies and malocclusion. Tooth number anomalies were noticed in two forms, agenesis and multiple supernumerary teeth. Enamel hypoplasia, premature and delayed eruption, and dental size anomalies were noticed as well. The patterns of oral and dental anomalies revealed previously described features, but also individual findings. **Conclusion:** Genetics has the potential to shape the dental education and practice. In the context of precision medicine, the dentists have to integrate the tools of genetics in their dental practice for prediction, prevention and personalized dental therapy. .

2977T

An amplicon panel for the detection of Noonan syndrome (NS). *M. Shafiekhani, R. Hrdlickova, J. Nehyba, D. Fox, R. Blidner.* Bioo Scientific, a PerkinElmer Company, Austin, TX.

A class of human genetic syndromes, RASopathies, is caused by germline mutations in genes of the Ras/mitogen activated protein kinase (MAPK) pathway. This pathway plays an important role in the regulation of the cell cycle, differentiation, and growth; therefore, it is at the center of many clinical research activities. Noonan syndrome (NS), one form of RASopathy, is an autosomal dominant disorder that prevents normal development in various parts of the body. It is typically characterized by unusual facial characteristics, short stature, congenital heart defects, blood clotting deficiency, developmental delays, and malformations of the bones of the rib cage. Individuals with NS also have an increased risk of cancer. The incidence of NS is approximately 1/1000 to 1/2500 newborns. At present, the most common genes associated with NS are *PTPN11*, *KRAS*, *SOS1*, and *RAF1*. We developed a panel with 90 amplicons to cover all of the coding exons of all isoforms of *PTPN11*, *KRAS*, *SOS1*, and *RAF1* genes with a 25 bp of buffer at the exon-intron boundaries. Amplicon sizes are between 92- 237 bp. The prepared libraries were validated with GIAB (Genome in a Bottle) on the Illumina® MiSeq® instrument. All targeted regions were successfully sequenced with 100% uniformity at $\geq 0.1\times$ mean coverage, and $\geq 99.3\%$ of the reads mapped to the targeted regions. The primer designs and library preparation method are platform-agnostic, with the exception of the instrument-specific adapters, and would easily allow compatibility with the Ion Torrent™ platform. The poster will also review the assay's ability to detect previously described mutations in *PTPN11*, *KRAS*, *SOS1*, and *RAF1* genes in other reference samples. The Noonan Syndrome Panel covers a 15.5 kb region and allows for the detection and differentiation of critical mutations associated with NS in a clinical research setting. This panel could serve as a base for larger panels that would also detect other, less frequent forms of RASopathy. Research use only. Not for diagnostic use.

2978F

L1CAM partial deletion in a Polish family with X-linked hydrocephalus: Clinical and molecular characterization. *M. Mlynek, A. Cie´sak,likowska, D. Wicher, M. Krajewska-Walasek, K. Chrzanowska.* Department of Medical Genetics, Children's Memorial Health Institute, Warsaw, Poland.

L1 syndrome describes a group of overlapping X-linked disorders caused by mutations in *L1CAM* located at Xq28. Clinical phenotypes within L1 syndrome include X-linked hydrocephalus with stenosis of the aqueduct of Sylvius (HSAS); mental retardation, adducted thumbs, shuffling gait, and aphasia (MASA) syndrome; spastic paraplegia type 1; and agenesis of the corpus callosum. Hundreds of mutations in *L1CAM* have been reported, however, only a few large gene deletions have been documented. Here we report on the clinical and molecular characterization of a four-generation Polish family in which six of the nine male members suffer from congenital hydrocephalus, short stature, microcephaly, strabismus, scoliosis, adducted thumbs, clubfoot, intellectual disability, aphasia, shuffling gate, and spasticity. Based on clinical and medical history (consistent with X-linked inheritance) a diagnosis of MASA syndrome was made. Additional features observed in the affected members include generalized seizures, cryptorchidism, intestinal obstruction or Hirschprung's disease. Initial failure of *L1CAM* gene sequencing suggested the possibility of a large rearrangement. Further whole-genome oligonucleotide microarray analysis revealed a 3.777-base pair hemizygous deletion within Xq28. The deleted region encompasses exons 3–5 of *L1CAM* with genomic coordinates chrX:153136766-153140542; GRCh37. This is the first report of array CGH used to diagnose partial *L1CAM* deletion. Given our results and other reported individuals with entire or partial *L1CAM* deletion, it appears that large gene rearrangements cause a more severe phenotypic expression of L1 syndrome. In this family, array CGH analysis was performed to detect asymptomatic female carriers. The identification of *L1CAM* deletion in family members with MASA syndrome will give them an opportunity for DNA-based counseling and precise prenatal diagnosis for all future pregnancies in this family, where appropriate. This study was supported by CMHI Grant No M23/17.

2979W

Molecular characterization of complex ring chromosome 11 in an infant with multiple congenital anomalies. R.S. Ebrahim^{1,2}, B. Kettinger², M. Hankered², M. Kristofice², A.I. Ahmed^{1,2}. 1) Wayne State University school of Medicine, Detroit, MI, USA; 2) Detroit Medical Center University Laboratories, Cytogenetic, Detroit, MI USA.

Ring chromosome 11 is an uncommon chromosomal abnormality reported in at least 20 cases in the literature. We report on the cytogenetic and molecular characterization of a rare, atypical, and complex ring chromosome 11 with deletions and duplications affecting both arms in an infant with multiple congenital anomalies. The infant was born at 40 weeks gestation to a 20 year old G2P1 mother. The pregnancy was complicated by gestational diabetes that was identified within the last three weeks of pregnancy. The infant presented with hypotonia, hypoglycemia, respiratory distress, dysmorphic features, cardiac defects, and was found to have a ring chromosome 11 on routine chromosome analysis shortly after birth. Follow-up array-CGH testing revealed two interstitial duplications and two terminal deletions, one at either end of the ring chromosome 11: 46,XX,r(11)(p15.5q25).arr[GRCh37] 11p15.5(205827_766624)x1,11p15.5p15.4(873973_8011711)x3,11q24.3q25(130341685_131845433)x3,11q25(131855948_134928849)x1. The deletion at 11q25 overlaps the Jacobsen syndrome region and includes the *JAM3* gene implicated in congenital heart defects as well as, the *OPCML* and *NTM* genes associated with intellectual disability. The duplication at the 11p15.5-p15.4 overlaps with the Beckwith-Wiedemann syndrome (BWS)/Silver-Russell syndrome (SRS) critical region. This is the first case report of a ring chromosome 11 with deletions and duplications overlapping the Jacobsen syndrome and the BWS/SRS critical region. This case demonstrates the effective use of array CGH for the diagnosis of ring chromosome 11 and the characterization of a complex phenotype for clinical management.

2980T

Beckwith-Wiedemann syndrome and Jacobsen syndrome caused by 11p duplication and 11q deletion inherited from paternal pericentric inversion. S. Lim¹, E. Seo^{1,2}, H. Yoo^{1,3}, B. Lee^{1,3}, C. Seol^{1,2}, J. Lee¹, J. Lee⁴. 1) Medical Genetics & Genomics Center, Asan Medical Center; 2) Department of Laboratory Medicine, Asan Medical Center; 3) Department of Pediatrics, Asan Medical Center; 4) Asan Institute for Life Sciences, Asan Medical Center.

We report on a female infant with a paternally derived recombinant chromosome 11 of 11p duplication and 11q deletion. This infant showed lung hypoplasia, generalized skin edema, polyhydramnios, club foot, DORV and VSD in prenatal ultrasound. At birth, the infant had respiratory problem. The conventional karyotype was 46,XX,rec(11)dup(11p)inv(11)(p15.1q24.2). Methylation-specific MLPA (MS-MLPA) result of genes located on 11p showed duplication of *H19* and *KCNQ1* alleles, and normal level of *KCNQ1* alleles and an increased level of *H19* alleles in methylation status, suggesting duplication of paternal alleles. The karyotype of the father showed the pericentric inversion of chromosome 11. In addition, array-CGH was performed in order to define the copy number changes more precisely and to rule out other chromosomal rearrangements. We identified an 11p15.5p14.3 duplication of 23.6 Mb and an 11q23.3q25 deletion of 13.8Mb. This finding is associated with Beckwith-Wiedemann syndrome (BWS) and Jacobsen syndrome (JBS). This patient presented with craniofacial features consistent with JBS: microcephaly, down slant palpebral fissure, webbed neck, low set ears. The others manifestations were thrombocytopenia, neutropenia, infection and double-outlet right ventricle (DORV) with VSD, interruption of central pulmonary artery. Meanwhile the clinical findings of BWS were polyhydramnios, macrosomia, splenomegaly, and renal abnormality with several cysts. Chromosome analysis, molecular test and array-CGH analysis for the diagnosis of syndromes will provide us with integrated insight into these complex rearrangements and the underlying mechanisms.

2981F

Brazilian experience of 22q11.2 deletion syndrome: Clinical variability and immunodeficiency. M.M. Montenegro¹, E.A. Zanardo¹, J. Damasceno¹, D.C.Q. Soares², G.M. Novo-Filho¹, A. Nascimento¹, L. Liro¹, Y. Gasparini¹, A.T. Dias¹, S.N. Chehimi¹, F.A.R. Madia¹, O.S. Akli¹, C.A. Kim², M.M.S. Sampaio³, L.D. Kulikowski¹. 1) Laboratório de Citogenômica, Departamento de Patologia, Faculdade de Medicina FM, São Paulo, Brazil; 2) Unidade de Genética, Departamento de Pediatria, Instituto da Criança, Hospital das Clínicas HCFMUSP, Faculdade de Medicina, Universidade de São Paulo, São Paulo, SP, BR; 3) Departamento de Pediatria, Instituto da Criança, Hospital das Clínicas HCFMUSP, Faculdade de Medicina, Universidade de São Paulo, São Paulo, SP, BR.

The 22q11.2 deletion syndrome (22q11.2DS) is the most common microdeletion syndrome estimated in approximately 1 in 4,000 births. The spectrum of 22q11.2DS DiGeorge syndrome, also include Velocardiofacial, Cayler's Cardiofacial Syndrome and conotruncal face anomalies. Classical clinical manifestations of a deletion in 22q11.2 include congenital cardiac anomalies, typical facies with prominent ears and enlarged nasal bridge and immune deficiency such as T cell lymphopenia. The variability of clinical manifestations encompassing altered genomic between 0.7 to 3 million Mb. We studied 63 Brazilian patients followed in the Genetics Unit of the Instituto da Criança HCFMUSP (Brazil). Cytogenomic study of the microdeletion was performed by MLPA using P064, P356 and P250 kits and array technique with CytoSNP-850K BeadChip. Clinical manifestations were collected in collaboration of Medical Geneticists and Immunogeneticist. To assess immunocompetence, tests with specific panel of antibodies were performed. Breakpoint size of del 22q11.2 was mapping in all patients. Typical deletions were found in 54 patients (3Mb) and atypical deletions in nine patients (2.3 Mb). The most prominent clinical characteristic observed both in patients with typical and in atypical deletions was neuropsychomotor development delay (NPDD - 22/63), High palate (14/63), tetralogy of Fallot (11/63), Repeat Infections (11/63), short stature (11/63) and others. The immunological analysis demonstrated alteration in the NK cells with elevated relative values of NK cells and NK CD161+ cells and low levels of NK CD4+ and NK CD8+ cells when compared to healthy controls. Clinical features of varying degree of expression occur in all patients, however all present NPDD and tetralogy of Fallot. The size of the deletion suggest that the variability is due to the loss of important genes that participate in the biological process of tissue and organ formation. In addition, such structural alteration may cause changes in chromatin conformation or impairing the function of other genes, affecting the expression of other genes involved in the pathogenesis of the 22qdel11.2. Lastly, the immunodeficiency observed in our patients with 22q11.2DS could be associated with B-cell defects and also NK-cell abnormalities. A multidisciplinary approach including cytogenomics and accurate immunological studies can improve the management of these patients and a better understanding of complexities in 22q11.2DS.

2982W

Integrated genomic and epigenomic analysis of CNVs involving the *NFIX* locus. M.L. Duque Lasio¹, I. Amarillo². 1) Department of Pediatrics, Washington University School of Medicine, St. Louis, MO; 2) Cytogenetics and Molecular Pathology Laboratory, Division of Laboratory Medicine, Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, MO.

Introduction: *NFIX* (nuclear factor I/X, MIM #164005) encodes for a highly conserved protein that has an N-terminal DNA binding domain and a C-terminal activation/repression domain. Sequence variations and microdeletions in *NFIX* are associated with allelic disorders Sotos syndrome 2 (Sotos 2) (aka Malan syndrome) and Marshall-Smith syndrome. Recently, a 19p13 microduplication including *NFIX* was reported in patients with phenotypic features described to be reciprocal to Sotos 2. This study aims to identify small deletions and duplications (<200 kb) involving *NFIX*, further characterize the genomic and epigenomic features of such CNVs, and correlate with the phenotypic findings. **Methods:** We performed a retrospective study using CMA data from patients referred to our clinical laboratory from 2008 to 2017 (n=4,700). Using the Affymetrix Chromosome Analysis Suite analysis software, a 1 kb/1 probe CMA filter was used to capture CNVs involving the *NFIX* locus (19p13.12 in hg19). Small CNVs were verified by qPCR. We aligned CNVs from published reports and DECIPHER database with those from our patient set. We performed genotype-epigenotype-phenotype analysis on small CNVs with breakpoints within *NFIX* and on CNVs encompassing the entire gene using the UCSC Genome and the WashU EpiGenome browsers. **Results:** We identified four duplications, two deletions and one concurrent deletion/duplication within our data set. Patients with deletions had phenotypic findings consistent with Sotos 2 syndrome. All duplications occurred in exon 1, and associated phenotypic findings were not consistent with the recently described 19p13 duplication syndrome. Alu-rich regions in exons 6 and 7 of *NFIX* have been implicated as hotspots for CNVs. Our study revealed Alu-repeats at recurrent breakpoints within introns 1, 2 and 9, suggestive of potential novel hotspots for CNVs. These small intervals also revealed significant genomic and epigenomic marks distributed across *NFIX* that are involved in spatiotemporal regulation, e.g. multiple CpG islands, repeat elements such as Alu SINEs, DNA and histone methylation, and enhancer/repressor regions. **Conclusion:** Our retrospective high-resolution CMA integrated with salient genomics and epigenomics data may improve the diagnostic yield and enhance the clinical management of patients with *NFIX*-related disorders.

2983T

Clinical interpretation of regions of homozygosity identified by SNP arrays in consanguineous families. J. Levy^{1,2}, C. Dupont¹, M. Rachid¹, B. Benzacken^{1,2}, Y. Capri¹, S. Drunat¹, F. Guimiot^{1,2}, P. Gressens², S. Grotto¹, D. Haye¹, S. Passemard^{1,2}, L. Perrin¹, A. Maruan^{3,4}, F. Renaldo⁵, M. Schiff⁶, M-A. Thaly¹, A. Verloes^{1,2}, E. Pipiras^{1,2}, A-C. Tabet^{1,4}. 1) Department of Genetics, AP-HP, Robert-Debre University Hospital, Paris, France; 2) INSERM UMR1141, Robert-Debre University Hospital, Paris Diderot University, AP-HP, Paris, France; 3) Department of Child and Adolescent Psychiatry, Robert Debré University Hospital, AP-HP, Paris, France; 4) Human Genetics and Cognitive Functions, Pasteur Institute, Paris, France; 5) Department of Child Neurology, Robert Debré University Hospital, AP-HP, Paris, France; 6) Reference Center for Inborn Errors of Metabolism, Robert Debré University Hospital, AP-HP, Paris, France.

Inbreeding is a common cultural practice in many parts of the world. Children born in these families have regions of homozygosity (ROHs) and an increased risk of autosomal recessive disease. SNP array allows whole genome analysis to identify both copy number variants (CNVs) and ROHs. These ROHs indicate ancestral homozygosity, uniparental disomy, or parental consanguinity depending on the size, the location and the chromosome(s) involved. In the context of several genetic disorder in children of consanguineous parents, the identification of the causal gene can be difficult given the great genetic and clinical heterogeneity. In these families, the mutated gene in the homozygous state is frequently found in a ROH. When SNP array revealed no genomic imbalances, we performed a clinical interpretation of ROHs in 213 patients with a broad range of clinical indications including intellectual disability, developmental delay, multiple congenital anomalies, or autism spectrum disorder. We analyzed these regions in order to identify genes that could be implicated in recessive pathology concordant with patients' phenotype. The analysis of ROH using the software "SNP array evaluation tool v3.0" found the mutated gene in 26/213 patients (12,2%). We further analyzed 30 multiplex families in which both affected and unaffected siblings were recruited. We studied common ROHs shared by affected patients, excluding ROHs present in unaffected siblings. Based on key words describing our patients' phenotype, we searched for OMIM genes implicated in recessive diseases concordant with the phenotype and selected candidate genes. ROHs analysis successfully lead to a diagnosis in 8/30 families (26,7%) in 15 patients. Finally, we performed whole exome sequencing in 7 selected multiplex families in which the ROH analysis was non-contributory and identified the causal gene in two out of three families analyzed. We identified two novel pathogenic variants in *ERBB3* reported to cause lethal congenital contractural syndrome 2 (MIM # 607598). The analysis of the remaining 4 families is currently ongoing. The diagnosis of genetic and chromosomal disorder in a context of consanguinity, ROH analysis must be considered in order to sequence the gene of interest, or to target WES analysis. Our strategy improves the efficiency of genetic diagnosis in a more targeted and cost-effective way.

2984F

Using array-based comparative genomic hybridization to diagnose Pallister-Killian syndrome: Are we there yet? I. Marcussi, C.M. Lourenco, J. Peral, K. Romanin, M.C. Silva. CENTRO UNIVERSITARIO ESTACIO DE RIBEIRAO PRETO, Ribeirao Preto, Sao Paulo, Brazil.

Introduction Pallister-Killian syndrome (PKS) is a rare genetic disorder characterized by craniofacial dysmorphisms including a prominent forehead with sparse frontotemporal hair and a wide range of developmental delays, intellectual disability, hypotonia, seizures, skin pigmentation, and other systemic anomalies **Material/Results** Here we describe two unrelated patients with PKS whose array CGH results pointed out to different results, maybe reflecting the degree of mosaicism that could be foreseen by skin pigmentation features **Discussion** Analysis of skin fibroblasts or buccal smears has been proposed as a diagnostic gold standard, but this notion is still controversial. To many experts, a skin biopsy sample should continue to be the diagnostic gold standard for PKS because of the variable and insufficient quality of buccal smears. Nonetheless, more and more it has been showed that in patients with suspected PKS, array CGH or FISH analysis of buccal smear cells seems to be the first diagnostic choice because these two methods are reliable, rapid, and noninvasive. So, skin biopsy should only be performed if buccal-smear analysis cannot detect i(12p) **Conclusion** PKS is a rare multisystem disorder that is difficult to diagnose prenatally and in early infancy. Suspicion of PKS can be made by physical examination if characteristic dysmorphic features and global developmental delay are present. Array CGH can be used to diagnose PKS in early infancy and during the prenatal period. Clinicians can also employ FISH analysis to diagnose this disorder in patients with visible PKS features. Subsequently and only if buccal smear analysis fails to produce results, then a skin biopsy is warranted.

2985W

Genotyping by NGS of 1200 known pathogenic mutations for extensive carrier screening of 228 monogenic recessive diseases helps avoid reporting VUS. *D. Bercovich¹, S. Horn-Saban², R. Gershoni².* 1) Human Molecular Genetics, Tel Hai College, Tel Hai, Israel; 2) Galile Genetic Analysis lab, Kazarin, Israel.

Full sequencing of genes related to monogenic recessive diseases revealed lots of VUS (variants of unknown significance) per individual which there are diabetes if they should be reported or not as there clinical significance is unclear. To avoid this problem, we generated a genotyping method using NGS of about 1200 known pathogenic mutations which are known to be related to the 228 most common monogenic recessive diseases in the Israeli and European populations. More than 80% of mutations are known rare mutations which are not screened by genetics labs on regular basis. As all the clinical information on these 1200 mutations is fully known, the TAT of summary letters is reduce to half the time, in comparison to full genes sequencing, as there is no need to address VUS and analyze their clinical significance. In screening several hundred healthy individuals and couples it was found that the carrier rates amounted to 1 to 2 mutations per DNA sample with an average of 1.3 mutations. To increase the genetic test sensitivity we added the SMA, FRAX (only women) and 6 most common microdeletions in the CFTR, GJB6, MCOLN, NEB, GALT & PAH genes as well. It was found that about 0.001% of couples were both carriers of different mutations in the same gene which can lead to approximately 7.5×10^{-5} offsprings with genetic diseases out of all births per year in Israel. These results indicate the importance of extensive genetic panel for common monogenic diseases for the detection of known recessive mutations in the general population and prevention of sick babies being born. We advise that such extensive genetic mutations panel will be applied as routine pre-conceptional screening, and should be followed by genetic counseling.

2986T

Distribution of MYCN pathogenic variants in Feingold syndrome 1. *S.M. Kirwin¹, K.M.B. Vinette¹, S.E. Mora^{1,2}, K.M. Robbins¹, K.W. Gripp^{1,2}, V.L. Funanage².* 1) Molecular Diagnostics Laboratory, Nemours/Alfred I. duPont Hospital for Children, Wilmington, DE; 2) Division of Medical Genetics, Nemours/Alfred I. duPont Hospital for Children, Wilmington, DE.

MYCN is a proto-oncogene and member of the MYC family. Located on chromosome 2p23-2p24, it encodes a transcription factor with a basic helix-loop-helix region. Somatic variants within *MYCN* are primarily known to cause neuroblastoma and retinoblastoma. Feingold syndrome 1 (FGLDS1; OMIM 164280) is an autosomal dominant disorder associated with heterozygous germline variants or deletions in *MYCN* leading to haploinsufficiency of the protein. These variants in *MYCN* have been identified in patients with microcephaly; short stature; short palpebral fissures; digital anomalies (thumb hypoplasia; brachymesophalangy 2 and 5; toe syndactyly 2-3 or 4-5); variable intellectual disability (spectrum from normal intelligence through mild-moderate disability); and 20-40% of patients may also present with esophageal atresia or tracheal esophageal fistula, cardiac or renal anomalies. We present a review of novel and previously reported variants found in an extensive cohort of FGLDS1 patients in our diagnostic laboratory including eight probands and two symptomatic parents. The preponderance of mutations are located within exon 3, as previously reported, and all variants found in exon 2 create an early termination of the protein due to a stop codon or frameshift variant. FGLDS1 is reported to have variable expressivity and further examination of family members found two parents with *MYCN* mutations, leaving 80% of this cohort predicted to be *de novo* in nature. Sequence analysis of *MYCN* is indicated in individuals with a phenotype suggestive of Feingold syndrome.

2987F

The utility of a next generation sequencing panel in severe pediatric onset inherited diseases. *E. Isik¹, H. Onay², T. Atik¹, E. Canda³, O. Cogulu^{1,2}, F. Ozkinay^{1,2}.* 1) Subdivision of Pediatric Genetics, Department of Pediatrics, Faculty of Medicine, Ege University, Izmir, Turkey; 2) Department of Medical Genetics, Faculty of Medicine, Ege University, Izmir, Turkey; 3) Subdivision of Pediatric Metabolism, Department of Pediatrics, Faculty of Medicine, Ege University, Izmir, Turkey.

Aim: Mendelian diseases are very rare, with many of them are genetically heterogeneous. Despite this however, they collectively account for about 20% of infant mortalities and about 18% of pediatric hospitalizations. Targeted next generation sequencing provides simultaneous sequencing of many genes, and is therefore able to play; it has an important role in the genetic diagnosis of Mendelian disorders. The aim of this study is to evaluate the diagnostic rate of a targeted next generation panel in severe pediatric onset and genetically heterogeneous autosomal recessive diseases. **Material and Methods:** In this study, molecularly analyzed cases between March 2015 and November 2017 were evaluated retrospectively. All cases had originally been diagnosed using a next generation sequencing panel (TruSight Inherited Disease® Panel) on Illumina MiSeq platform. **Results:** A hundred and eighty five cases (175 affected, 10 carrier) were analyzed using TruSight Inherited Disease® Panel. Referral reasons for molecular analysis of the cases were: Inborn errors of metabolism (47.6%), genetic syndrome (23.9%), neurological disorder (7%), mental retardation / multiple congenital anomalies (5.9%) and others (15.6%). The molecular diagnostic rate of the next generation sequencing panel in all cases was 40%. The highest diagnostic rate was achieved in neurological disorders (69.2%). In the cases not having a clinical diagnosis prior to molecular analysis no pathogenic variant was found. Fifty four mutations in 39 different genes, defined in 37 cases, were novel. **Conclusion:** This is the first study to evaluate the diagnostic effectiveness of on inherited disease panel (TruSight Inherited Disease® Panel). This targeted next generation panel achieved a high molecular diagnostic rate which provides the opportunity for advanced genetic counseling in severe pediatric onset and genetically heterogeneous autosomal recessive diseases. Comprehensive clinical examination is recommended to achieve specific clinical diagnosis for more efficient molecular analysis.

2988W

Copy number variation analysis for newborn screening using genome-wide SNP array and its application study on Korean population. *J.S. Bae¹, J. Joo², B.R. Park¹, H.J. Kwon¹, J. Kim¹, H. Kim¹, H. Im¹, R.P. Mohanty¹, A. Goyal¹, S. Lee¹, M.J. Kim², M.S. Lee^{1,2,3}.* 1) Eone-Diagnomics Genome Center, Inc., Incheon, Republic of Korea; 2) Diagnomics, Inc., 5775 Kearny Villa Rd., San Diego, CA 92123, USA; 3) Incheon National University, 119 Academy-ro Yeonsu-gu, Incheon, 22012, Republic of Korea.

Copy number variation (CNV) is the major cause of structural variation in the human genome, involving both duplications and deletions of sequences leading to several disorders. Analysis of neonatal CNV is crucial in managing CNV-related chromosomal disorders. In this study, we have developed and validated a newborn screening test for genomic abnormalities in 138 CNV-related chromosomal disorders. We used customized Illumina Infinium Global Screening Array (GSA) BeadChip (version 1), which includes 680,960 single nucleotide polymorphism (SNP) makers, and CNVs surveyed in 4507 Korean subjects. For CNV data analysis, we used our own analytical pipeline. To confirm the analytical performance, we analyzed genomic DNA from 22 human cell lines with CNV and successfully detected all corresponding CNV regions using our analysis pipeline. In total, 4507 newborn screenings were conducted with this developed test. The CNVs related to genetic disorders were identified in 69 cases (1.53%) with 27 disorders. The most frequently detected region for microdeletions were 2q13 Joubert syndrome-4, and the 15q11.2 deletion syndrome, corresponding to 11 out of 69 cases (0.244%) in each region. The next most frequently detected CNV region was 22q11.2 duplication syndrome in 7 cases (0.155%). CNVs in the CNV regions of 1q21.1 microdeletion syndrome, 16p11.2 duplication syndrome, 17q12 duplication syndrome, 17p13.3 microduplication syndrome, downsyndrome, triple X syndrome, and Klinefelter syndrome corresponds to 3 cases (0.067%) in each. The CNV regions related with 15 other syndromes were found in one case each (0.022%). In conclusion, genome-wide SNP array is useful and powerful method for detecting the neonatal CNV, by accurate screening and subsequently identifying the risk of the disease in advance, thereby discovering new etiologies.

2989T

Copy Number Variations (CNVs) data from Brazilian patients with neuropsychomotor developmental delay and congenital malformations obtained through SNP array methodology. L.P.G. Bertollo, E.A. Zanardo, C. Ae Kim, A.T. Dias, L.D. Kulikowski. Faculdade de Medicina da Universidade de São Paulo, Av. Dr. Arnaldo, 455, São Paulo, Brazil.

Introduction and objectives: neuropsychomotor developmental delay (NDD) associated with congenital malformations (CM) is a frequent condition in clinical practice that often remains undiagnosed. We investigated undiagnosed NDD/CM patients using the SNP array methodology, aiming to search for Copy Number Variations (CNVs) relevant to clinical phenotype, as well as describing the CNVs characteristics in this cohort of Brazilian patients. **Methods:** we studied SNP array and clinical data from 60 patients referred from a high complexity genetic outpatient clinic. The results were analyzed according to the American College of Medical Genetics guidelines using independent tests and were compared with the following databanks of CNVs: the Database of Genomic Variants, the Database of Chromosomal Imbalance and Phenotype in Humans Using Ensemble Resources and the UCSC Genome Bioinformatics database. The genomic positions are reported according to their mapping on the GRCh37/hg19 genome build. The CNVs were then classified as benign, pathogenic or VUS (variants of uncertain clinical significance) and a qualitative frequency analysis was performed. **Results:** in 27 patients (45%), at least one pathogenic CNV was found. In 19 patients (31.7%), only benign CNVs were found. In 4 patients, no pathogenic variants were found, but there was at least one VUS. In 10 patients, no CNVs were found. The chromosomes which presented a higher frequency of pathogenic CNVs were chr.4 (7 pathogenic CNVs, especially involving the 4q35 and 4p16 regions), chr.5 (6 pathogenic CNVs, specially involving the 5q35 and 5p15.33 regions), chr.9 (4 pathogenic CNVs, 3 in the 9p24 region) and chr.X (4 pathogenic CNVs, 3 in the Xq28 region). 44% of the pathogenic CNVs represented deletions. **Conclusions:** SNP array is an extremely useful tool for investigating patients with NDD/CM, providing a higher diagnostic conclusion rate for our patients compared to CGH Array (around 10-14% according to the literature). The accurate assessment of the CNVs in these patients consist of an important database with representative information from Brazilian people. **Funding Agency:** FAPESP - process 2017 / 21659-1.

2990F

Precise breakpoint detection of balanced and unbalanced structural variation in whole genome sequencing data using haplotype blocks created by linked-reads. J. Knijnenburg¹, M. Kriek¹, H.A. van Duyvenvoorde¹, N. Pappas², Y. Ariyurek³, H.P.J. Buermans³, K.M.L. Hertoghs¹, M.E.Y. Laurensse-Bik¹, K. Kuipers-Heijboer¹, F. Baas¹. 1) Clinical Genetics, LUMC, Leiden, Netherlands; 2) Biomedical Data Sciences, LUMC, Leiden, the Netherlands; 3) Human Genetics, LUMC, Leiden, the Netherlands; 4) GenomeScan, Leiden, the Netherlands.

Introduction: Routine diagnostics of genome wide balanced and unbalanced structural variation is still dependent on classical techniques such as karyotyping and (SNP-)array. Both techniques result in a rough breakpoint detection where especially karyotyping can result in significant misinterpretation of chromosomal breakpoints. New sequencing technologies such as long range haplotyping by mapping linked-reads to generate haplotype blocks (Chromium Genome Solution, 10x Genomics, San Francisco USA) are expected to overcome this problem, although repeat sequence elements at breakpoints in the genome may occasionally interfere with a precise detection. **Materials and Methods:** To test whether this NGS based approach with linked-reads is suitable for routine diagnostics, a small series of patients, including one with a de novo balanced translocation (8;17) and one with a duplication detected with SNP-array, were analysed. **Results:** The expected structural variants based on previous knowledge were identified and in both patients precise breakpoints of these variants could be detected. In the translocation patient, the breakpoint differed at least 15 Mb from the estimated breakpoint and proved SOX9 to be causative for the phenotype of the patient. In the second patient the orientation of the duplication fragment could be defined to be in tandem, making a gene disruption to be causal for the phenotype unlikely. **Conclusions:** Structural variant analysis using WGS with linked-reads promises to be a feasible technique for detection of balanced and unbalanced structural variants in the genome and may be a next step in replacing classical cytogenetic techniques for genome wide screening.

2991W

FISH is still the most reliable diagnostic tool for Pallister-Killian syndrome: A case based review. S. Tug Bozdogan^{1,2}, A. Bisgin^{1,2}, G. Ceylaner³, I. Boga³, S. Ozer³. 1) Cukurova University, Faculty of Medicine, Medical Genetics Dept of Balcali Hospital and Clinics, Adana, Turkey; 2) Cukurova University AGENTEM (Adana Genetic Disease Diagnosis and Treatment Center), Adana, Turkey; 3) Intergen Genetic Diagnosis Center, Ankara, Turkey.

Objective: Pallister-Killian syndrome is a rare sporadic genetic syndrome with distinctive dysmorphic features caused by chromosomal mosaicism of chromosome 12p tetrasomy. Estimated prevalence of disease is 5.1 per million in live birth with the 50% mortality rate during the first year of their lives. Clinical heterogeneity with short life time and tissue limited mosaicism make the diagnosis difficult. However, cytogenetic testing is the only methodology for genetic diagnosis. Though, due to the mosaicism cytogenetic analysis from leukocyte may not capture the abnormality. Therefore, proper clinical genetic evaluation is substantial for diagnosis with proper diagnostics. Within this case based presentation the diagnostic algorithm for a Pallister-Killian syndrome was reviewed from cytogenetics to far distant but misused of a whole exome sequencing. **Methods:** We report a 5-years old female patient who referred to our clinical genetics policlinics with dysmorphological features of overgrowth, macrocephaly and coarse face with intellectual impairment and developmental delay. Cytogenetics, molecular cytogenetics and whole exome sequencing was performed to the patient. **Results:** Firstly cytogenetic investigations from peripheral blood was performed with normal karyotype (46,XX). However as a next step diagnostics, whole exome sequencing was applied in another out patient clinics with no genetics specialist. During the follow up we performed the cytogenetics from skin biopsy. Finally, patient had a diagnosis with Pallister-Killian syndrome due to mosaic tetrasomy in chromosome 12p according to skin biopsy cytogenetic investigation and the result was confirmed with metaphase FISH on cultured skin fibroblasts (46,XX [18] / 47,XX, +i(12)(p10)[12], nuch ish (AMLx2) (ETOx4) [40/100]). **Conclusion:** Whole exome sequencing might be easily misused by non-experienced clinicians as a final solution for diagnosis. However, such cases as in Pallister-Killian syndrome, cytogenetic testing or FISH analysis are still the most reliable ones within the hands of medical/clinical geneticists.

2992T

Identification of *HPSE2* variant in urofacial syndrome by whole exome sequencing. I. Cherkaoui Jaouad^{1,2}, A. Zrhidi², J. Lyahya², MC. Ferreira³, S. Kemeny³, P. Pugniere³, E. Page³, P. Lochu³, A. Sefiani^{1,2}, G. Egéa³. 1) Department of Medical Genetics, National Institute of Health, Rabat, Rabat, Morocco; 2) Center of Human Genetic, Medical and Pharmacy faculty, Mohammed V University, Rabat, Morocco; 3) Department of Medical Genetics, Imagénome, Gen-Bio, Clermont-Ferrand, France.

Urofacial syndrome (UFS) or Ochoa syndrome (OMIM : 236730) is characterized by infantile onset of urinary bladder voiding dysfunction, abnormal facial movement with expression. A peculiar facial dysmorphism is related to an unusual inversion of facial expression that occurs when the child smiles or cries. Bladder voiding dysfunction increases the risk for urinary incontinence, megacystis, vesicoureteric reflux, hydronephrosis, urosepsis, and progressive renal impairment. We applied whole exome sequencing technology to a 7 year old moroccan boy, born from consanguineous parents, presenting severe voiding dysfunction and a characteristic facial expression of UFS. By using a trio whole exome sequencing, we identified a homozygous nonsense mutation, c.1414C>T (p. Arg472*) in exon 10 of *HPSE2* (which encode the protein Heparinase 2). The same variant has previously described like pathogenic by Saly et al. in one Turkish family. This mutation has been observed to lead to nonsense mediated decay of the transcript and absent protein. The gene is normally expressed in both the central nervous system and the bladder. The parents were both confirmed as heterozygotes for the mutation. We describe the second family in our knowledge with the homozygous nonsense mutation c.1414C>T in *HPSE2* causing Urofacial syndrome.

2993F

Novel pathogenic mutation of *ST3GAL5* gene identified in Moroccan family with Salt and Pepper syndrome. G. Egéa¹, A. Zrhidri², I. Cherkaoui Jaouad^{2,3}, J. Lyahyai², MC. Ferreira¹, S. Kemeny¹, P. Pugniere¹, E. Page¹, P. Lochu¹, A. Seffiani^{2,3}. 1) Department of Medical Genetics, IMAGENOME, GEN-BIO Laboratory, Clermont-Ferrand, France; 2) Center of Human genomic, Medical and pharmacy faculty, Mohammed V University, Rabat, Morocco; 3) Department of Medical Genetics, National Institute of Health, Rabat, Morocco.

We identified an autosomal recessive Salt and Pepper syndrome characterized by severe intellectual disability, epilepsy with absence of speech, decreased eye contact, irritability and choreoathetosis in three siblings observed on consanguineous family, two brothers and one sister. In our study, whole exome sequencing was performed on one brother and sister affected and the parents unaffected. The analysis revealed a novel, homozygous missense mutation (c. 674C>T) in exon 5 of the *ST3GAL5* gene, resulting from p.Ala225Val (NM_003896.3). Whole exome sequencing data was further validated using Sanger sequencing by designing the primers of the identified mutation. Moreover this mutation was found homozygous for the other brother affected. The *ST3GAL5* gene (OMIM : 609056) encodes for a sialyltransferase (also known as GM3 synthase) that synthesizes ganglioside GM3, a glycosphingolipid enriched in neural tissue. The gene might then have a role for normal brain function. Our report presents the results of the quartet exome data of both unaffected parents and their affected son and daughter with the phenotype describing the salt and Pepper syndrome. A homozygous missense variant in the *ST3GAL5* gene was identified as a candidate causal variant for the clinical phenotype in this family. The complementary investigation in analytical chemistry is ongoing in order to measure the plasma GM3, and its downstream metabolites, GD3 gangliosides. This case expanded the clinical and genetic spectrum of ultra-rare disease, Salt and Pepper Syndrome with *ST3GAL5* mutation.

2994W

A pitfall in the molecular diagnostics of Kabuki syndrome: Functional redundancy of exon 15 in *KDM6A*. S.K. Fiordaliso¹, A. Otsubo¹, A.L. Ritter¹, L. Lulis², S. Baker², K. Helbig³, I. Helbig³, M. Dulik², K. Izumi^{1,2}. 1) Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Division of Genomic Diagnostics, Children's Hospital of Philadelphia, Philadelphia, PA; 3) Division of Pediatric Neurology, Children's Hospital of Philadelphia, Philadelphia, PA.

The *KDM6A* gene encodes a histone demethylase and pathogenic variants in this gene are found in around 5-8% of individuals with Kabuki syndrome. Although *KDM6A* resides on the X-chromosome, it escapes X-inactivation, allowing Kabuki syndrome to occur in both female and male patients. While the introduction of comprehensive NGS-based genetic diagnostic testing has revealed a wider phenotypic spectrum for Kabuki syndrome, its phenotypic spectrum has yet to be defined. Occasionally, a genetic testing finding not consistent with the clinical findings creates a diagnostic dilemma. Here, we report a 12-year-old girl with developmental delay, moderate intellectual disability, short stature, autistic features, expressive speech and language disorder, and focal seizures beginning at age 4 years. Clinical trio exome sequencing identified a heterozygous, likely pathogenic *KDM6A* nonsense variant (NM_021140.3, c.1519C>T; p.Gln507*) that was paternally inherited and predicted to create a premature stop codon in the 15th of 29 coding exons, occurring upstream (5') of previously reported loss-of-function, disease-causing variants. Although this molecular finding was highly suggestive of Kabuki syndrome, neither the proband nor her father had any physical features suggestive of Kabuki syndrome. A lymphoblastoid cell line was established from the hemizygous father, and although quantitative RT-PCR revealed a lack of exon 15 transcript, exon 14 and exon 16 transcripts were still present, suggesting that *KDM6A* transcripts lacking exon 15 can compensate for the function of exon 15-containing isoforms. Therefore, we concluded that exon 15 of *KDM6A* is functionally redundant, and our patient's nonsense variant does not cause Kabuki syndrome nor explain her clinical findings. Despite evidence implicating *KDM6A* as a dosage-sensitive gene in humans, our findings suggest that exon 15 of *KDM6A* is functionally redundant, and that a variant introducing a premature stop codon in exon 15 does not cause Kabuki syndrome. From a molecular diagnostics standpoint, it is important to recognize that variants which seemingly result in loss-of-function may only affect a subset of gene isoforms and not critically affect a gene's overall function. As molecular diagnostic tools become more widespread, our finding emphasizes the importance of functional analysis and clinical correlation.

2995T

Repeat expansion disorders in patients evaluated by clinical whole exome sequencing. S. Gu^{1,2}, J. Scull^{1,2}, B. Yuan^{1,2}, C. Shaw^{1,2}, J. Zhang^{1,2}, C. Eng^{1,2}, Y. Yang^{1,2}, P. Liu^{1,2}. 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, US; 2) Baylor Genetics, Houston, TX, US.

Clinical whole exome sequencing (WES) has been demonstrated to improve diagnosis and alter patient management by numerous large cohort studies. It has become a routine genetic testing approach, with a diagnostic yield of 25-45% based on different studies. One category of genetic diseases that may escape detection by WES is the repeat expansion disorder (RED). We retrospectively reviewed ~10,000 consecutive clinical WES patients analyzed between November 2011 and April 2018, and identified 446 patients who also underwent testing for potential REDs in our diagnostic laboratory. This included 9 patients tested for spinocerebellar ataxia type 1 (SCA1), 9 for SCA10, 31 for Friedreich ataxia (FRDA), 116 for myotonic dystrophy type 1 (DM1), 6 for dentatorubral-pallidoluysian atrophy (DRPLA), 257 for Fragile X syndrome (FMR1), 5 for Kennedy Disease (KD), and 13 for Huntington Disease (HD). Among these patients with a positive RED result, three had concurrent WES and RED testing due to suspected genetic disorders with a non-specific clinical presentation, including two patients with expansions on both alleles for FRDA and one patient with an expansion for DM1. Three patients had molecular diagnoses of REDs before being tested for WES, and WES was ordered due to additional phenotypes unexplained by their RED diagnosis. This included one patient with an expansion for SCA10, one patient with an expansion for FMR1, and one unusual patient with expansions on both alleles for DM1. One particular patient had DM1 testing after a non-diagnostic WES due to atypical DM1 presentation, yet she was found to have an allele with approximately 1350 repeats. Additionally, six patients had mutable/premutation/intermediate alleles for autosomal dominant/X-linked REDs, while two other patients were FRDA carriers, informing disease risk estimates for individual families. Our analysis demonstrated the importance of concurrent WES and REDs testing in a substantial proportion of patients. Recent advancement in computational algorithms demonstrated successful estimation of repeat numbers based on short-read whole genome sequencing (WGS) data (Dolzhenko E. *et al.*, Genome Res. 2017). Therefore, for patients with non-specific phenotypes that are not typical of well-characterized disorders, it is promising to identify the underlying various types genetic abnormalities, e.g. small nucleotide changes (SNVs), structural variations, copy-number variants and REDs, simultaneously by clinical WGS.

2996F

Rapid turnaround whole exome sequencing for critically ill neonates.

C.S. Gubbels^{1,2,3,4}, G.E. VanNoy⁵, D. Copenheaver⁶, S. Yang⁶, M. Wojcik^{1,2,3,6}, N.B. Gold^{1,2}, C. Genetti^{1,4}, J. Stoler^{1,2}, A.H. Beggs^{1,2,4}, O. Bodamer^{1,2,3}, J. Juusola⁵, P.B. Agrawal^{1,2,4,6}, T.W. Yu^{1,2,3,4}. 1) Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA; 2) Harvard Medical School, Boston, MA; 3) Broad Institute of MIT and Harvard, Cambridge, MA; 4) The Manton Center for Orphan Disease Research, Boston Children's Hospital, Boston MA; 5) GeneDx, Inc., Gaithersburg, MD; 6) Division of Newborn Medicine, Boston Children's Hospital, Boston, MA.

In this study, we investigate the impact of applying rapid-turnaround trio-based WES in a cohort of critically ill neonates. Key inclusion criteria include corrected age <6 months, and one or more of the following clinical features: hypotonia, seizures, multiple congenital anomalies and/or a complex metabolic phenotype. Clinically relevant genomic variants (diagnostic of the phenotype, or otherwise relevant to the current hospitalization) are returned within seven days; no incidental findings are reported. Additional mitochondrial DNA (mtDNA) analysis is performed in cases when clinical suspicion warrants. 21 of 25 approached families decided to enroll. Families that declined cited reasons related to timing and privacy concerns surrounding genomics. One family withdrew after enrollment. To date, twenty WES and four additional mtDNA analyses have been performed. Sixteen out of these twenty cases (80%) received a definitive molecular diagnosis. In one additional case a variant contributing to the phenotype was found, and in another patient a strong candidate gene was identified. Fifteen of the diagnoses were found on rapid WES; the sixteenth arose from mtDNA analysis. Our patients were diagnosed earlier than cases of these disorders reported in the literature, and they had phenotypes on the severe side of the spectrum, especially for the metabolic and neuromuscular disorders. Outcome analyses demonstrate that in all cases the rapid result inspired one or more changes in management and/or had consequences for parents or other family members. Knowledge about the diagnosis and its prognosis prompted a shift to comfort care in half of the patients. Also, additional specialists were involved in the patients' care anticipating future needs. Other effects included medication changes, such as a switch in anti-epileptic drug leading to reduced seizure frequency in one patient, and the consideration of new experimental therapies in two others. For families with an inherited variant, consequences were mostly related to reproductive options for future pregnancies. In one family, mutation carriers were referred to cardiology for evaluation. This algorithm-driven approach identified candidate patients with a high likelihood of a diagnosis for rapid WES. These results suggest that rapid turnaround WES can lead to faster and more accurate diagnoses in neonates in the ICU setting, reduce unnecessary testing and procedures, and impact the course of medical care.

2997W

Biallelic HDAC3 mutations in a patient with developmental delay, hypotonia and seizures. A. Otsubo¹, S. Yum², L. Medne¹, J. Balciuniene³, B. Krock³, K. Izumi^{1,3}. 1) Division of Human Genetics, Children's hospital of Philadelphia, Philadelphia, PA; 2) Division of Pediatric Neurology, Children's hospital of Philadelphia, Philadelphia, PA; 3) Division of Genomic Diagnostics, Children's hospital of Philadelphia, Philadelphia, PA.

Gene regulation is a biological process essential to all organisms. Many neurodevelopmental disorders are associated with abnormalities in gene regulation. Histone deacetylase 3 (HDAC3) is one of the enzymes that controls gene expression by deacetylating histone proteins, leading to chromatin in transcriptionally inactive state. Previously, one individual with a primary seizure disorder was reported to have a *de novo* HDAC3 missense mutation. However, the association between HDAC3 mutation and genetic diagnosis has not been established yet. Here we report a patient with a severe epileptic encephalopathy with biallelic HDAC3 mutations. The proband is a 10 year-old female with severe and global developmental delays, hypotonia, periventricular heterotopias and intractable seizures. She developed infantile spasms at ~8 months of age that progressed to intractable myoclonic epilepsy. Brain MRI demonstrated subependymal gray matter heterotopias. Her initial presentation included hand-wringing and hand-mouthing behaviors, but she developed dyskinesia with choreoathetoid movements by ~18 months of age. Her length and weight have been below the 5th centile, although her head circumference measurements are in the normal range. Clinically, she was suspected of having a Rett syndrome, however, *MECP2*, *CDKL5*, *ARX*, and *FOXP1* sequencing was unremarkable. Exome sequencing revealed biallelic novel missense mutations (c.64C>G; p.H22D and c.317G>T; p.R106L) in *HDAC3* in the proband. Neither of these variants has been reported in healthy controls (ExAC or gnomAD). *In silico* prediction programs indicate that both missense variants are likely pathogenic. Although the proband here demonstrates some clinical overlap with the previously reported individual, her clinical symptoms are more severe. In our case, a heterozygous HDAC3 variant is seen in each of her healthy parents; therefore, the pathogenic mechanism of mutations may be different between the two cases. Interestingly, this proband was suspected of having Rett syndrome clinically, which is mainly caused by *MECP2* mutations. HDAC3 interacts with *MECP2*, and its genomic localization is compromised in Rett syndrome patient-derived cell lines. Therefore, biallelic mutations of *HDAC3* may have caused similar effects to those of *MECP2* mutations seen in Rett syndrome. In summary, we propose that the biallelic HDAC3 mutations may cause a novel neurodevelopmental disorder characterized by developmental delay, hypotonia and seizures. .

2998T

The VetSeq study: A pilot study of expanded exome sequencing in the Veterans Health Administration. J.L. Vassy^{1,2,3}, L.E. Hull¹, M.H. Crawley¹, A. Muirhead Oza¹, S. Amr¹, P. Pomianowski¹. 1) VA Boston Healthcare System, Boston, MA; 2) Department of Medicine, Harvard Medical School, Boston, MA; 3) Division of General Internal Medicine and Primary Care, Brigham and Women's Hospital, Boston, MA; 4) Laboratory for Molecular Medicine, Partners HealthCare.

Background Whole exome sequencing (WES) is increasingly demonstrating utility in a range of clinical contexts. We conducted a pilot study of WES at VA Boston Healthcare System (VABHS), returning both indication-based and secondary results. **Methods** VABHS providers refer patients for indication-based WES via a research protocol. Patients complete baseline surveys and interviews, using validated instruments to measure genetic literacy, quality of life, and trust in the healthcare system. WES is performed at a CLIA-certified lab, and the WES report includes results related to the indication for testing, secondary pathogenic or likely pathogenic (P/LP) variants in >4600 genes associated with monogenic disease risk, carrier status variants, and pharmacogenomic results for 20 medications. The referring provider discusses the results with the patient and documents medical decision-making in the medical record. Patients complete follow-up interviews 3 months after receiving results. **Results** Two patients have enrolled to date. P1 is a 72-year-old man with a long history of sensory neuropathy and myalgias treated with chronic opioid therapy. In his interview, he hoped that WES would identify a diagnosis and possible treatment for his children and grandchildren, who suffer similar symptoms. WES identified a pathogenic p.Arg298Cys variant in *LMNA* associated with Charcot-Marie-Tooth hereditary neuropathy type 2. Although other *LMNA* variants are associated with dominantly inherited neuromuscular disease, the patient's variant was inconclusive for his symptoms because it is typically associated with autosomal recessive inheritance. He had no P/LP variants in any of the 59 actionable ACMG genes, but secondary heterozygous frameshift variants were identified in *KIAA0586* (Joubert syndrome 23) and *NPHP1* (nephronophthisis 1) and are awaiting assessment and Sanger confirmation. P2 is a 33-year-old woman with a history of severe thrombocytopenia in her and 1 of 2 sons. In her interview, she endorsed hope that WES would suggest a treatment to stabilize their platelet counts. WES identified a variant of uncertain significance in a conserved DNA-binding domain of *ETV6*, a gene that is often rearranged or fused with other genes in certain leukemias. Testing of family members to assess segregation is pending for both patients. Both endorsed high trust in VABHS. **Conclusions** It is feasible to use WES in clinical care at VABHS and study its impact on patient care and outcomes.

2999F

FISH following aCGH analysis identifies unexpected results: Further support for FISH characterization of copy number changes. S.G. Sulpizio¹, S.A. Morton¹, D.A. VanGundy¹, V.M. Cawich¹, K.A. Leiser¹, A.J. Flage¹, C. Burson², K. Fry³, L.A. Bailey⁴, K.K. Bachmann³, L. Bailey⁵, L. Escobar⁶, K. Tsuchiya⁶, G. Bellus³, A. Seeley², B. Salbert⁶, B. Torchia¹. 1) Allele Diagnostics, Spokane, WA; 2) Rocky Mountain Hospital for Children at Presbyterian/St Luke's, Denver, CO; 3) Pediatric Genetics, Geisinger Medical Center, Danville, PA; 4) Genomic Medicine Institute/Maternal Fetal Medicine, Geisinger Health System, Danville, PA; 5) Medical Genetics & Neurodevelopmental Center, St. Vincent Hospital & Health Services, Indianapolis, IN; 6) Seattle Children's Hospital, Seattle, WA.

Array CGH (aCGH) analysis provides information regarding the size and gene content of copy number changes, but aCGH results are not always informative as to the chromosomal structure of the alteration. Our laboratory routinely employs fluorescence *in situ* hybridization (FISH) analysis to visualize alterations detected by aCGH. Samples from 416 unique probands with copy number changes identified by aCGH and specimens that could be cultured were analyzed using FISH with BAC probes to visualize chromosomal structure. Many copy number changes, such as recurrent deletions/duplications arising via non-allelic homologous recombination (NAHR), had predictable underlying chromosome structure; however, we identified unexpected complexities in 16 (3.8%) cases. In nine cases, FISH analysis revealed unexpected findings including insertions, cryptic translocations, and supernumerary marker chromosomes, after aCGH detected copy number changes of clinical significance or unclear clinical significance. In the additional seven cases, FISH analysis revealed cryptic insertions or translocations after aCGH detected copy gains that were expected to be benign based on gene content. Although the gene content of the inserted material remains not clinically relevant, these rearrangements may cause gene expression changes at or near the breakpoint sites. Our data is consistent with previously published reports which indicate that up to 3% of copy gains identified by aCGH represent insertions or other unbalanced rearrangements. Clarifying the structure of the aberration by FISH is helpful in determining the most appropriate parental follow-up testing and providing the most accurate genetic counseling, including estimating recurrence risk. This data provides additional evidence to support the importance of FISH to assess cryptic structural changes, which can contribute to abnormal phenotypes.

3000W

Genome sequencing for NICU newborns across the south. K.M. Bowling¹, S.M. Hiatt¹, C.R. Finnila¹, M.L. Thompson¹, M.D. Amaral¹, W.V. Kelley¹, K.M. East¹, D.E. Gray¹, J.M. Lawlor¹, L.G. Hendon³, H. Wienpahl⁴, B.M. Shaw², A.C. Hurst², R. Savich³, G.S. Barsh¹, B.R. Korf², G.M. Cooper¹. 1) HudsonAlpha Institute for Biotechnology, Huntsville, AL; 2) University of Alabama at Birmingham, Birmingham, AL; 3) University of Mississippi Medical Center, Jackson, MS; 4) Woman's Hospital, Baton Rouge, LA.

We here describe SouthSeq, a Clinical Sequencing Evidence-Generating Research (CSER2) effort to perform whole genome sequencing (WGS) for infants in neonatal intensive care units (NICUs) at hospitals in Alabama, Mississippi, and Louisiana. SouthSeq aims to sequence 1,500 newborn babies presenting with multiple congenital anomalies, or other signs of a genetic disorder. The primary goals of SouthSeq are to provide early diagnoses for conditions that might otherwise lead to diagnostic odysseys and demonstrate the utility of WGS as a frontline test with potential to improve intermediate and long-term clinical outcomes. SouthSeq also aims to ensure that genomic testing is applied across diverse communities. Recruitment efforts will focus on patients from rural and minority populations that have been historically under-represented in both medicine and genomics research. Sequencing and analysis for SouthSeq is being conducted at the HudsonAlpha Institute for Biotechnology in Huntsville, Alabama. Newborns with signs of genetic disease will be enrolled at three different clinical sites, including NICUs at the University of Alabama at Birmingham, the University of Mississippi Medical Center in Jackson, Mississippi, and Woman's Hospital in Baton Rouge, Louisiana. Both biological parents will be enrolled when available so that follow-up testing can be performed to determine inheritance. Further, patient families may choose to opt-into return of secondary findings, which focus on pathogenic and likely pathogenic variation identified within ACMG SFv2.0 genes. Seventeen newborns have thus far been enrolled, seven of which have undergone sequencing and analysis. For one patient, *de novo* pathogenic variation was identified in *GATA6*, associated with pancreatic agenesis and congenital heart defects; *de novo* likely pathogenic variation in *PPP2R1A*, associated with intellectual disability, was identified in another. A third individual harbored a maternally inherited variant of uncertain significance in *TPM3*, which has been linked to congenital myopathy. These three individuals all exhibit phenotypes that overlap those that have been previously associated with the given gene. Enrollment has recently started at one NICU, and in June 2018 the two other sites will begin, with an expectation of enrollment of 5-10 infants per week. Preliminary results suggest that we are likely to identify dozens of medically relevant variants in the coming months, and hundreds in the coming years.

3001T

Whole genome sequencing as a tool for genetic diagnosis and gene discovery in children with medical complexity.

M. Snell¹, G. Constain^{2,5}, R.Z. Hayeems^{3,4}, M. Marano⁶, M. Reuter⁷, D. Veenma⁷, S. Walker⁷, S. Lucas⁴, E. Cohen^{3,4,5,6}, R.D. Cohen^{2,5,6,9,10}, S.W. Scherer^{7,9,10}, C. Shuman^{2,9}, R. Mendoza^{1,2,5,6,9}, D.J. Stavropoulos^{8,11}, C.R. Marshall^{1,7,8}, J. Orkin^{3,6}, M.S. Meyn^{1,2,5,9,12} 1) The Centre for Genetic Medicine, The Hospital for Sick Children, Toronto, ON, Canada; 2) Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, ON, Canada; 3) Institute of Health Policy Management and Evaluation, University of Toronto, Toronto, ON, Canada; 4) Program in Child Health Evaluative Sciences, The Hospital for Sick Children, Toronto, ON, Canada; 5) Department of Paediatrics, University of Toronto, Toronto, ON, Canada; 6) Department of Paediatrics, The Hospital for Sick Children, Toronto, ON, Canada; 7) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, ON, Canada; 8) Paediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, ON, Canada; 9) Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada; 10) Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, ON, Canada; 11) Department of Laboratory Medicine and Pathology, University of Toronto, Toronto, ON, Canada; 12) UW Center for Human Genomics and Precision Medicine, University of Wisconsin, Madison, WI, USA.

Children with medical complexity have at least one chronic condition, technology dependence, multiple subspecialist involvement, and high healthcare utilization. We hypothesized that whole-genome sequencing (WGS) has the potential to effectively establish genetic diagnoses for children with medical complexity, and that cohorts of children with medical complexity are enriched for novel genetic disorders. Screening the medical records of 545 children with medical complexity from a tertiary-care based Complex Care Program yielded 106 children (19%) suspected of having a genetic condition despite negative prior genetic testing. Eligible participants were evaluated through a clinical genetic assessment; WGS was performed in parallel with any outstanding conventional genetic testing (including whole-exome sequencing, when appropriate). We used our previously published WGS methods, including bioinformatics pipeline, for clinical genome annotation. Of the first 21 probands with WGS data (including 13 trios, 2 dyad, and 6 singletons), 11 have reportable primary diagnostic variants. Seven pathogenic/likely pathogenic variants in known disease genes were identified, including one intronic variant predicted to affect splicing, one small structural variant missed by targeted clinical testing, and one mosaic variant. There also was a promising variant of uncertain significance. Additionally, *de novo* missense variants were identified in three genes without published disease associations: *FBXW7*, *H3F3B*, and *RAC3*. We addressed the N of 1 issue for these candidate disease genes by using public databases (DECIPHER, GeneMatcher, and ClinVar) to rapidly identify similarly affected individuals around the globe. Collaborations have been established and manuscripts are being drafted that will shortly present our combined evidence for disease causality for these three genes (See 2018 ASHG abstract by Chitayat, et al.) Our initial experience applying WGS to children with medical complexity suggests that trio-based genome-wide sequencing is a high yield testing strategy for this patient population, even when prior genetic testing has been negative. Interestingly, this population appears to be enriched for *de novo* mutations and novel genetic disorders. Future work includes analysing unsolved cases using RNA sequencing, in order to identify and evaluate candidate non-exonic variants present in the WGS data.

3002F

Highly-sensitive quantitative testing enhances molecular diagnosis of mosaic Beckwith-Wiedemann syndrome.

S.W. Baker¹, J. Kelly¹, J. Richards-Yutz¹, M. Hernandez-Meadows¹, K.A. Duffy², M.A. Deardorff², J.M. Kalish², A. Ganguly¹ 1) University of Pennsylvania, Philadelphia, PA; 2) Children's Hospital of Philadelphia, Philadelphia, PA.

Background: Beckwith-Wiedemann Syndrome (BWS) is a multisystem overgrowth disorder caused by genetic and epigenetic alterations of imprinting control regions (ICR) on chromosome 11p15.5. Though great phenotypic diversity is observed, cardinal features of Beckwith-Wiedemann Spectrum (BWSp) include macroglossia, omphalocele, lateralized overgrowth, hyperinsulinism, and embryonal tumor predisposition. We present the results of high-sensitivity molecular diagnostic testing of a large series of BWS patients. **Methods:** Molecular diagnosis of BWS was made using a testing strategy that included methylation-specific quantitative real-time PCR (MS-PCR) of bisulfite-treated DNA to evaluate methylation levels at the paternally methylated ICR1 and maternally methylated ICR2, identification of ICR copy number changes using array comparative genomic hybridization, and Sanger sequencing of the *CDKN1C* gene. BWS molecular diagnostic testing for 968 patients was performed by our laboratory. Blood samples from all patients and tissue samples from 66 patients were tested. **Results:** We confirmed a molecular diagnosis for 293 of 968 patients tested, representing a 30.3% diagnostic yield. Evaluation of the clinical features based on indications for testing demonstrated that this testing cohort represents patients from across the BWSp. Among patients with a molecular diagnosis, 46.1% exhibited loss of methylation at ICR2, 21.5% had gain of methylation at ICR1, 25.9% had paternal uniparental disomy, and 4% of patients exhibited copy number changes of ICR1 or ICR2 or a pathogenic variant in *CDKN1C*. About 20% of patients with methylation defects exhibited evidence of low-level mosaic disease, defined as differential methylation of <10% at ICR1 and/or ICR2. We demonstrate that our assay is highly sensitive and is capable of detecting methylation defects of as low as 3% at ICR1 and ICR2. Additionally, we report a 22.7% increase in diagnostic yield enabled through testing of affected tissue samples from patients whose blood testing was negative. **Discussion:** Our results demonstrate that MS-PCR has an increased sensitivity for mosaicism detection in BWSp (down to 3%) compared to methylation sensitive multiplex ligation probe amplification (MS-MLPA), a technique used by most diagnostic laboratories (sensitivity of 10-15%). In addition, this study highlights the increased sensitivity of BWSp diagnostic testing enabled by identification of low-level mosaicism in blood and tissue.

3003W

De novo missense variants in *TRAF7* cause developmental delay, congenital anomalies, and dysmorphic features. X. Wang^{1,12}, M. Tokita¹, C. Chen¹, D. Chitayat², E. Macnamara³, J. Rosenfeld⁴, N. Hanchard^{1,4}, A. Lewis^{1,4}, C. Brown^{5,6}, R. Marom^{1,4}, D. Novacic^{3,7}, L. Wolfe^{3,7}, C. Wahl⁸, C. Tiff⁸, C. Toro³, J. Bernstein⁹, C. Hale⁹, J. Silver¹⁰, L. Hudgins⁸, A. Ananth¹¹, A. Hanson-Kahn^{8,9}, P. Magoulas^{1,4}, P. Liu^{1,12}, C. Eng^{1,12}, D. Batkovskyye¹, A. Ronza¹, M. Sardiello¹, B. Lee¹, C. Schaaf^{1,13,14,15}, Y. Yang^{1,12}, Undiagnosed Diseases Network. 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) The Prenatal Diagnosis and Medical Genetics Program, Department of Obstetrics and Gynecology, Mount Sinai Hospital, University of Toronto, Toronto, Ontario M5G 1Z5, Canada; 3) NIH Undiagnosed Diseases Program, National Institutes of Health, Bethesda, MD 20892, USA; 4) Texas Children's Hospital, Houston, TX 77030, USA; 5) University of Tennessee Health Science Center, Memphis, TN 38163, USA; 6) Le Bonheur Children's Hospital, Memphis, TN 38103, USA; 7) Office of the Clinical Director, National Human Genome Research Institute, Bethesda, MD 20892, USA; 8) Stanford University School of Medicine, Stanford, CA 94305, USA; 9) Lucile Packard Children's Hospital Stanford, Stanford, CA 94305, USA; 10) Prenatal Diagnostic Center, University of California San Francisco, San Francisco, CA 94158, USA; 11) University of Alabama, Birmingham, AL 35294, USA; 12) Baylor Genetics, Houston, TX 77021, USA; 13) Institute of Human Genetics, University Hospital Cologne, Cologne, 50931, Germany; 14) Center for Molecular Medicine Cologne, University of Cologne, Cologne, 50931, Germany; 15) Center for Rare Diseases, University Hospital Cologne, Cologne, 50931, Germany.

TRAF7 is a multi-functional protein involved in diverse signaling pathways and cellular processes. The phenotypic consequence of germline *TRAF7* variants remains unclear. Here we report missense variants in *TRAF7* in seven unrelated individuals referred for clinical exome sequencing. The seven individuals share substantial phenotypic overlap with developmental delay, congenital heart defects, limb and digital anomalies, and dysmorphic features emerging as key unifying features. The identified variants are *de novo* in six individuals and comprise four distinct missense changes, including a c.1964G>A (p.Arg655Gln) variant that is recurrent in four individuals. These variants affect evolutionarily conserved amino acids and are located in key functional domains. Gene-specific mutation rate analysis showed that the occurrence of the *de novo* variants in *TRAF7* ($P=2.6 \times 10^{-3}$) and the recurrent *de novo* c.1964G>A (p.Arg655Gln) variant ($P=1.9 \times 10^{-6}$) in our exome cohort was unlikely to have occurred by chance. *In vitro* analyses of the observed *TRAF7* mutations showed reduced ERK1/2 phosphorylation. Our findings suggest that missense mutations in *TRAF7* are associated with a multisystem disorder and provide evidence of a role for *TRAF7* in human development.

3004T

Mosaic Somatic KRAS mutation is responsible for Oculo-cerebro-cutaneous syndrome (Delleman-Oorthuys syndrome): A seventh patient report. A. De Leener¹, V. De Keuleneer², A-F. Dekairelle¹, P. De Potter³, P. Clapuyt⁴, Y. Sznajer¹. 1) Centre de Génétique Humaine, Cliniques Universitaires St Luc, Brussels, Belgium; 2) Dermatology, Cliniques Universitaires St Luc, Brussels, Belgium; 3) Ophthalmology, Cliniques Universitaires St Luc, Brussels, Belgium; 4) Radiology, Cliniques Universitaires St Luc, Brussels, Belgium.

Background: Oculo-cerebro-cutaneous syndrome (Delleman-Oorthuys syndrome) represents a multisystemic condition with skin (aplasia cutis congenita, lipomatosis on subcutaneous regions located from any subcutaneous regions), ophthalmic (choristoma and epibulbar cyst), intra medullary and brain lesions. Non ossifying fibromas may develop during the first decade on jaw and throughout life on long bones. Follow-up should guarantee precise diagnosis and prompt management. DNA extracted from femoral fibroma compared to blood DNA, applying whole genome shotgun sequencing, allowed to select variants in 2 genes and KRAS was selected since found in other tissues and absent in unaffected relatives. These findings were confirmed in the second available report so far on 4 patients : this last underlines absence of germline mutation, various degrees of mosaicism and hot spot mutations at codon 13,19,146. Expressivity is highly variable, prognosis factors are not defined so far. **Patient Report and Result:** We report the natural history of girl second child born from unrelated healthy parents. From birth were noted: aplasia cutis on left parieto-occipital region, epibulbar dermoid kyst of left conjunctiva and a subcutaneous mass on nuchal region. At the age of 18 months, milestones were normal, she has no seizure and no behavior problem. Ophthalmic lesion as lipoma's and aplasia cutis are not creating morbidity so far. Standard array CGH was normal. Ultrasound at the posterior cervical mass identified lipomatous tissue. Extracted DNA from lipoma and targeted KRAS gene sequencing identified presence of the c.38G>A (p.Gly13Asp) in 47% of these cells. This mutation is one of the 3 (hotspot) reported so far. **Conclusion:** Present patient is the 7th patient with OCCS identified carrier of somatic mutation in KRAS gene. Clinical diagnosis together with histology and molecular studies enabled appropriate genetic counselling. Precise gain-of-function effect of the missense change in the RAS pathway related to phenotype and indication for mTOR inhibitor therapy remain to be evaluate. **References:** Moog U. et al. J Med Genet 2005;42:913-921 Peacock J. et al. Am J Med Genet 2015;167A:1429-1435 Boppudi S. et al. Clin Genet 2016;90(4):334-342 Marcelis A. and Verstraete L. Pediatric Neurology 2018;83:58-59.

3005F

Rare mosaic variant caller allows for low level variant detection in unmatched mosaic overgrowth samples: Results from thirty-five individuals. J.J. Johnston, J.N. Dudley, M.J. Lindhurst, J.C. Sapp, K.M. Keppler-Noreuil, L.G. Biesecker. Medical Genomics and Metabolic Genetics Branch, National Institutes of Health, Bethesda, MD.

Somatic activating mutations in the phosphatidylinositol-3-kinase (PI3K)/AKT signaling pathway have been implicated in a number of overgrowth disorders including Proteus syndrome and the *PIK3CA*-related overgrowth spectrum (PROS). Detection of these variants is complicated by the mosaic nature of the alterations, often resulting in low levels of the pathogenic variant in samples available for testing. As well, an increasing number of genes connected to these disorders adds to the diagnostic complexity. We have designed a next-generation sequencing panel to detect variants in 11 pathway genes with the ability to expand the panel as necessary. The initial panel targeted exons in *AKT1*, *FGFR1*, *GNAQ*, *KRAS*, *PHLPP1*, *PIK3CA*, *PIK3R1*, *PIK3R2*, *PDGFRB*, *RASA1*, and *TEK*. Libraries were constructed with dual index adapters and run on the Illumina MiSeq with 300 cycles of V2 chemistry on nano flow cells with an average read depth over the targeted exons of 271 reads. Samples without likely pathogenic variants identified at this read depth were rerun on a standard flow cell to achieve an average read depth of 1,735 reads. Using in-house sequence data from a control cohort, statistically significant single nucleotide variation (SNV) was differentiated from position specific background using Fisher's exact test with multiple testing correction. Analysis was restricted to variants with an alternate read count greater than 1. The use of data from a control cohort eliminated the need to use matched samples. LoFreq was employed to detect mosaic insertion-deletion variants in *PIK3CA*. Thirty-five cases with mosaic overgrowth were analyzed. From the nano flow cell data, a total of 6,957 SNVs passed initial filters and 38 variants showed significance that survived multiple testing correction. Of these 38 variants, 16 were in *PIK3CA* with variant allele frequencies (VAF) between 1.8% and 47.4%. Four individuals had variants in *AKT1* with VAF between 7.4% and 30.7%. Variants in *RASA1* (2) and *PIK3R1* (1) were considered likely to be pathogenic. A single *PIK3CA* variant was detected upon rerunning negative samples on standard flow cells, and had a VAF of 0.7%. LoFreq detected two indel variants in *PIK3CA* considered likely to be pathogenic. In total, variants were detected in 26 individuals for a detection rate of 74%, with the majority of individuals having variants in either *PIK3CA* (19) or *AKT1* (4).

3006W

Landscape of chromosomal aberrations among pesticides exposed agriculture workers of Punjab, India. M. Ahluwalia, A. Kaur. Department of human genetics, Guru Nanak Dev University, Amritsar, Punjab, India.

Punjab is an agrarian state of India and in order to mitigate the crop damage caused by pests, use of pesticides is done extensively. Farmers are occupationally exposed to pesticides thus, are more prone to genotoxicity and oxidative stress. In human system, glutathione-S-transferases are important detoxification enzymes involved in the pesticide metabolism. Due to inherited genetic variation, individuals have different metabolic capacity. Literature suggests that polymorphism in *GSTT1*, *GSTM1*, *GSTP1*, and *GSTA1* lead to loss of metabolic function in individual. So, the present study was aimed towards analysis of chromosomal aberrations (CAs), oxidative stress, and distribution of *GST* polymorphisms among agricultural workers. A total of 200 subjects were enrolled in the study which included agricultural workers (n=100) and non-exposed controls (n=100) from Punjab. CAs were analyzed using peripheral lymphocyte culturing technique. Calorimetric techniques were used to analyze the oxidative stress by estimation of malondialdehyde (MDA), reduced glutathione (GSH) levels and superoxide dismutase (SOD) activity. Null *GSTT1* and *GSTM1* distribution was analysed through multiplex-PCR method. For *GSTP1* (rs1695) and *GSTA1* (rs3957356), PCR-based restriction fragment length polymorphism was done. Increased frequency of total aberrant metaphases (TAM) were observed in exposed subjects; 15.2 ± 10.6 (mean $\pm$ SD) as compared to non exposed controls (3.7 ± 3.9); $p < 0.0001$. Significant increase in frequency of metaphases with aneuploidy as well as structural aberrations were observed among exposed subjects (6.98 ± 6.7 ; 7.1 ± 6.2) in comparison to controls (1.2 ± 2.2 ; 2.1 ± 2.8). Also, increased number of clonal and non clonal CAs were observed among exposed subjects. Higher oxidative stress was confirmed in exposed group by significantly higher MDA levels ($p < 0.05$) and lower GSH levels ($p < 0.01$). Complete gene deletion in either of *GSTT1* or *GSTM1* was strongly associated with an increased risk (OR=2.01; 95%CI=1.1–3.6; $p < 0.05$) of genetic damage. Increased frequency of TAMs was observed irrespective of *GST* genotypes except combined null genotype of *GSTT1* and *GSTM1* and homozygous mutant (rs3957356) genotype. Thus, a positive association of pesticide exposure with CA frequency and oxidative stress was observed. Due to health effects of genetic damage and oxidative stress on reproduction, progression of cancer, and other disorders, preventive measures on usage and strict regulations are recommended.

3007T

Non-invasive preimplantation genetic testing for monogenic diseases and aneuploidies using cell free embryonic DNA. S. Madjunkova¹, V. Kuznyetsov¹, R. Antes¹, R. Abramov¹, S. Chen¹, Z. Ibarrientos¹, G. Motamedi¹, K. Glass^{1,2}, A. Baratz^{1,2}, C. Librach^{1,2,3,4,5}, 1) Reproductive Genetics, CReATe Fertility Centre, Toronto, Ontario, Canada; 2) Department of Obstetrics and Gynecology, University of Toronto, Toronto, Canada; 3) Institute of Medical Sciences, University of Toronto, Toronto, Canada; 4) Department of Physiology, University of Toronto, Toronto, Canada; 5) Department of Gynecology, Women's College Hospital, Toronto, Canada.

Introduction: Advances in preimplantation genetic testing (PGT) of human embryos has improved outcomes of assisted reproduction with euploid embryo transfers and selection of embryos free of monogenetic disorders. Along, the limitations of PGT were clearly identified, mainly linked to mosaicism, sampling bias of trophectoderm (TE) biopsy and safety of the procedure. We have showed that using cell free embryonic DNA (cfeDNA) for non-invasive PGT for aneuploidies (NIPGT-A) released into blastocyst culture conditioned medium and blastocoel fluid (BCCM+BF) has the potential to overcome the limitations mentioned above. Our aim was to develop an approach for NIPGT-M and -A screening of embryos from couples at risk and identify the areas to be improved. **Methods:** PGT-A and PGT-M on TE biopsies using NGS (VeriSeq, Illumina), SNP-Karyomapping (Illumina) and direct mutation analysis was performed in four couples at risk for: Fabry disease (*GLA* c.1000-2A>G) (6 embryos), familial exudative vitreoretinopathy (*TSPAN12* del Ex8) (4 embryos), breast cancer (*BRCA1* c.68_69delAG) (2 embryos) and Neu-Laxova Sy (*PHGDH* c.488G>A) (4 embryos). We collected cfeDNA from BCCM+BF from 16 embryos (released within 24-48 hours of culture after d4) under strict aseptic and clean room conditions and blank media cultured on the same dish as a negative control. For PGT-M using TE biopsy Repli-G (QIAGEN) was used for whole genome amplification (WGA) of embryonic DNA. NIPGT samples were amplified with Repli-G WGA (n=4) and SurePlex (Illumina) (n=12). Same assays for mutation detection were used for PGT-M and NIPGT-M. Concordance rates for ploidy between NIPGT-A and mutation detection of NIPGT-M to corresponding TE biopsy samples results were evaluated. **Results:** All NIPGT samples had cfeDNA amplified (23.3±5ng/ul). PGT-A results from TE biopsy showed 11 euploid, 3 aneuploid and 3 mosaic embryos. NIPGT-A results were obtained only for samples amplified with SurePlex at 100% concordance. Samples amplified with RepliG didn't give interpretable NGS results. Full concordance of mutation analysis was present for PGT and NIPGT for *PHGDH* (4/4) and *TSPAN12* (4/4) assays and partial for *GLA* (2/6) and *BRCA1* (1/2) NIPGT-M assays, due to maternal contamination. **Conclusion:** NIPGT-M and NIPGT-A using cfeDNA from BCCM+BF are feasible approach for selection of euploid/unaffected embryos for transfer. WGA systems and sample collection need improvement to reduce maternal contamination and increase sensitivity of NIPGT.

3008F

Semi-automated variant prioritization pipeline for retinal dystrophy diagnostics. B. Guan, D.M. McGaughey, M.J. Reeves, E. Ullah, W.M. Zein, K. Wetherby, D. Blain, A. Turriff, B.P. Brooks, S.J. Tumminia, R.B. Hufnagel, eyeGENE® Working Group. NEI/NIH, Bethesda, MD.

Over 300 genes contribute to various forms of inherited retinal degenerations (IRDs). A major limitation in NGS diagnostics is bioinformatic analyses to sort pathogenic variants from those of benign or uncertain significance. Here, we developed a reliable semi-automated pipeline featuring filtering and prioritization of single nucleotide variants (SNV) and small insertions-deletions (indels) in exonic and intronic regions. We also incorporated tools for structural variants, copy number variation (CNV), sex determination, and sample switch or contamination control. 147 IRD genes were custom-captured using the Nimblegen SeqCap EZ choice from 100 IRD samples from the eyeGENE® program. The sequences were processed for SNV and indel calling following GATK best practices. Variants were annotated, filtered using a custom script incorporating ANNOVAR/InterVar and Variant Effect Predictor/Gemini. Panel sequencing provided 98% of 2374 targeted exons with average depth of >500x. There were 794 rare variants with allele frequency less than 1% in ExAC per sample on average. Semi-automated variant filtering and prioritization produced 0 to 7 variants (median: 3) per sample for manual variant classification based on 2015 ACMG/CAP guidelines. There was 92% of agreement between ClinVar-annotated variants and the pipeline. Correlated genotype to pre-test diagnosis was found in nearly half of patients, in line with previous published reports in literature. Manual examination of all rare variants did not identify additional pathogenic or likely pathogenic variants, suggesting that our automated pipeline did not overlook those. Intronic and exonic variants not in canonical splicing sites but predicted to affect splicing were found in 15% of cases, offering additional potentially significant variants for consideration. Large CNVs were detected in 11% of cases and provided direct evidence for molecular diagnosis in an additional 8% of cases, which includes a previously reported 353bp *Alu* insertion in the *MAK* gene and a 36.4 kb deletion in the *TRPM1* gene. Lesser allele fraction quality control identified one sample contamination and another sample with likely mosaic chrX deletion. Our pipeline improves efficiency of clinical variant interpretation without sacrificing sensitivity of detecting reportable variants. Pathogenic or likely-pathogenic CNVs and splicing variants are also present in a significant number of samples.

3009W

A reliable and robust DNA library preparation chemistry in an automated workflow for improved sequencing performance. A. Kothandaraman, S. Eun, M. Carter, S. Piehl, K. Kushiro. Bioo Scientific, a PerkinElmer company, 7050 Burleson Rd. Austin, TX 78744.

For many clinical research labs, next-generation sequencing is a staple in many of its applications and projects. For this indispensable molecular technique, many researchers rely on commercial DNA library preparation kits for convenience, ease of use, and consistency in performance, which can greatly affect the quality of a researcher's data. Many would be quick to agree that though library preparation may seem theoretically simple for all its steps of end-repair of fragmented DNA, ligation of adapters, and amplification, each process must be carefully tuned to avoid bias, which would result in an uneven representation of some segments of the starting materials. Thus, efforts are made to increase efficiency and minimize bias throughout each step in the library preparation protocol. These reagents and workflow give library yields at 102 nM from 250 ng and 125 nM from 1 ng input of fragmented human male DNA. Using *E. Coli* DNA, the resulting library yield was at 84 nM and 126 nM from 1 ng inputs. Genome coverage ranged from 96% to 94% using bacterial DNA with differing GC-content at a 1x coverage and maintained a robust coverage ranging from 89% to 93.5% at a 10x coverage, which was comparatively higher than another commercial reagent on market. The robustness of PCR-free workflows was also investigated. On top of robust performance of the reagent chemistry, the additional benefit of automation for a complex process prone to error that can come from pipetting, mastermix setup, among other processes is more and more visible in the clinical research application field. This poster reviews the specifications of the library prep chemistry in regards to yield, mapping rate, coverage, and GC-bias mitigation and its optimization on an automated platform.

3010T

Mate pair sequencing: Ushering cytogenetics into the era of personalized medicine. N. Hoppman¹, B. Pitel¹, S. Smoley¹, S. Johnson², J. Smadbeck², M. Webley¹, S. Koon¹, L. Baughn¹, G. Vasmatzis², E. Thorland¹, H. Kearney¹. 1) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN; 2) Center for Individualized Medicine, Mayo Clinic, Rochester, MN.

For many decades chromosome studies were the gold standard of care for patients with abnormal congenital phenotypes. Chromosomal microarray has replaced karyotype as a first-tier test for these patients; however, chromosome studies are still utilized to some degree, particularly following a normal microarray study. Chromosome studies will detect apparently balanced rearrangements such as translocations, but without a molecular understanding of the breakpoints of the rearrangement, those findings most often have uncertain clinical significance. Similarly, chromosomal microarrays often detect copy number gains that may or may not disrupt genes at the breakpoints associated with a phenotype resulting from haploinsufficiency. Parental testing to establish inheritance is recommended in such situations; however, patients are still frequently left with unclear results and no additional clinical testing available to further characterize and classify the abnormality. The Mayo Clinic Genomics Laboratory recently launched a novel clinical test, mate pair sequencing (MPseq), to help clarify uncertain results obtained from chromosome or microarray studies. This assay couples whole genome MPseq with patient-specific, customized analysis to clarify findings detected by other assays in order to provide more clear and molecularly precise results to the patient. In addition to our extensive research and development experience, we have now performed MPseq clinically on nine patients with uncertain karyotype (n=8) or microarray (n=1) findings. MPseq was able to identify and further characterize the abnormality of interest in all nine cases. Of these, four (one translocation and three inversions) were re-classified as pathogenic or likely pathogenic, one (a translocation) was re-classified as likely benign, and four (two translocations, one inversion, and one duplication) remained classified as uncertain. In one case, the rearrangement of interest (a paracentric inversion involving chromosome 5) was significantly more complex than originally appreciated. This summary of our clinical experience with MPseq thus far highlights the unique ability of this novel assay to clarify uncertain results obtained by other cytogenetic methods. Based on this experience, our laboratory has developed algorithms that will be shared for performing MPseq as a follow-up assay for cytogenetic results of uncertain clinical significance in a constitutional setting.

3011F

Telomeric association involving chromosomes Y and 19 in a patient with premature ovarian insufficiency. B. Kar¹, B. Linda¹, P. Aishwarya², S. Kundav². 1) Center for Genetic Studies & Research, The Madras Medical Mission, Chennai, India, Asia,; 2) Institute of Reproductive Medicine & Women's Health, The Madras Medical Mission, Chennai, India, Asia.

Background: Premature ovarian insufficiency (POI) affects about 1% of women below the age of 40 and these patients have a lower chance of getting pregnant. POI has been observed in association with X chromosome abnormalities and with Turner syndrome wherein the patients have a high rate of follicle atresia leading to the termination of ovarian function. Here we present a case of telomeric association between chromosomes Y and 19 observed in a 24-year-old patient with primary amenorrhoea and POI. To the best of our knowledge, this is the first case report of tas(Y;19) that has been associated with POI. **Case description:** The patient visited The Madras Medical Mission after being diagnosed with POI with the hope of conceiving through fertility treatment after 1.5 years of infertility. Clinical evaluation performed on the patient showed webbed neck, high arched palate and a short stature of 145 cm. No axillary hair was present and she had a breast development of Tanner IV stage. Hormone level of AMH was low whereas LH and FSH were high. Clinical pelvic examination revealed a small uterus with an infantile cervix. Ultrasonography also confirmed the presence of a small uterus and ovaries were not visualized. MRI was done for further evaluation and this showed a septate uterus, streak gonads and no evidence of ovotestis. Laparoscopy performed showed streak gonads and histopathology of the gonads confirmed the presence of an ovotestis. The patient was referred for genetic analysis and family history was collected. She had two unmarried sisters who had regular menstruation. Cytogenetic analysis showed mos 46,X,tas(Y;19)(p11.3;q13.4)[20]/45,X[10] karyotype with the short arm of the Y chromosome attached to the end of the long arm of chromosome 19. C banding was done and this showed that the derivative chromosome 19 had two centromeres. FISH for SRY gene (Yp11.32) showed its presence on the derivative chromosome. AZFa, AZFb and AZFc regions that are usually located at Yq were present on the derivative chromosome with no deletion. **Conclusion:** Features of Turner syndrome and the presence of an ovotestis in the patient can be attributed to her mosaic karyotype. Gonadectomy was performed to avoid future risk of gonadoblastoma. This case highlights the need for genetic diagnosis in patients with POI to help in coming up with a proper management of infertility.

3012W

How does DNA break form of three-way translocations? T. Kato¹, H. Inagaki¹, Y. Shinkai¹, A. Kato¹, R. Kawamura¹, M. Tsutsumi¹, M. Taniguchi-Ikeda¹, S. Mizuno², Y. Makita³, N. Saito⁴, N. Ochi⁵, H. Kurahashi¹. 1) Fujita Health University, Toyoake, Japan, Select a Country; 2) Department of Pediatrics, Central Hospital, Aichi Human Service Center, Aichi, Japan; 3) Education Center, Asahikawa Medical University, Hokkaido, Japan; 4) Shin-Koga Hospital, Fukuoka, Japan; 5) Aichi Prefectural Mikawa Aitori Medical and Rehabilitation Center for Developmental Disabilities, Aichi, Japan.

Complex chromosomal rearrangements (CCRs) are defined as rearrangements involving three or more chromosomes with three or more breaks. CCRs can be classified into three groups based on the number of breakpoints and type of arrangement: (i) three-way rearrangements, (ii) double two-way translocations and (iii) exceptional CCRs involving multiple chromosomes and breakpoints. Three-way rearrangements involving terminal exchanges with one breakpoint per derivative chromosome have been thought to be the most common category of CCRs. However, recent advance in molecular methodology has greatly improved the characterization of CCRs, with high-density SNP array and next generation sequencing uncovering more complexity among some simple three-way translocations than previously reported. In this study, we identified six individuals with three-way translocations originally identified using standard chromosome banding and fluorescence in situ hybridization. To characterize the breakpoint junctions of three-way translocations, we applied whole genome mate-pair sequencing. None of the six cases carried a simple structural aberration mediated by three breakages and three repairs. Breakpoint fragments of one or two derivative chromosomes consisted of multiple genomic segments arranged in a disorderly array. On the other hand, the remaining derivative chromosomes had simple, single breakpoint rearrangements. The mechanisms underlying these CCRs may be associated with a chromothripsis-like event. Pulverized chromosomes might be then translocated into another chromosome, resulting in the establishment of a derivative chromosome with complex rearrangements. Unrepaired segment of translocated chromosome might be shifted to different chromosomes with simple rearrangements. Therefore, these previously simple three-way rearrangement cases are more consistent with exceptional CCRs, and suggest this type of highly complex arrangements is more common than previously recognized.

3013T

Sodium acetate induces superoxide dismutase activity change and human chromosome aberrations in lymphocyte. *M. Pongsavee.* Department of Medical Technology, Faculty of Allied Health Sciences, Thammasat University, Patumthani, Thailand.

Sodium acetate is the food additive which can prevent bacterial cultivation in a wide range of acidic region. It is often used to give potato chips a salt and vinegar flavor. Study the effects of sodium acetate on superoxide dismutase activity and human chromosome aberrations in lymphocyte were performed. Four human lymphocyte samples from Thai volunteers were treated with sodium acetate at concentration of 0.04, 0.08, 0.16 and 0.32 mg/ml for 24 and 48 hours. Superoxide dismutase activity assay and standard chromosome culture technique were studied about superoxide dismutase activity and chromosome aberrations. The results showed that 0.08, 0.16 and 0.32 mg/ml of sodium acetate were significantly affected superoxide dismutase activity change on human lymphocyte at 24 and 48 hours ($P < 0.05$). Sodium acetate at concentration of 0.16 and 0.32 mg/ml at 24 and 48 hours caused chromosome aberrations ($P < 0.05$). Sodium acetate significantly effects superoxide dismutase activity change and chromosome aberrations in lymphocyte. Type of food additives, the concentration and the incubation time affect enzyme activity and chromosome aberrations in human.

3014F

Molecular cytogenetic characterization of a female patient with a dicentric X chromosome at breakpoints Xq27. *R.M. Diaz^{1,7}, D. Chung³, A. Nguyen³, D. Palencia², D. Trejo², L. Villalba², A. Lopez², A. Duarte², W. Herrera^{1,7}, C.A. Tirado^{1,3,4,5,6}.* 1) The International Circle of Genetic Studies, Chapter Lambayeque, Chiclayo, Peru; 2) Laboratorio de Citogenética, CARPER-MOR – Laboratorio de referencia internacional, México; 3) The International Circle of Genetic Studies, Chapter Los Angeles, USA; 4) HPA, Minneapolis, MN, USA; 5) Allina Health, Inc. Minneapolis, MN; 6) University of Minnesota, Minneapolis, MN, USA; 7) Universidad Nacional PEDRO RUIZ GALLO, Lambayeque, PERU.

A dicentric X chromosomes with breakpoints at Xq are rare. We report a 16-year-old female patient, who presented at the clinics for a very mild intellectual disability. Cytogenetic studies of peripheral blood (GTG banding) showed one normal copy of the X chromosome. The second copy of the X chromosome was characterized as a dicentric X chromosome. To confirm this abnormality, FISH (In Situ Fluorescence Hybridization) using the centromeric probes CEPX (DXZ1), CEPY (DYZ3) on interphase nuclei and metaphases was performed. FISH results showed three CEPX (DXZ1) signals (centromere). The abnormal X was described as 46, X, dic (X; X) (q27; q27). ish dic (X; X) (q27; q27) (DXZ1 ++). This patient behaves as a Triple X Syndrome (XXX Syndrome). The probability of this patient having offspring with Klinefelter syndrome is 50% if the fetus is male and segregates the dicentric X chromosome, and 50% Triple X Syndrome (XXX Syndrome) if the fetus is female. Genetic counseling and chromosome studies of the mother were recommended.

3015W

Prevalence of ACMG 59 associated conditions for 1,031,941 patients in a large rural hospital system. L. Baas¹, C. Hajek¹, J. Smith², G. Frost¹, K. Rageth¹, M. Moore¹, C.F. Boerkoel^{1,2}, M. Sincan^{1,2}. 1) Sanford Health, Sioux Falls, SD; 2) Sanford Research, Sioux Falls, SD; 3) Sanford EDA, Sioux Falls, SD.

Background: The American College of Medical Genetics and Genomics (ACMG) recommends that laboratories conducting clinical exome or genome sequencing report secondary findings from a "minimum list" of 59 genes associated with 27 rare conditions known to be highly penetrant yet clinically actionable. Multiple studies have investigated the frequency at which these secondary genetic findings occur, reporting figures ranging from 0.5 - 20%. However, a scarcity of literature exists regarding the prevalence of the associated conditions in the general population. Clarifying the prevalence of ACMG 59 associated conditions is useful not only for establishing an evidentiary threshold fully to support the ACMG recommendations, but also for assessing the resources necessary for responding to secondary genetic findings in a clinical setting. To this end, we used International Classification of Disease version 10 (ICD-10) data from the electronic medical record (EMR) to approximate the prevalence of ACMG 59 associated conditions in a population of patients in a large rural hospital system. **Hypothesis:** We expect that our approximation will reflect discordance between the frequency of secondary findings reported in genetic testing and the rate at which the associated conditions are documented in the EMR. **Methods:** We developed a text-matching algorithm linking ACMG 59 associated conditions with ICD-10 codes recorded in the EMR. We estimated prevalence from the number of occurrences of ICD-10 codes matched to ACMG 59 associated conditions. **Results:** We evaluated 10,208,234 total ICD-10 diagnoses for 1,031,941 adult patients seen at a rural healthcare system between July 2016 - July 2017. We identified 4,222 diagnoses corresponding to ACMG 59 associated conditions. These data suggest that ACMG 59 associated conditions occur in approximately 0.4% of individuals in our reference sample. **Conclusion:** Our results are consistent with the expectation that secondary genetic findings are reported during genetic testing at higher frequencies than ACMG 59 associated conditions are recorded in the EMR. We conclude that genetic data must be evaluated in combination with clinical reports in order to comprehend the full disease burden imposed by these conditions.

3016T

Rapid exome-based testing for disorders of neonatal or infantile onset. K. Bliven, M. Kyriss, R. Steiner, D. Allingham-Hawkins, G. Londre, R. Vanneste, F. Xu, J. Weber. PreventionGenetics, Marshfield, WI.

In the United States, monogenic disorders resulting in severe neonatal or infantile illness account for ~20% of infant deaths and ~18% of pediatric hospitalizations (Kingsmore. 2012. PubMed ID: 22872815; Saunders et al. 2012. PubMed ID: 23035047). Several studies have shown that genetic testing in such patients can be a very beneficial first-line test, as the results may provide a molecular diagnosis that allows for early intervention for the patient, proper genetic counseling of the family, and may reduce the time spent in the diagnostic odyssey (Kingsmore. 2012. PubMed ID: 22872815; Saunders et al. 2012. PubMed ID: 23035047; Ceyhan-Birsoy et al. 2017. PubMed ID: 28079900). In an effort to provide a molecular diagnostic test for these individuals, PreventionGenetics developed a Neonatal Crisis Sequencing Panel with Copy Number Variation (CNV) detection via Next Generation Sequencing (NGS). This panel includes 1,155 genes that have been linked to disorders with severe neonatal or infantile onset, and is indicated for patients who present with symptoms such as respiratory distress, hypotonia, gastrointestinal distress, difficulty feeding, failure to thrive, lethargy, seizures, encephalopathy, cardiac defects, organomegaly, unusual facial features, abnormal odor, and/or metabolic disturbances. To date, ~22% of cases had a clear positive result. A possible or contributory genetic cause, such as at least one plausible causative variant of uncertain significance in a gene relevant to the patient phenotype, was identified in another ~56% of cases. Of the positive cases, one proband harbored a *de novo* likely pathogenic variant in *TBX5*, a gene associated with Holt-Oram syndrome. Additionally, we identified an indication of maternal UPD15 in one proband who presented with Prader-Willi syndrome as a neonate. ~22% of submitted samples have been for ongoing pregnancies or fetal demise, demonstrating the value of this diagnostic test to aid in delivery management of fetuses presenting with multiple anomalies. Thus far, the positive yield for this panel is comparable to our whole exome sequencing (WES) rate (26%). Coupled with a shorter turnaround time (<21 days) compared to WES (~6 weeks), this panel is ideal for patients in the neonatal or pediatric intensive care unit who present with a suspicion of a rare genetic disease. If desired (i.e., when a negative or indeterminate report is issued), clients also have the potential to reflex to whole exome sequencing.

3017F

Pipeline improvements enable fast genome sequencing and clinical services. Y. Ding, S. Chowdhury, L.V.D. Kraan, L. Puckett, C. Yamada, S. White, R. Hovey, S. Batalov, N. Veeraraghavan, J. Maria, J. Braun, D. Oh, P. Mulrooney, M. Tokita, K. Ellsworth, J. Cakici, S. Breeding, W. Benson, M. Lurie, S. Cox, D. Dimmock, S. Nahas, S. Kingsmore. Rady Children's Institute for Genomic Medicine, San Diego, CA.

Rapid, clinical Whole Genome Sequencing (rWGS) at Rady Children's Institute for Genomic Medicine (RCIGM) was recently enhanced by implementation of multiple improvements: *In toto*, these enabled achievement of a Guinness World Record for fastest genetic diagnosis by rWGS (~19.5 hours from specimen receipt to provisional diagnosis). As part of this effort, we evaluated the analytic performance of the Novaseq 6000 utilizing various library construction kits and flowcells, for single nucleotide variant (SNV), small insertion and deletion (indel), and larger structure variant (SV) genotypes. Artificial Intelligence (AI) was employed to accelerate the analysis of variants of interest by integration of phenotype and genotype data. At an average coverage of >40X, rWGS provided $\geq 10X$ coverage for 100% coding domain nucleotides in >95% of OMIM genes. The sensitivity of SNV and Indel calls was 99.5-99.9% and 96.1-99.3%, respectively; precision was 99.99%. To date, we have completed rWGS in ~950 individuals, yielding molecular diagnoses in ~35%. Among diagnoses, 72% were determined to have changed medical management. Further pipeline developments will focus on achieving scalability through sample preparation by liquid handling robots, and integration of an end-to-end laboratory informatics management system (LIMS), with a centralized database of sequences and metadata. These improvements will provide a Turnaround Time (TAT) of rWGS in a routine clinical setting, including in-house confirmatory tests, such as Multiplex ligation-dependent probe amplification (MLPA) for detection of deletion and insertion variants, of <1 week. In addition, to meet the needs of on-going clinical trials and research, rapid Whole Exome Sequencing (WES) with spike-in of the Mitochondrial Genome and RNAseq are in development. RCIGM is increasingly providing rWGS for diagnosis of infants in neonatal and pediatric intensive care units at hospitals in the Sanford consortium and other children's hospitals nationwide empowered by these approaches. Ongoing improvements in clinical rWGS instrumentation, sample preparation, AI and pipelines are improving diagnostic yield, shortening time to diagnosis, and decreasing testing cost.

3018W

Assessment of the utility of functional studies in variant classification. I. Thibodeau, E. Kaseniit, B. Mar-Heyming. Counsyl, South San Francisco, CA.

Proper interpretation of genetic variants is crucial for devising appropriate patient and health management strategies. According to ACMG guidelines, experimental evidence describing the impact of a variant on protein function can contribute to either benign or pathogenic classifications. However, the clinical relevance of functional studies is not well-established. Laboratories weigh experimental evidence differently, leading to discrepancies in variant classifications that may result in uncertainty in patient and health management. Therefore, a more standardized approach to assessing the clinical relevance of functional studies is needed. Focusing on GJB2 (non-syndromic hearing loss) and ATP7B (Wilson's disease), we devised a methodology to interrogate the clinical relevance of experimental assays. We identified variants classified as pathogenic or benign based on their allele frequencies in patients and controls but, critically, independent of functional studies. Next, the reported functional impacts of these variants were retrieved from the literature and analyzed. For GJB2, localization to gap junctions, lucifer yellow (LY) transfer, and junctional conductance (JC) assays were interrogated. For ATP7B, complementation studies, localization to the Golgi apparatus, and Cu uptake assays were analyzed. For both genes, the outcomes of localization studies showed the highest concordance with pathogenic classifications. For GJB2, the results of LY and JC assays exhibited significant variations between laboratories and experimental systems for the same variant. Additionally, at least two benign variants were shown to have activities similar to variants that are classified as pathogenic, challenging the strength of the predictive value of these assays. Moreover, several pathogenic variants showed activities similar to the wild type protein in various assays regardless of the gene studied, suggesting that a neutral functional impact in an assay may not be strong evidence of non-pathogenicity. In summary, we devised a strategy to interrogate the utility of functional studies for variant interpretation and tested it on two genes. We found several study-specific caveats that indicate a need for careful and comprehensive assessment of experimental assays. We suggest that the performance of functional studies be evaluated in a gene-by-gene and study-by-study basis prior to weighing the results in the variant classification process.

3019T

Detection rates of postnatal chromosomal microarray testing based on clinical indication. T. Guo^{1,2}, M. Shafiq^{1,2}, R. Singh^{1,2}, Y. Mito^{1,2}, A. Babu^{1,2}, L. J. Edelmann^{1,2}, S. A. Scott^{1,2}, L. Shi^{1,2}. 1) Icahn School of Medicine at Mount Sinai, NY, NY; 2) Sema4, a Mount Sinai venture, Stamford, CT 06902.

Chromosomal microarray (CMA) is a first-tier test for patients with multiple congenital anomalies (MCA), developmental delay/intellectual disability (DD/ID), and autism spectrum disorders (ASD); however, the detection rates of CMA across indications is currently unclear. To determine postnatal CMA detection rates based on clinical indication, deidentified data from 12,589 CMA tests from 2013 to 2017 were assessed. Among these samples, 8,014 (63.7%) were prenatal and 4,575 (36.3%) were postnatal, including 849 (18.6%) with MCA, 1,206 (26.4%) with DD/ID, 669 (14.6%) with ASD, and 1,851 (40.5%) with heterogeneous other indications. All identified copy number variants (CNVs) were classified based on American College of Medical Genetics and Genomics (ACMG) guidelines. CMA testing evolved over time with respect to analytical reporting criteria and included three microarray designs: 44K, 105K, and 180K CGH+SNP (Agilent Technologies). Interestingly, the overall detection rate for variant of uncertain clinical significance-likely pathogenic (VOUS-LP) and pathogenic CNVs was consistent across CMA designs (11.3% for 44K/105K vs. 10.0% of 180K); however, the frequencies of VOUS CNVs significantly increased using the 180K CGH+SNP array (9.4% for 44K/105K vs. 14.8% of 180K). The increased VOUS rate using the higher resolution microarray was due, in part, to the additional reporting of absence of heterozygosity (AOH) that was enabled by the 180K CGH+SNP microarray. Importantly, the detection rates also varied across clinical indications, with VOUS-LP and pathogenic CNVs being reported for MCA, DD/ID, ASD, and other indications being 11.2%, 10.1%, 3.4%, and 11.3%, respectively. In addition, ASD also had the lowest detection rate of AOH (9.0% MCA, 7.1% DD/ID, 4.8% ASD, and 10.0% other), as defined by a single region greater than 10 Mb or when the total autosomal AOH proportion is greater than 3%. However, the overall VOUS rates were similar across the four clinical indications (13.7% MCA, 14.0% DD/ID, 16.0% ASD, and 15.5% other). Taken together, our results indicate that overall postnatal detection rates using CMA platforms designed for clinical cytogenomic testing are ~10%; however, higher resolution platforms also result in an increased VOUS rate. In addition, these data also underscore the genetic heterogeneity between different indications, in particular highlighting the need for higher resolution exon-level CMA and/or gene panel sequencing among individuals with ASD.

3020F

Detection of clinically relevant copy number variation in whole exome sequencing data. F. Xu, C. Glodosky, M. Benzschawel, D. Allingham-Hawkins, J. Weber. PreventionGenetics, Marshfield, WI.

Next-generation sequencing (NGS) approaches such as whole exome sequencing (WES) have been widely used as a standard tool for detecting disease causing single nucleotide variants or small insertions/deletions (indels). However, copy-number variants (CNVs) including large cytogenetic events and exon level deletions and duplications, although a common genetic cause of disease, have not routinely been identified from NGS data in clinical laboratories. In present study, CNV analysis using Golden Helix's VarSeq CNV caller (VS-CNV) was first validated to establish its performance compared to microarray for 109 rare CNVs and LOH events. These positive CNV/LOH controls have a diverse type of clinically relevant CNVs, including triploidy, trisomy, interstitial and terminal deletions and duplications, recurrent microdeletions and microduplications, mosaic chromosomal deletions, unbalanced chromosomal translocations, loss of heterozygosity, and exon level deletions and duplications. The detection rate was 97% for all CNVs detected by chromosomal microarray (CMA) and 77% for exon level CNVs (ranging from single exon to whole gene deletion/duplication) detected by high density aCGH. This approach allowed the detection of expected deletions and duplications of varying sizes, ranging from 196 bp to whole genome duplication, with the highest sensitivity for the detection of CNVs greater than 50 Kb in length. We then applied the VS-CNV algorithm to WES tests for 170 patients affected by a range of genetic disorders and successfully identified five clinically relevant CNVs including two *de novo* CNVs and one sex chromosome aneuploidy (47,XXY). Our results demonstrate that the CNV calling algorithm is capable of detecting a diverse type and size of CNVs based on high quality WES data. The high sensitivity of our WES CNV data supports its utility for CNV detection in clinical settings, although the specificity is expected to be low due to the greater variability in depth of coverage across an exome. Our experience in a clinical diagnostic laboratory supports the importance of the integration of CNV detection with routine WES and shows this strategy results in an increase in diagnostic yield with reduced turnaround time (TAT) but without additional testing and cost.

3021W

The clinical impact and analytical validity of copy number variation calling in rapid genomic sequencing. *S. Chowdhury, S. Batalov, Y. Ding, R. Hovey, S. Nahas, M. Bainbridge, M. Wright, K. Ellsworth, M. Tokita, T. Wong, N. Veeraraghavan, D. Dimmock, S. Kingsmore, RCIGM Investigators.* Rady Children's Institute for Genomic Medicine, San Diego, CA.

Introduction: Implementation of rapid whole genome sequencing (rWGS) has been shown to result in improved diagnostic rates and improved outcomes. A benefit of using rWGS is the potential to simultaneously identify multiple types of genetic variation. The added impact of the inclusion of copy number variation (CNV) within a pediatric patient cohort was assessed. Methods: After consent, blood samples were collected and rWGS was performed via Illumina SBS chemistry to an average genome-wide coverage of 40x. Small variant calling was performed via the Edico DRAGEN, and CNVs were detected using a combination of the CNVnator and Manta callers. Union of the CNV calls generated and custom filtering was applied to ease interpretation burden. The pipeline was optimized to complete alignment and variant calling for small and large genetic variation within 3.5 hours. All reported CNV calls were validated via microarray, Q-PCR, or MLPA. Results: Using a custom filtering workflow, the average number of CNVs to interpret was six per analysis. To date, 443 patients that underwent rWGS resulted in 150 diagnoses for a diagnostic rate of 33%. Of these diagnoses, 24 of the 150 involved CNVs (16%). CNV pathogenic variation ranged from 1.7 kB to whole chromosome abnormalities. All identified variants were orthogonally confirmed with no false positive calls. Seven of the events were less than 50 kB, and thus potentially below the standard resolution and interpretation of microarrays. Three cases involved integrated reporting of a CNV and SNV within the same gene. Conclusion: This study outlines the impact CNV calling has on a rWGS cohort. The ability to rapidly detect multiple types of genetic variation via rWGS will allow greater diagnostic rates, fewer genetic test orders, and faster genetic results, which has been shown to have a significant impact on the clinical utility of genetic testing.

3022T

Enzyme activity determination and DNA extraction obtained from a single 3.2 mm DBS punch. *S.S. Dallaire, C. Gutierrez-Mateo.* Molecular Diagnostics R&D, PerkinElmer, Waltham, MA.

The challenge for laboratories to decide which assays to implement is ever-present with the continual advances in science. With more options available than most laboratories have resources, there is a demand to make testing even more efficient than previously required. In newborn screening, testing can be readily performed using a punch from a dried blood spot (DBS) made from a newborn heel stick. Traditionally, each assay performed in the lab needs an individual punch from the DBS card. There is a limited number of times that a DBS card can be punched; thus capping the number of tests that can be done on that specimen. A common way to maximize the number of analytes than can be tested in a single DBS punch is to favor a multiplex assay over a single-plex test, which also decreases the cost per analyte. A new approach, presented here, is the ability to perform two multiplex assays on the same DBS punch, which will allow obtaining adequate results while reducing the required sample material by half. This study demonstrates how a single DBS punch can be exposed to the assay conditions necessary to measure certain enzyme activities via mass spectrometry followed by immediate processing to extract and measure the DNA by PCR amplification. This new sequential elution method allows one to obtain multiple results on two separate platforms using a single 3.2 mm punch.

3023F

Improving continuity of care and reducing healthcare costs by focusing on how *CYP2C19* and other pharmacogenetic tests are ordered. *M. Moore, B. Tucker, A. Mroch, L. Mullineaux, C. Flom, A. Aifaoui, A. Schultz, J. Reiner, R. Pyatt.* Imagenetics, Sanford Health, Sioux Falls, SD.

Electronic medical record (EMR) systems are designed to improve efficiency and safety by creating automatic responses to specific situations. These responses include grouping certain tests together for specific clinical situations and discontinuation of open test orders when a patient's status changes, such as being transferred or discharged. Although the intention of the automatic response is to reduce decision-making time and the number of outstanding orders in the system; unfortunately, the response can actually increase costs and the patient's time involved. Redundant charges arise from standing orders that include testing that has already been performed and automatically cancelled orders may include testing still relevant to clinical care. Sanford Health's genomic medicine program, Imagenetics, has created an inclusive team of medical doctors, molecular laboratory directors, laboratory genetic counselors (lab GCs), laboratory compliance and quality specialists, the laboratory technicians, and pharmacists. The team has devised methods to evaluate specific pharmacogenetic orders to improve overall quality assurance, continuity of care, and reduce healthcare system redundancies. These implementations have already demonstrated increased quality assurance. Data from the first quarter of implementation noted 308 *CYP2C19* tests, which included 14 duplicate orders (4.5%), 14 automatically cancelled orders (4.5%), and one changed order (0.3%). Of the automatically cancelled orders due to patient discharge, all but 2 had the testing either reordered ($n=7$; 50%) or had already been performed ($n=5$; 35.7%). Through discussions with the hospital clinical informatics operations team, this process has started additional in-patient unit interventions to reduce the number of relevant orders being cancelled. Expanding this process to include other genetic testing performed in our laboratory is also being considered.

3024W

Phenotype clustering and exome reanalysis approach to aid identification of gene-disease associations: A pilot study using genes in the RNA exosome pathway. *R. Ghosh^{1,2}, B. Yuan^{1,2}, J. Salvo², F. Xia^{1,2}, C. Shaw^{1,2}, C. Eng^{1,2}, Y. Yang^{1,2}, P. Liu^{1,2}.* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA; 2) Baylor Genetics, Houston, TX 77021, USA.

The current overall diagnostic rate for clinical exome analysis is around 30%. To potentially improve the diagnostic rate, we performed a retrospective analysis of exomes of patients referred for clinical exome sequencing who shared similar clinical features. As a pilot study, we focused our analysis on the genes in the RNA exosome pathway. The RNA exosome pathway is a multiprotein complex that is essential for proper RNA metabolism. It is composed of nine core subunits, catalytic components with exoribonuclease activity and other associated proteins. Of these, four of the genes in the core subunit have been associated with disease. We annotated 223 patients with variants in the exosome pathway genes with HPO terms corresponding to their clinical features. We also included HPO term annotated phenotype profiles of previously published patients with causative variants in the exosome core genes in our analyses. We next performed cluster analysis to identify groups of patients with similar clinical features and then analyzed the exomes of groups of individuals that shared phenotype profiles. Patients harboring variants in the EXOSC2 gene have been reported to have distinct clinical features compared to patients with variants in the other exosome core genes. Consistent with these published results, we identified patients reported in the literature harboring EXOSC2 pathogenic variants in a different cluster from the published patients with pathogenic variants in the other exosome core components. We identified three patients in our cohort in the same cluster with overlapping features as the published patients with variants in EXOSC3 including motor and speech delay, cerebellar atrophy, joint contractures and hypotonia. Two of these patients harbored the known homozygous pathogenic c.395A>C (p.D132A) variant in the EXOSC3 gene and the other patient harbored a homozygous rare variant in the EXOSC3 gene. We also identified deleterious variants in genes not currently associated with core exosome components in individuals whose clinical features significantly overlapped with clinical features of patients with pathogenic variants in known exosome genes implicated in human disease. Using a phenotype clustering method, our pilot study show that this approach helps identify cohorts with unique combinations of clinical features on which exome reanalysis could be conducted to identify novel disease genes and/or reclassify previously existing variant assertions.

3025T

Report of five families with autosomal recessive bestrophinopathies

(ARB). I. Habibi¹, Y. Falfoul¹, M.G. Todorova², S. Wyrsh³, M. Helfenstein³, D.F. Schorderet⁴, L. El Matri¹. 1) 1.Hedi Rais Institute of Ophthalmology (Department B), LR14SP01, Tunis, Tunisia; 2) 2.Department of Ophthalmology, University of Basel, Basel, Switzerland; 3) 3.Eye Clinic, Lucerne Cantonal Hospital, Lucerne, Switzerland; 4) 4.Institute for Research in Ophthalmology, Sion, Switzerland.

Mutations in *BEST1* cause several phenotypes including autosomal dominant (AD) Best vitelliform macular dystrophy type 2 (BVMD), AD vitreoretinopathy (ADVIRC), and retinitis pigmentosa-50 (RP50). These retinopathies are characterized by vitelliform dystrophy for which ophthalmologists like to describe 5 stages, going from pre-vitelliform to the atrophic final stages. A rare subtype of bestrophinopathy exists with biallelic mutations in *BEST1*. Its frequency is estimated to be 1/1'000'000 individuals. We report 5 cases and establish a phenotype-genotype correlation. All patients were referred because of reduced best-corrected visual acuity, ranging from 0.1/10 to 3/10. All patients showed vitelliform lesions present at the macula, sometimes extending into the midperiphery and along the vessels and optic disc. Onset of the disease varied from the age of three to 25 years. Electrooculogram (EOG) revealed reduction in the EOG light rise in all patients. Molecular analysis revealed the previously reported mutations p.[R141H];[A195V], p.[E35K];[E35K], and p.[R202W];[R202W]. Two families showed novel mutations: p.[Q220*];[Q220*] and p.[L31M];[L31M]. All mutations were heterozygous in the parents. In one family, heterozygous children showed various reduction in the EOG light rise and autofluorescent deposits. ARB, also rare, can be recognized by its phenotype and should be validated by molecular analysis. Phenotype-genotype correlations are hard to establish and will have to wait for more cases.

3026F

Expanding the options of secondary findings in whole exome sequencing. J. Machado, D. Allingham-Hawkins, B. Dorshorst, G. Fischer, R. Steiner, R. Vanneste, J. Weber. PreventionGenetics, Marshfield, WI.

Individuals undergoing whole exome sequencing (WES) often have complex clinical presentations or disorders with significant genetic and phenotypic heterogeneity. At the same time, they are often asked to decide whether they would like to be informed of secondary findings (SF), medically important discoveries unrelated to the reason for testing. The most common option for SF is the American College of Medical Genetics and Genomics' (ACMG) recommended list of 59 genes for disorders that have established interventions for reducing morbidity and mortality. Outside of these 59 genes, however, there are many other disorders that may have current and potential medical management. Moreover, individuals may also harbor variants for recessive disorders, which may have implications for future family planning. In our PreventionGenetics' PGxome® WES analysis, patients may opt-in for variants in the ACMG 59 genes, and options for assessing variants causing other Mendelian predisposition disorders, and recessive diseases for clinically relevant disorders (e.g. OMIM, HGMD). In the PGxome® analysis, 69%, 40%, and 39% of individuals consented to the assessment of ACMG 59 gene variants, predisposition to other Mendelian disorders, and carrier status for recessive diseases, respectively. Likely pathogenic or pathogenic (LP/P) variants were identified in 2%, 24%, and 90% of cases, respectively. Examples of SF for genes in predisposition disorders include *NOTCH3* (CADASIL), *CHEK2* (Breast cancer), *VWF* (von Willebrand disease), *COL5A1* (Ehlers-Danlos syndrome), and *FLG* (Ichthyosis vulgaris). Additional genes beyond the ACMG 59 genes may alert the patient and healthcare provider to future disease risk, which there may or may not be a treatment for healthcare, and for personal management. For recessive disorders, individuals who opted-in to these SF were, on average, carriers of 3 LP/P recessive variants (range 1-8). One trio analysis detected a heterozygous $\Delta F508$ *CFTR* variant in the proband and both parents. The parents were previously unaware of their carrier status, and this knowledge provided additional family planning considerations for them. The option of additional SF beyond the ACMG 59 genes empowers individuals and their families by giving them more information for potential future health care decisions.

3027W

24 hours rapid diagnosis for critically ill infants through trio whole exome sequencing. Y. Qian, B. Wu, Q. Qin, Y. Lu, P. Zhang, H. Wang, W. Zhou. Children Hospital of Fudan University, Shanghai, Shanghai, China.

Objective: Genetic diseases in childhood, especially in neonatal period are difficult to receive clearly diagnosis from clinical information. Some disease require distinctive treatments in a timely manner. Therefore, a rapid and precise molecular diagnosis is necessity for patients under intensive care. We aimed to establish a pipeline, from receiving blood sample to oral report within 24 hours through trio-whole exome analysis (Trio-WES). **Study designs:** The rapid WES pipeline was set up by sequencing of the standard DNA sample of NA12878 on IonTorrent S5 XL platform, and thirty-four patients from Pediatric Intensive Care Units (PICU)/Neonate Intensive Care Units (NICU) with rapid progression without definitive diagnosis were enrolled. The samples were sequenced in parallel with the rapid and the routine WES test pipeline. **Result:** Using the standard DNA, we found the precision was 88.08%, and sensitivity was 81.33%, when compare with the golden standard variants in GIAB for single nucleotide variants (SNVs). Then, we successfully completed thirty-four Trio-WES, the mean sequence depth is 70x. Among the thirty-four seriously illness infants under test, definitive diagnosis was made in twenty-three cases, which including immunodeficiency, inherited metabolic diseases, early infantile epileptic encephalopathy and multi-malformation syndrome. The oral reports were sent to clinicians within 24 hours, Sanger sequencing for variants confirmation as fast as nine hours and routine genetic reports will return within 21 days. Clinical treatment strategies were immediately adjusted according to the rapid genetic diagnosis. There are three cases missed the diagnosis from the rapid pipeline due to insufficient coverage or mismatch mapping, which were detected the pathogenic mutations by routine pipeline. So the diagnosis ratio of the infants was 67.6% (23/34), and the false negative ratio was 8.8% (3/34). **Conclusion:** Using the 24 hours rapid Trio-WES, we successfully diagnosed twenty-three out of thirty-four critically ill infants stayed in ICU. Thus, this rapid genetic diagnosis method can serve as a rapid response genetic diagnosis tool, which can guide clinical decision-making in urgent cases.

3028T

Development and validation of a 30 gene pharmacogenetic genotyping panel: Comprehensive allele and copy number detection. S.A. Scott^{1,2}, E.R. Scott^{1,2}, A.J. Chen¹, Y. Seki¹, M.R. Botton^{1,2}, R. Wallsten¹, A. Owusu Obeng^{2,3}, G. Zhao¹, Y. Yang^{1,2}, P. Nicoletti^{1,2}, H. Shi¹, P. Brake¹, M. Delio¹, L. Shi^{1,2}, R. Kornreich^{1,2}, E.E. Schadt^{1,2}, L. Edelmann^{1,2}. 1) Sema4, a Mount Sinai venture, Stamford, CT; 2) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 3) Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY.

Increasing enthusiasm for implementing clinical pharmacogenetic testing is evidenced by the recent availability of clinical practice guidelines and personalized medicine programs that pre-emptively genotype pharmacogenetic variants. However, there are inconsistencies in content among the available clinical pharmacogenetic panels and some genes and variants implicated in drug response variability are challenging to interrogate due to sequence homology. To develop a novel and clinically relevant pharmacogenetic genotyping panel, a multidisciplinary team evaluated the available evidence and identified 138 common and rare sequence (n=131) and copy number (n=7) variants with clinical validity for drug response testing. These functional variants intersected 30 genes and gene regions, including 11 Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline genes, which together inform on 160 medications. Targeted genotyping and copy number interrogation were accomplished by multiplex single-base extension using the MassARRAY[®] system and multiplex ligation-dependent probe amplification (MLPA), respectively. Analytical validation was accomplished by testing 122 unique DNA samples, which included sequence and copy number variant accuracy (n=119), reproducibility (n=24), and specimen (n=8) controls. Among the accuracy controls were 32 Coriell samples from the 1000 Genomes Project that were selected based on their enrichment of variants included in the pharmacogenetic panel. Coupled with publicly available GeT-RM pharmacogenetic controls, these accuracy samples enabled reference validation material for the majority of interrogated variant alleles, including complex *CYP2D6* structural alleles. Accuracy and reproducibility were both determined to be >99%, indicating that the optimized panel platforms were precise and robust. Importantly, multi-ethnic allele frequencies of the interrogated variants indicate that the vast majority of the general population carries at least one of these clinically relevant variants, supporting the use of this panel for pharmacogenetic research and/or clinical testing programs.

3029F

A rapid and scalable platform for direct modeling of human developmental disorders in mice using CRISPR/Cas9 genome editing. S.A. Murray¹, K.A. Peterson¹, C.S. Heffner¹, C.N. Baker¹, X. Liu², R.J. Manak³, C. Lo². 1) The Jackson Laboratory, Bar Harbor, ME; 2) University of Pittsburgh School of Medicine, Pittsburgh, PA; 3) University of Iowa, Iowa City, IA.

The pace of discovery of the genetic basis of human disease continues to accelerate. However, ascribing causality to a given gene variant is often challenging and thus many cases remain unsolved. Animal models provide powerful tools to both support gene discovery and to serve as research platforms for mechanistic investigation with the mouse providing the ideal mammalian system to model developmental disorders. While CRISPR/Cas9 genome editing has simplified and accelerated pipelines for mouse model creation, timelines are often incompatible with the needs of human gene discovery programs. To address this issue and to support human disease discovery efforts, we have developed a rapid, scalable platform to directly evaluate the causality of novel genes/variants for developmental disorders, taking advantage of the high-throughput mouse embryo phenotyping platform built for the KOMP2/IMPC program. We demonstrate the feasibility of direct phenotypic analysis of F0 ("founder") mouse embryos generated by CRISPR/Cas9 reducing the time required to model a mutation from over a year to a matter of weeks. In proof-of-principle studies, we show that we can efficiently generate and definitively identify expected morphological phenotypes in a mosaic population of edited embryos. Our platform can generate both knockout (null) and precision knock-in (missense) alleles in a single experiment, creating an allelic series to support our understanding of variant function. High-throughput sequencing methods allow us to quantitate the degree of mosaicism in individual specimens, permitting genotype-phenotype associations. Finally, we apply our platform in support of multiple human disease gene discovery efforts, helping solve these cases, and demonstrating the value of our approach. Together with current experiments to increase the efficiency in generation of precision alleles, our platform has the potential to improve the speed and utility of the mouse as a tool to validate and model novel variants for human disease.

3030W

Exome and genome sequencing for Ontario adults with undiagnosed complex hereditary disorders. M. Gutierrez Salazar^{1,2}, K.R. Zakoor¹, S. Shickh^{1,2}, J. Gu¹, N. Watkins^{1,3}, M. Care^{3,4}, S. Horsburgh⁴, O. Morar⁴, J. Murphy^{2,4}, M. Szybowska⁴, D.L. Nevay⁴, M. Aaronson^{3,5}, S. Panchal^{3,6}, J. Stavropoulos⁷, C. Marshall¹, A. Noor¹, S. Khalouei¹, R. Godoy⁶, S. Armel⁸, K. Semotiuk¹, C. Elser⁹, Y. Bombard³, D. Chitayat¹, R. Kim^{1,4,8,9}, J. So^{4,9}, H. Faghfoury^{4,9}, J. Silver^{3,4}, C. Moreh^{1,9}, J. Lerner-Ellis^{1,10}. 1) Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, Ontario, Canada; 2) Genomics and Health Policy Research, Li Ka Shing Knowledge Institute, St. Michael's Hospital, Toronto, ON; 3) Department of Molecular Genetics, University of Toronto, Toronto, ON; 4) Fred A. Litwin Family Centre in Genetic Medicine, Toronto General Hospital and Mount Sinai Hospital, Toronto, Canada; 5) Familial Gastrointestinal Cancer Registry, Mount Sinai Hospital, Sinai Health System, Toronto, ON; 6) Marvella Koffler Breast Centre, Mount Sinai Hospital, Sinai Health System, Toronto, ON; 7) Hospital for Sick Children, Toronto, ON; 8) Princess Margaret Cancer Center, Toronto, ON; 9) Department of Medicine, University of Toronto, Toronto, Canada; 10) Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON.

Since the advent of whole exome sequencing (WES) and whole genome sequencing (WGS), the identification of the genetic bases of disease has improved. Although the clinical utility of WES and WGS has been highly debated, several studies have shown that for individuals with complex medical presentation and extensive clinical work-ups, WES/WGS allows clinicians to circumvent the step-wise targeted single-gene or gene panel tests in favour of a more cost- and time-effective diagnostic tool. Advances in WES/WGS platforms and bioinformatic pipelines have greatly improved scientists' ability to conduct WES/WGS, understand human genetic variation associated with disease, as well as improve the ability to provide molecular diagnoses for individuals with complex medical presentations. In Ontario, the Ministry of Health and Long-Term Care (MOHLTC) lacks eligibility criteria for publicly-funded WES/WGS in adults. Therefore, adults with suspected hereditary conditions are often denied coverage. This lack of access has significant clinical implications. In the adult genetics clinic in Toronto, which services over 1,600 patients/year, <30% reach a diagnosis despite extensive diagnostic investigations. As a result, providing guidance for prognosis, appropriate clinical management and treatment, and recurrence risk for patients and families remains an enormous challenge. In response, our research initiative provides WES/WGS for adult patients that were denied funding by the MOHLTC but for whom standard testing was uninformative. We currently recruit from three Mount Sinai Hospital-affiliated clinics and have enrolled 87 families (155 individuals) presenting with breast and gastrointestinal cancer, cardiovascular, neurological, musculoskeletal, or renal disorders. To date, 42 families (71 individuals) have undergone WES or WGS analysis and have been counselled by genetic professionals on their results. Here, we report a diagnosis yield of 23.8% using WES or WGS, a figure that is consistent with the 20-25% diagnostic yield of ES reported in other comparative studies using only pediatric neurodevelopmental or adult cardiac cohorts. Furthermore, this study addresses the importance of both disease type and analytical context (singleton vs. trios) to develop recommendations for improving analysis of WES/WGS. These findings will not only help to generate guidelines for appropriate use in adults, but will strengthen justifications for using WES/WGS as a first tier clinical test.

3031T

The impact of depth of coverage from sample quality in whole genome sequencing. *K. Park, H. Kwon, S. Lee, M. Roh, D. Seo.* MacroGen Corp., Rockville, MD.

The depth of coverage in sequencing data is one of the important factors to determine whether the data are suitable for scientific researches and clinical studies. As a sequencing service provider, we investigated the depth of coverage using more than 5,000 samples from whole genome sequencing (WGS). Many samples had passed, but some had failed from our QC criteria, called the DOC, measuring difference from Poisson distribution, interquartile range, and mode of depth of coverage. One of the factors that caused DOC failure was the percentage of G and C bases (GC%) in sequencing data, which should be about 41% in the human genome. Among the samples with DOC failure, the percentages of GC in most samples had increased, ranging up to 50% in sequencing data. To identify the cause of GC% alteration, we measured the quality of samples using KAPA hgDNA Quantification and QC kit (Roche, CA, USA), and we observed a negative correlation between Q ratio from this kit and GC%, which indicated that lower quality samples were more likely to have a higher percentage of GC contents. In conclusion, it is challenging, but maintaining a high quality of sample can guarantee to obtain a high quality of sequencing data. .

3032F

Improving quality, cost and turnaround time using the Illumina NovaSeq 6000 System for rapid whole genome sequencing. *L. Van Der Kraan, L. Puckett, C. Yamada, S. White, R. Hovey, S. Batalov, J. Braun, P. Mulrooney, M. Janes, N. Veeraraghavan, M. Tokita, K. Ellsworth, S. Chowdhury, S. Nahas, Z. Bezares-Orin, Y. Ding, S. Kingsmore.* Rady Children's Institute for Genomic Medicine, 7910 Frost St, San Diego, CA, 92123.

The Clinical Genome Center (CGC) at Rady Children's Institute for Genomic Medicine (RCIGM) is a CAP and CLIA certified clinical laboratory specializing in rapid whole genome sequencing (rWGS) for acutely ill infants and children with suspected genetic conditions. Historically, for diagnostic laboratories, delivering sequencing data in a timely manner while maintaining sequencing performance at a low cost has been challenging. At RCIGM's inception, initial instrumentation included Illumina's HiSeq 2500 (Rapid Mode) and 4000 sequencing platforms. Early last year, Illumina introduced the NovaSeq series to the genetic community. At RCIGM we sought to compare the performance of the NovaSeq 6000 to the standard Illumina sequencing platforms, the HiSeq 2500 and 4000. We validated the NovaSeq 6000 for clinical laboratory use and implemented it for routine rWGS testing. Validation of the NovaSeq 6000 was performed using the same methods as used for the HiSeq 2500 and 4000 with Coriell gold standard (CEPH and AJ) trios, in addition, to previously-tested research-consented patient samples reported from the CGC. Library preparations were completed using either the Illumina TruSeq PCR-Free or KAPA Hyper Prep (Roche) kits. Sequencing on the NovaSeq 6000 was performed using the S2 flowcell with a 6-plexed, single indexed 13.5 nM library loaded on one side with 2x101 cycles. Metrics following sequencing on the NovaSeq 6000 produced an average yield of 932 gigabases (Gb) per flow cell, >78% clusters passing filter and >93% Q30. These metrics were comparable and better than that following clinical validation and performance of the HiSeq 2500 and 4000 systems. In particular, with regards to sequencing run time and cost, the average sequencing run time for rWGS on the NovaSeq 6000 is ~24 hours for 6 genomes, on the HiSeq 2500 ~26 hours for 1 genome and on the HiSeq 4000 ~40 hours for 6 genomes. The cost savings per genome on the NovaSeq 6000 is 66.8% when compared to the HiSeq 2500 and 5.6% when compared to the HiSeq 4000. Thus, the NovaSeq 6000 system decreases run time and cost per genome while maintaining and improving the quality of its sequencing data; moreover, it dramatically advances our rWGS approach in the neonatal (NICU) and pediatric intensive care (PICU) units by providing us with faster, higher throughput and better quality sequencing data in a more economical way.

3033W

Solving the mystery of rare genetic disorders: An integrated approach to accelerate diagnosis of rare diseases, facilitate gene discovery and improve management. X. Wang¹, M. Harr¹, D. Watson¹, F. Mafra¹, C. Chung¹, E. Bhoji¹, K. Stumm¹, A. Wenocur¹, L. Tian¹, M. March¹, D. Li¹, C. Hour¹, R. Pellegrino¹, P. Sleiman^{1,2,3}, H. Hakonarson^{1,2,3}, A. Santani^{1,4}. 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Department of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA; 3) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 4) Department of Pathology, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

To benefit patients with rare genetic disorders, it is increasingly critical to create an infrastructure where genome analysis is integrated with functional analysis of novel genes and variants. Integration of sequence data with comprehensive clinical evaluations and advanced machine learning-based algorithms are needed to facilitate rapid and efficient analysis of genomic data. Candidate gene and variants subsequently require further evaluations to demonstrate an impact on protein function. To address these challenges, we have established an infrastructure that integrates a biorepository, a sequencing and microarray facility, an informatics and analysis team, a clinical team and a translational research laboratory. The workflow includes: 1) sample preparation and sequencing; 2) a genomics pipeline including sequence mapping, alignment, variant calling, variant annotation and filtration, 3) A custom-developed phenotype-driven genome analysis algorithm and 4) infrastructure to facilitate in vitro cell studies and functional assays in model systems. The sequencing and informatics workflow was validated by detection of known variants in reference samples with an analytical sensitivity and specificity for SNVs and indels of >99%. The analysis of genomic data primarily relies on leveraging the clinical presentation of the patient. Our analysis strategy utilized our custom variant prioritization algorithm and was validated by performing blinded exome analysis on 32 randomly selected probands who had previously undergone diagnostic exome sequencing by reference laboratories. The causal variant was correctly detected in all cases, and ranked at No.1 for 29/32 cases and in the top 3 for the remaining 3 cases. We report substantially reduced turn-around times with identification of high priority variants within 2 hours after sequencing. This system accelerates the comprehensive analysis of patients with unsolved rare disorders and, coupled with translational follow-through on variants and novel candidate genes, supports identification of novel therapeutics. Using specific case vignettes, we will describe how this infrastructure enabled the diagnosis and understanding of rare genetic disorders. We will highlight successful strategies that other programs can use to facilitate disease diagnosis, impact clinical management and advance medical research.

3034T

Genotype-phenotype analysis of 523 pediatric patients referred for genetics evaluation who underwent whole exome sequencing. M.N. Ziats¹, A. Ahmad², E. Ames², J.A. Bernat², R. Fisher², M. Glassford², M. Hannibal², J. Jacher², C.E. Keegan^{2,4}, K. Lee², T. Marzulla⁴, B.C. O'Connor², S.C. Quinonez^{1,2}, L. Seemann⁴, L. Turner⁴, S. Bielas⁴, N.L. Harris⁵, J.W. Innis^{1,2,4}, D.M. Martin^{2,4}. 1) Department of Internal Medicine, University of Michigan; 2) Division of Pediatric Genetics, Metabolism, and Genomic Medicine, Department of Pediatrics and Communicable Diseases, University of Michigan; 3) Division of Medical Genetics, Stead Family Department of Pediatrics, University of Iowa; 4) Department of Human Genetics, University of Michigan; 5) Children's Clinical Trial Support Unit, University of Michigan.

Clinical whole exome sequencing (WES) has high potential for uncovering phenotype-genotype relationships that guide discovery and clinical care. Here we present WES genotypes and clinical phenotypes from 523 patients referred to the Pediatric Genetics clinic at Michigan Medicine. Patients were systematically phenotyped by presence or absence of abnormalities for 13 body/organ systems, including autism spectrum disorder (ASD), developmental delay/intellectual disability, cancer, multiple congenital anomalies, dysmorphic features, and behavioral phenotypes. Overall, 30% of patients who underwent clinical WES had an identified pathogenic variant. Notably, 67% of patients had a variant of uncertain significance (VUS; n = 244) or gene of uncertain significance (GUS; n = 116), and 23% had no pathogenic or uncertain variants identified. The most common clinical phenotypes among patients with a pathogenic variant were developmental delay (83%) and neuromuscular system abnormalities (81%); 20% had ASD and 42% had multiple congenital anomalies. The average number of body systems affected for patients with a pathogenic variant was 5.89, vs 6.0 for patients with VUS only and 6.12 for patients with GUS only. For patients with no identifiable variant, this average was 4.6, a significant difference when compared to patients with a pathogenic variant (one-way ANOVA P<0.00001). Among patients with ASD and a pathogenic variant (N=31), 42% had a variant in a gene involved in transcriptional regulation, whereas 23% of patients had a variant involved in synaptic functioning, 16% were involved in ion channels, and 6% were involved in another pathway. In patients with ASD, we analyzed the relationship between implicated molecular pathway and severity of clinical phenotype as inferred from the number of body systems affected. Interestingly, ASD patients with a pathogenic variant in a gene involved in transcription had a higher number of systems affected than those with a gene involved only in synaptic functioning or ion channels (7.4, 4.9, and 5.2, respectively), although this difference did not reach statistical significance (p=.12) perhaps owing to small sample size. In summary, our cohort of 523 pediatric patients with WES exhibits clinical features that are enriched for multiple organ system involvement. Variants identified in this cohort of patients can be leveraged to inform mechanistic understanding of individual disorders with a genetic component.

3035F

The use of multi-mode reader for the quantification of libraries for next generation sequencing systems. *J. Adigun, C. Nolan, F. Davis, J. Kennedy, W. Lowe, S. Ayres, J. Hatton, S. Hofherr.* GENEDX, Gaithersburg, MD.

Next generation sequencing continues to be an important tool in the determination of causative variants in the pathology of many genetic disorders. The accurate quantification of completed libraries for sequencing on next generation sequencing systems is an important factor in the successful and cost-efficient generation of sequencing data. While the cost of data for generation for massive parallel sequencing has fallen considerably in recent years, inaccurate quantification of libraries continues to lead to costly repeat of the sequencing process. Over the years, many technology and processes have been implemented by laboratories to achieve accurate quantification of libraries. Such include the use of Real-time (quantitative) polymerase chain reaction process, QPCR and/or the use of instrumentations such as the Bioanalyzer and TapeStation systems. More often than not, these processes are cumbersome, not amenable to high throughput laboratories, time consuming and may lead to high recurring cost in the laboratory. Here we describe the use of a reasonably priced microplate reader, the Synergy HTX Multi-Mode reader, with fluorescence-based dye, that is amenable to high throughput laboratories for accurate and dependable quantification of libraries. This workflow is very quick, accurate and cost-efficient in the clinical as well as research laboratories.

3036W

Genome-wide SNP array analysis reveals contribution of mosaicism and chimerism in pediatric genomic diagnostics. *B. Weckselblatt¹, L.K. Conlin¹, J. Fan¹, E. Zackai², M. Dulik¹, N.B. Spinner¹, M. Luo¹.* 1) Division of Genomic Diagnostics, Department of Pathology and Laboratory Medicine, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Division of Human Genetics, Department of Pediatrics, Children's Hospital of Philadelphia, Philadelphia, PA.

The chromosomal microarray is a first-tier diagnostic test for patients with congenital anomalies, neurodevelopmental disorders, and/or other clinical phenotypes. We have previously validated a SNP-based array to detect mosaicism and chimerism in at least 5% of cells, and to interpret the timing of these events as pre-zygotic or post-zygotic in origin. In this retrospective analysis, we reviewed the SNP array data from patients received by our clinical laboratory to investigate the contribution of mosaicism to pediatric genomic diagnostics. Since May 2008, we have performed SNP array analysis for 14,078 samples from 13,816 patients. Mosaic changes were detected in ~1.8% (N=251) of cases, accounting for ~15% (N=206) of abnormal reports, while ~18% of the mosaic findings have uncertain clinical significance at this time. When possible, cytogenetic studies confirmed array results and resolved variants that were associated with structural rearrangements. The most common mosaic finding was chromosomal aneuploidy, with the sex chromosomes being most frequently involved, followed by autosomal trisomies involving 13 different chromosomes. Whole genome mosaicism consistent with fertilization errors, including mosaic triploidy and several types of chimerism, were detected in multiple tissues from 10 individuals, with mosaic whole genome paternal isodisomy (androgenetic chimerism) as the most frequently observed genome-wide finding. The remaining results involved supernumerary marker chromosomes, interstitial deletions and duplications, unbalanced translocations, uniparental isodisomy (entire chromosome or segmental), as well as more complex patterns that do not fit known rearrangement models. For 13 individuals, the array had sufficient sensitivity to detect a large abnormality of the B allele frequency present in less than 5% of the cell population, though this level is too low to detect a significant change in the LogR ratio, thus we could not determine whether the region denotes a copy number change or loss of heterozygosity. This review of the last ten years of SNP array testing demonstrates that mosaicism is a significant contributor to the diagnostic yield of SNP arrays, and provides insights into the mechanisms and origins of chromosomal abnormalities.

3037T

The potential significance of Alu elements in genetic diseases as assessed through identification of rare insertion events via whole genome sequencing in a cohort of pediatric patients. *M.S. Wright, D.G. Tamang, S. Chowdhury, D.P. Dimmock, S.F. Kingsmore, M.N. Bainbridge.* Rady Children's Institute for Genomic Medicine, San Diego, CA.

Alu elements are abundant in the human genome, but it is estimated that less than 0.5% of insertion sites are polymorphic. The current estimate is that de novo Alu insertion events arise in 1:20 births. Alu elements have been implicated in multiple monogenic diseases and cancer development when inserted into new chromosomal locations. We therefore sought to address the potential clinical significance of Alu elements in pediatric disease using genome wide surveys of whole genome sequence data from a cohort of 309 families sequenced through Rady Children's Institute for Genomic Medicine. This included 702 individual genomes comprising 178 trios, 32 proband/parent duos, 94 proband only, and 5 matched germline-somatic samples. We developed a computational pipeline that utilizes a split-read mapping approach to detect Alu insertions. For paired reads in which one read maps to an Alu sequence, the chromosomal location for the other read was identified. To minimize mapping artifacts and remove common polymorphic Alu sites, count data for each insertion site in each individual genome was normalized to population-wide counts, where only locations that exceeded greater than 2 standard deviations above the normalized population counts were retained. This enabled us detect rare events. Using this approach we detected >16,000 rare intergenic Alu insertions in our cohort, as well as >2000 intragenic sites in 2008 genes. Strikingly, we detected Alu insertions in 313 OMIM disease genes, many of which were in UTR sequences. The availability of WGS data from trios allowed us to further estimate de novo retrotransposition rates and the potential contribution of de novo events to genetic diseases. Our results highlight that Alu insertions may play a larger role in the genetic disease burden than previously estimated in pediatric populations, as well as demonstrates the role Alu elements play in genome diversification.

3038F

A simplified, rapid approach to targeted resequencing of dried blood spot samples. *D. Mandelman, S. Rozenzhak, A. Kraltcheva, C. Van Loy, A. Harris.* Thermofisher Scientific, 5781 Van Allen Way, Carlsbad, CA . 92008.

Dried blood punches (BPs) are a common sample type in hospitals - especially for newborns – and in remote field studies. We present a simplified protocol for preparing libraries from dried blood punches for analysis on next-generation sequencing (NGS) platforms using Ion AmpliSeq™ chemistry. AmpliSeq™ is a PCR-based NGS library prep method that uses targeted primer panels to amplify DNA or RNA regions of interest for most any research application, including inherited and infectious disease. Current BP protocols require laborious, time consuming sample purification from the BP followed by sample quantification. This problem is only exacerbated as sample number increases. Our protocol enables direct use of BPs in PCR. To verify the efficacy of our method, a variety of AmpliSeq™ panels were tested on BPs using the new protocol. When compared to traditionally purified, quantified samples, no differences in sequencing read quality or variant calling were observed between the BPs and purified genomic DNA. Simplified sample prep methods like this will facilitate the uptake of NGS in settings where BPs are a common sample type. *For Research use only. Not for use in diagnostic procedures.*

3039W

Identification of novel *CYP2D6* haplotypes that interfere with TaqMan copy number analysis. A. Turner^{1,2}, P. Aggarwal^{1,2}, E. Boone³, C. Haidar⁴, M. Relling⁴, U. Broeckel^{1,2}, A. Gaedigk². 1) The Medical College of Wisconsin, Milwaukee, WI; 2) RPRD Diagnostics, Wauwatosa, WI; 3) The Children's Mercy Hospital, Children's Research Institute, Kansas City, MO; 4) St. Jude Children's Research Hospital, Memphis, TN.

Cytochrome P450 2D6 (*CYP2D6*) is a highly polymorphic gene encoding an enzyme critical in the metabolism of up to 25% of commonly prescribed drugs. There are over 100 described star allele (*) haplotypes, including frequent gene conversions, deletions, duplications and gene-fusions (D6/D7 hybrid), many of which have altered enzyme activity. Accurate copy number (CN) calling is critical in determining patient drug response. TaqMan CNV assays are a commonly used methodology for CN analysis, utilizing specific primers and labeled probes to evaluate the CN state of a genomic region of interest. Polymorphisms within the primer or probe target sequence can generate a false positive for a CN loss. Currently three *CYP2D6* intra-gene regions can be tested using commercially available TaqMan assays: intron 2, intron 6 and exon 9. Gene-fusions or conversions present as discrepant CN calling between the three regions. During clinical testing, we identified samples with CNV inconsistent with known haplotypes. To characterize the haplotypes that lead to these CNV results, we performed long range PCR, Next-Gen sequencing and allele-specific Sanger sequencing. Sequence analysis of 20 patients identified six novel sub-alleles containing polymorphisms within the TaqMan assay probe binding sites, causing a false-positive copy number loss. In total, four novel haplotypes (sub-alleles) with SNPs within the intron 2 and 6 probes were identified (*1 sub-allele 1: rs770138443, rs78854695; sub-allele 2: rs781257354, rs37522240. *2 sub-allele 1: rs180847475, rs186133763; sub-allele 2: rs370010370). Analysis of samples which showed CN loss in only exon 9, identified new *10 and *17 sub-allele haplotypes. Two additional samples with exon 9 CNV contained hybrid/deletion arrangements within the probe region that affected CN analysis. In summary, performance of widely used TaqMan CNV assays is affected by the presence of sequence variation within the probe target regions. The identification of these haplotypes and further characterization of these structural variations will be important for clinical interpretation and could directly impact patient care.

3040T

Evolution of dihydropyrimidine dehydrogenase (DPD) diagnostics in a single center in a time-period of eight years. B.J.C. van den Bosch¹, A.D.C. Paulussen¹, M. Breuer¹, M. Lindhout¹, D.C.J. Tserpelis¹, A. Steyls¹, J. Bierau¹, M.J.H. Coenen². 1) Department of Clinical Genetics, MUMC+, Maastricht, The Netherlands; 2) Radboud university medical center, Radboud Institute for Health Sciences, Department of Human Genetics, Nijmegen, The Netherlands.

Background: Treatment with fluoropyrimidines, the main chemotherapeutic agents used in many types of cancer, is not well tolerated in a subgroup of patients. The enzyme responsible for degradation of fluoropyrimidines is dihydropyrimidine dehydrogenase (DPD), the first and rate-limiting enzyme of the pyrimidine degradation pathway. DPD dysfunction leads to an increased exposure of active metabolites, which can result in severe or even fatal toxicity. **Objective:** Fluoropyrimidine treatment can be optimized based on DPD activity.

Design: We provide an overview of eight years DPD diagnostics (n=1194). In this time-frame the test has evolved from a single enzyme measurement using Ultra-High Performance Liquid chromatography (UHPLC) in peripheral blood mononuclear cells (PBMCs) to a combined enzymatic and genetic test of four variants in the *DPYD* gene (*DPYD**2A, *DPYD**13, c.2846A>T and 1129-5923C>G). **Results:** In the group tested for four variants (n=814), patients with either one variant have a lower enzyme activity than the overall patient group. The majority of patients with the *DPYD**2A variant (83%) consistently showed a decreased enzyme activity. Only 24 (25.3%) of the 95 patients with a low enzyme activity (tested for four variants) carried a variant. Complete *DPYD* sequencing in a subgroup with low enzyme activity and without *DPYD**2A variant (n=47) revealed 10 genetic variants, of which four have not been described previously. In summary we show that there is a weak correlation between DPD enzyme activity and the presence of the variants tested. With respect to this correlation we observed a sensitivity of 0.25 (24/[24+71]) and a specificity of 0.93 (667/[667 + 52]), i.e. most patients do not carry a variant as expected and all these patients have a normal DPD enzyme activity (high specificity). **Conclusions:** We did not observe a strong link between *DPYD* genotype and enzyme activity. There is no doubt that DPD status should be determined before treatment with fluoropyrimidines to save patients' lives and prevent unnecessary side effects. Our study in combination with literature shows that there is a discrepancy between the DPD enzyme activity and the presence of clinically relevant SNPs. Therefore at this moment a combination of a genetic and enzymatic test is preferable for diagnostic testing.

3041F

A novel NGS solution for combined PGT-M and PGT-A: Analysis for β -thalassemia. *M. Brockman, K. Warren, M.J. Jasper.* RHS Ltd, Thebarton, South Australia, Australia.

Thalassemia, caused by mutations within genes involved in haemoglobin production, including the haemoglobin subunit beta gene (*HBB*), has become one of the most prevalent blood disorders around the world and is a frequent target for Preimplantation Genetic Testing for Monogenic disorders (PGT-M). The ability to combine aneuploidy detection (PGT-A) and PGT-M, by adding gene-specific PCR primers to a Whole Genome Amplification (WGA), maximises the screening opportunity for a single embryo biopsy. The aim of this study was to demonstrate the application of a novel Target Sequence Enrichment (TSE) protocol using the DOPlify® kit (RHS Ltd) for combined PGT-M for β -thalassemia and PGT-A by Next Generation Sequencing (NGS). Manually sorted 5-cell aneuploid samples (Coriell Institute) modelling a trophectoderm biopsy were subjected to either standard WGA (no enrichment) or WGA using the DOPlify® TSE protocol (WGA+TSE) with *HBB*-specific PCR primer sets targeting exons 1-3. Successful sequence enrichment was determined by performing a *HBB*-specific PCR, and further enrichment of *HBB* sequence was achieved by pooling *HBB*-specific PCR products with the WGA+TSE prior to library preparation and NGS. All samples were sequenced according to the standard PG-Seq™ kit protocol (RHS Ltd) for 48 samples using a MiSeq® instrument (Illumina). Correct aneuploidy diagnoses were achieved for all samples. The NGS protocol used was typical for PGT-A, but was not expected to yield the depth of reads required for PGT-M in the absence of enrichment. As expected, <1x depth of coverage and <10% breadth of coverage was achieved for *HBB* with no enrichment. In contrast, WGA+TSE achieved 5x depth of coverage with 80% breadth of coverage, while further enrichment achieved a depth of coverage of 200x with 100% breadth of coverage. The targeted and concurrent enrichment of *HBB* sequence during WGA provided a suitable depth of coverage for PGT-M without compromising PGT-A results. RHS' TSE protocol provides a unique opportunity to develop a multiplex panel targeting the most common disorders tested by PGT-M.

3042W

A simple NGS method for detection of sequence variants causing alpha and beta thalassemia. *F. Oliv, M. Uzunel.* Devyser AB, Stockholm, Sweden.

AIM: To develop a simple NGS method for detection of mutations causing alpha and beta Thalassemia. **BACKGROUND:** Thalassemias are inherited blood disorders characterized by abnormal hemoglobin production. Depending on the type and number of mutations, the symptoms can vary from none to severe forms where stem cell transplantation is the only curative treatment. Today, many different methods are available for mutation analysis of thalassemia patients. These include ARMS-PCR, restriction-enzyme PCR, GAP-PCR, Sanger sequencing, LIPA and MLPA. The use of different methods may be time-consuming and expensive, especially when analyzing less common deletions. **METHOD:** We have developed an amplicon based NGS method using only ONE oligo-mix to detect virtually all known mutations for alpha and beta Thalassemia. The method includes sequencing of the *HBB*, *HBA1* and *HBA2* genes. We have also included primers for direct detection of the most common deletions in alpha and beta Thalassemia ("Gap-PCR"). In addition, we amplify regions upstream and downstream of the above-mentioned genes to find less common deletions using copy number variation (CNV) analysis. The upstream regions also include regulatory sequences, i.e. HS-40 (*HBA*) and LCRB (*HBB*). The protocol is user-friendly, requires less than one hour of hands-on time and only 10 ng of DNA is needed. **RESULTS:** Libraries have 100% uniformity at $\geq 0.3x$ mean coverage. We have analyzed 125 clinical samples with known mutations. All mutations were also detected with the Devyser Thalassemia kit. Interestingly, we could find additional mutations in 15% of the samples; mutations not previously detected or tested by the clinical laboratory. CNV analysis could successfully detect deletions and triplications (anti-3.7 triplication). **SUMMARY:** The Devyser Thalassemia kit is simple, user-friendly and efficiently detects mutations causing Thalassemia.

3043T

CNVs inherited from insertion carriers are rare in a large cohort. R.D.

Burnside¹, M.K. Rudd¹, J. Schleede¹, I. Gadi¹, P. Papenhausen¹, K. Phillips¹, V.R. Potluri², H. Risheg³, J.H. Tepperberg¹, J. Wiley¹, S. Schwartz¹. 1) Laboratory Corporation of America, Research Triangle Park, NC; 2) Laboratory Corporation of America, Houston, TX; 3) Laboratory Corporation of America, Seattle, WA.

Familial follow-up studies to CNVs are important for interpreting the pathogenicity of some CNVs and for proper assessment of recurrence risk. qPCR as a follow up assay has been adopted in many labs due to faster turn-around-time and significant cost savings over FISH. One limitation to qPCR is that this assay cannot determine whether an apparently de novo CNV is actually inherited from a parent with a balanced structural rearrangement, nor can it determine the location of copy number gains. The goal of this project was to ascertain how many parents with follow-up studies carried a structural rearrangement, either balanced or unbalanced, to determine appropriate FISH usage in the context of qPCR as a primary follow up methodology. The current study includes 25,696 CNVs in 21,473 families reported from 111,362 postnatal arrays performed from August 2008 to September 2016. Follow up studies were performed by FISH, chromosomes, or microarray, and in April 2015 qPCR was added as the primary follow-up methodology except for instances of clearly pathogenic CNVs, microscopically visible CNVs, complex abnormalities, or CNVs for which a commercial FISH probe was available. Of those families, 6,634 individuals (31%) either had follow-up studies or were presumed to have obligate parental carriers if CNVs were found in siblings. We identified 18 families (~1:370) in which CNVs resulted from balanced parental insertions, with slightly less than half being microscopically visible. There were five additional families (~1:1,300) in which a parent carried the same unbalanced insertion; consistent with a benign CNV. None of the insertions involved recurrent LCR-flanked CNVs. Microscopically visible CNVs are typically followed up with chromosome analysis, although the resolution of CNVs in the 5-10 Mb range may benefit from molecular techniques. Taken together, 23/6,634 probands (~1:280) with follow up studies have CNVs related to unbalanced structural rearrangements. Given the rarity of balanced insertion carriers, we propose that additional FISH follow up may be reserved for future prenatal studies in families with apparently de novo qPCR CNVs that are not LCR-flanked.

3044F

Cytogenetically detected inversions are rarely formed by ectopic recombination between inverted repeats. C. Carvalho¹, M. Pettersson², J. Eisfeldt³, J. Wincent^{2,6}, J.R. Lupski^{1,4,5}, E.S. Lundberg^{2,6}, D. Nilsson^{2,3,6}, A. Lindstrand^{2,6}.

1) Dept. of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Department of Molecular Medicine and Surgery, Center for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden; 3) Science for Life Laboratory, Karolinska Institutet Science Park, Solna, Sweden; 4) Human Genome Sequencing Center, BCM, Houston, TX; 5) Texas Children's Hospital, Houston, TX; 6) Department of Clinical Genetics, Karolinska University Hospital, Stockholm, Sweden.

The relevance of inversions for disease causation, speciation and adaptation, is broadly recognized. Cytogenetically detected pericentric and paracentric inversions are the hallmark of copy number neutral inversions (INVcnNs), which, in humans, are associated with congenital anomalies in ~9.6% of carriers. Yet, despite biological relevance, inversions are generally assumed to be copy number and phenotypically neutral events. The major mechanism of formation of INVcnNs is proposed to be non-allelic homologous recombination (NAHR) between inverted repeats in which large blocks of sequencing homology is estimated to contribute to the formation of 67% of INVcnNs, whereas shorter repeat sequences underlie formation of the remaining. We performed whole genome sequencing (WGS) with the Illumina 30X paired-end protocol of 15 large inversions detected by classical karyotyping in 14 carriers who were referred to the Karolinska University Hospital diagnostic laboratory with various clinical phenotypes. Data were analyzed using FindSV, a pipeline that combines CNVnator and TIDIT. Using this approach, we obtained the breakpoint junctions of 9 out of 15 INVcnNs. Contrary to expectations, data analyses revealed that the vast majority (8/9) of the resolved inversions were not mediated by inverted repeats, except for one *Alu-Alu* case. Instead, the majority of those inversions form simple breakpoint junctions, except in 1 case with a 4 kb deletion and a few kb duplication located on each side of an apparently reciprocal inversion. Importantly, in 4 out of 9 (44%) inversions, there was disruption of genes associated with diseases which may potentially contribute to that individual clinical phenotype. The gene disruption frequency is similar to the frequency obtained for balanced translocations (48%), some of which contribute to the carrier phenotype. In summary, our study indicates that: 1. high-coverage, short-read WGS can detect ~60% of the INVcnNs and resolve breakpoint junctions at the nucleotide level; 2. NAHR is not the major mechanism underlying formation of inversions as previously proposed; 3. Inversions are likely relevant contributors to Mendelian diseases. Inversions are considered balanced and clinically neutral genomic rearrangements but our study challenges this historic view, shows that gene disruption occur in almost half of the cases, and provides evidence that molecular mechanisms other than ectopic recombination are prominent contributors.

3045W

Whole-genome sequencing analysis of copy-number variation (CNV) using low-coverage and paired-end strategies is efficient and outperforms array-based CNV analysis. B. Zhou¹, S.S. Ho², X. Zhang³, R. Pattni¹, R.R. Haraksingh^{1,2}, A.E. Urban^{1,3}. 1) Department of Psychiatry and Behavioral Sciences, Department of Genetics, Stanford University School of Medicine, Stanford, CA; 2) Department of Life Sciences, The University of the West Indies, Saint Augustine, Trinidad and Tobago; 3) Program on Genetics of Brain Function, Stanford Center for Genomics and Personalized Medicine, Tasha and John Morgridge Faculty Scholar, Stanford Child Health Research Institute, Stanford University, Stanford, CA.

Background: CNV analysis is an integral component to the study of human genomes in both research and clinical settings. Array-based CNV analysis is the current first-tier approach in clinical cytogenetics. Decreasing costs in high-throughput sequencing and cloud computing have opened doors for the development of sequencing-based CNV analysis pipelines with fast turn-around times. We carry out a systematic and quantitative comparative analysis for several low-coverage whole-genome sequencing (WGS) strategies to detect CNV in the human genome. **Methods:** We compared the CNV detection capabilities of WGS strategies (short-insert, 3kb-, and 5kb-insert mate-pair) each at 1x, 3x, and 5x coverages relative to each other and to 17 currently used high-density oligonucleotide arrays. For benchmarking, we used a set of Gold Standard (GS) CNVs generated for the 1000-Genomes-Project CEU subject NA12878. **Results:** Overall, low-coverage WGS strategies detect drastically more GS CNVs compared to arrays and are accompanied with smaller percentages of CNV calls without validation. Furthermore, we show that WGS (at $\geq 1x$ coverage) is able to detect all seven GS deletion-CNVs >100 kb in NA12878 whereas only one is detected by most arrays. Lastly, we show that the much larger 15 Mbp Cri-du-chat deletion can be readily detected with short-insert paired-end WGS at even just 1x coverage. **Conclusions:** CNV analysis using low-coverage WGS is efficient and outperforms the array-based analysis that is currently used for clinical cytogenetics.

3046T

Scaling the resolution of sequence variant classification discrepancies in ClinVar. S.M. Harrison¹, J.S. Dolinsky², W. Chen³, A.J. Coffey⁴, C.D. Collins^{5,23}, S. Das⁶, J.L. Deignan⁷, K.B. Garber⁸, J. Garcia⁹, O. Jarinova⁹, I. Karbassi¹⁰, A.E. Knight Johnson⁶, J.W. Koskenvuo¹¹, H. Lee⁷, R. Mao¹², A. Mav-er¹³, R. Mar-Heyming¹⁴, A. McFaddin¹⁵, K. Moyer¹⁴, N. Nagan³, S. Rentas¹⁶, A.B. Santani^{16,17}, C. Saunders¹⁸, E.H. Seppala¹⁹, B. Shirts¹⁵, T. Tidwell¹², S. Topper²⁴, L.M. Vincent¹⁹, K. Vinette²⁰, H.L. Rehm^{1,21,22}. 1) Laboratory for Molecular Medicine, Harvard Medical School, Cambridge, MA; 2) Ambry Genetics, Aliso Viejo, CA; 3) Integrated Genetics, Laboratory Corporation of America® Holdings, Westborough, MA; 4) Illumina Inc., San Diego, CA; 5) EGL Genetics, Tucker, GA; 6) Department of Human Genetics, The University of Chicago, Chicago, IL; 7) Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA; 8) Invitae Corporation, San Francisco, CA; 9) Department of Genetics, Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada; 10) Quest Diagnostics, Athlona Diagnostics, Marlborough, MA; 11) Blueprint Genetics, Helsinki, Finland; 12) ARUP Laboratories, Salt Lake City, UT; 13) Clinical Institute of Medical Genetics, University Medical Centre Ljubljana, Ljubljana, Slovenia; 14) Counsyl, South San Francisco, CA; 15) Department of Laboratory Medicine, University of Washington, Seattle, WA; 16) Division of Genomic Diagnostics, The Children's Hospital of Philadelphia, Philadelphia, PA; 17) Department of Pathology and Laboratory Medicine, Perelman School of Medicine at The University of Pennsylvania, Philadelphia, PA; 18) Center for Pediatric Genomic Medicine, Children's Mercy Hospital, Kansas City, MO; 19) GeneDx, Gaithersburg, MD; 20) Molecular Diagnostics Laboratory, A. I. duPont Hospital for Children, Wilmington, DE; 21) Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA; 22) The Broad Institute of MIT and Harvard, Cambridge, MA; 23) Global Laboratory Services, PerkinElmer Genomics, Branford, CT; 24) Color Genomics, South San Francisco, CA.

Sharing data in ClinVar provides open access to variant classifications from many clinical laboratories. While most classifications are consistent across laboratories, classification differences exist. By aggregated submitted classifications, ClinVar provides an opportunity for submitters to collaborate on deriving the most accurate conclusions. A pilot project from ClinGen's Sequence Variant Inter-Laboratory Discrepancy Resolution team focusing on four clinical labs found that 53% of classification differences were resolved by either updating ClinVar with current internal classifications or reassessment of an older classification with current classification criteria (PMID: 28301460). With these findings in mind, our working group prioritized variants that reach a majority consensus ($\geq 2/3$ of submitters agree) with an outlier classification and expanded the scope to include more clinical labs. The first round compared classifications from 41 clinical labs and identified 24,445 variants classified by ≥ 2 clinical labs (April 2017). The majority of classifications were concordant (84.6%; 20,677 variants) and only 2.7% (650 variants) of variants were medically significant differences (MSDs) with potential to impact medical management. Labs with outlier classifications were sent a custom report and encouraged to reassess. Labs have returned results for 204 variants, of which 62.3% (127 variants) were resolved. A second round of this expanded effort including classifications from all clinical labs submitting to ClinVar (108 submitters; April 2018) is underway and preliminary data shows 54.3% (150/276) resolution rate. In total, 67.7% (488/721) of reassessed variants have been resolved through these rounds of discrepancy resolution and 22 clinical labs are now actively involved in this process. By comparison, within one year, 24.3% (106/437) of variants that previously had MSDs but did not reach a majority consensus are now concordant, showing that a subset of differences will be resolved over time by routine reassessment; however, notification of outlier classifications can serve as a prompt for reassessment and accelerate resolution. In conclusion, this process capitalizes on the value of data sharing within ClinVar and will help the community move toward more consistent variant classifications.

3047F

Analytical validation of qPCR detection of CYP2D6 copy number variation. J.M. Lubieniecka^{1,2,3}, K.P. Lubieniecki^{1,2,3}, A.L. Mohror¹, M.M. Soe¹, J. Reiner¹, D. Lyalin^{1,3}, K. Rageth¹, M. Mboob¹, L. Mullineaux¹, C.F. Boerkoel¹, R. Pyatt^{1,2,3}, L. Carmichael¹, P. Cherukuri^{1,2,3}, 1) Imagenetics, Sanford Health, Sioux Falls, SD; 2) Sanford School of Medicine, University of South Dakota, Sioux Falls, SD; 3) Sanford Research Center, Sioux Falls, SD.

Background: CYP2D6 copy number variation (CNV) is a commonly assessed pharmacogenetic variant. Reported variation includes 0, 1, 2, 3 and 4 copies. Frequently this variation is determined by qPCR; however, qPCR has difficulty resolving less than 2-fold differences in DNA copy number, although the resolution can be improved by increasing the number of sample replicates in a qPCR assay. CYP2D6 CNV falls short of the 2 fold difference at higher CNVs raising questions as to whether the commonly recommended 4 replicates per sample is sufficient. **Hypothesis:** More than four sample replicates are needed to detect higher CYP2D6 CNV robustly. **Methods:** Using a TaqMan probe for exon 9 of CYP2D6 and a QuantStudio 12K Flex Real-Time PCR system, we generated qPCR results for 12 replicates of each member of a set of Coriell samples and calculated relative copy number (RCN) as described by Weaver, et al. [1]. To determine the number of replicates needed for confident copy number calls, we used two different approaches: (a) random sub-sampling of 12 replicate data, and (b) median data point approach to minimize the effect of outliers. We also compare varying approaches to reliably determine CYP2D6 CNV. **Results:** This analysis finds that, in order to detect 3 copies of CYP2D6 gene within a 99% CI by qPCR, a minimum of 10 replicates are required. To ensure high accuracy alternative methods of copy number quantification may be necessary. **Conclusions:** The variability of qPCR data confounds detection of higher CYP2D6 CNV indicating that more than four replicates are needed for analytical validity. 1. Weaver, S., et al., *Taking qPCR to a higher level: Analysis of CNV reveals the power of high throughput qPCR to enhance quantitative resolution*. *Methods*, 2010. **50**(4): p. 271-6.

3048W

A retrospective analysis of clinical whole genome sequencing results after non-diagnostic microarray or whole exome sequencing. A. Scocchia, M. Rajan, M. Bennett, E. Thorpe, C.M. Brown, D.L. Perry, R.J. Taft. Illumina Inc., San Diego, CA.

Patients with a rare and undiagnosed genetic disease (RUGD) often undergo years of serial testing, commonly referred to as the "diagnostic odyssey". Often, clinicians must navigate the patient's health insurance policy and choose whether to investigate copy number variation (CNVs) using microarray (CMA) or target single nucleotide variants (SNVs)/indels using whole exome sequencing (WES). The use of clinical whole genome sequencing (cWGS) may provide simultaneous interrogation of the genome for both gene dosage and small base-pair changes, combining detection of SNVs, indels, CNVs, aneuploidy, and other chromosomal changes in a single clinical test. Here, we provide a case series of probands who underwent cWGS testing through the Illumina Clinical Services Laboratory from 2015-2018, in whom cWGS identified molecular findings with clinical congruence to the reported phenotype following prior non-diagnostic CMA or WES. As part of routine cWGS RUGD case analysis for 262 probands, details about prior genetic testing were collected from test requisition forms and medical notes submitted by ordering clinicians to the laboratory. At least 64% (n=167) of these probands were reported to have pursued at least one genetic test prior to cWGS, including CMA (n=104) and WES (n=17). Over half of these probands (57%; n=96) received two or more genetic tests, including nine probands who received both CMA and WES. After cWGS analysis, new information was reported for 62.5% (n=65) of probands who had received prior CMA and 53% (n=9) who received prior WES. Of these cases where new information was provided, most probands received pathogenic or likely pathogenic findings (66%, n=49). Most often, a SNV or indel was identified by cWGS after non-diagnostic CMA (n=36). In five cases, CNVs >100 bp but <87kb were identified by cWGS that were not reported to have been detected on CMA. In two cases, CNVs were detected after non-diagnostic WES. In five cases, new gene-disease relationships had emerged in the literature since last testing. Importantly, a compound heterozygous SNV/CNV pair was detected by cWGS in one case, which may not have been detected until a multi-platform approach was pursued. This case series highlights the added value of simultaneous detection of multiple variant types by cWGS over the serial or reflexive target-capture molecular testing approach often employed to investigate genetically heterogeneous differential diagnoses.

3049T

Implementation of copy number variant detection from existing exome and genome samples. *E. Farrow*^{1,2,3}, *I. Thiffault*^{1,2,3}, *M. Cadieux-Dion*¹, *G. Twist*¹, *B. Yoo*¹, *N. Miller*¹, *T. Pastinen*^{1,3}, *C. Saunders*^{1,2,3}.

1) Center for Pediatric Genomic Medicine Children's Mercy - Kansas City, Kansas City, Missouri; 2) School of Medicine, University of Missouri-Kansas City, Kansas City, Missouri; 3) Department of Pathology, Children's Mercy - Kansas City, Kansas City, Missouri.

Copy number variant (CNV) detection from next generation sequencing data has been a long standing promise of the technology. Indeed, numerous publications have demonstrated the ability to detect CNVs, ranging from small indels to aneuploidy. However, many of the studies utilize whole genome sequencing (WGS), which remains limited in clinical testing, or small panels. Additionally, there is no consensus on a standard bioinformatics approach. In order to be fully implemented into clinical testing, CNV detection ideally would be applied to both WES and WGS, and would not require modifications to established laboratory workflows. To test advances in CNV detection, with the goal of clinical implementation, we evaluated NxClinical (BioDiscovery), Dragen (Edico), and GATK (Broad Institute) with a validation set of 16 samples with clinically confirmed CNV; including those with existing NGS data and a subset prepared with Sciclone automation for library and enrichment. An initial assessment of these platforms performed surprisingly well: NxClinical detected 13/16, Dragen 12/16, and GATK 13/16 known CNVs, 9 of which were consistent between all three platforms. No CNV was missed by all, indicating that a combined approach may have some benefit. Deletions ranged from 20bp to 14Mbp, and duplications from 195bp to 1.7Mbp. Both Dragen and GATK are amenable to bioinformatic automation, but require custom analysis software for filtering of CNVs. In contrast, NxClinical has a graphical user interface, but is less amenable to automation. To further test CNV detection, an analyst blinded to the results was provided phenotypic information for each sample and recapitulated the diagnosis using NxClinical in 9/12 samples. Evaluation of an additional validation set resulted in a similar 'diagnostic' rate. Small exonic CNV in WES remained challenging for accurate detection, dependent on the amount of neighboring sequence. We therefore analyzed WGS for the WES samples with missed CNV calls. As expected, WGS offered increased resolution of CNV breakpoints; however 2 small exonic CNVs in *STK11* and *PLOD1* remained undetected. Additional testing to validate assay sensitivity and specificity is underway. In sum, CNV detection using existing laboratory workflows and commercially available software can provide accurate detection from both WES and WGS. What is becoming increasingly clear is that integration of CNV detection during clinical NGS testing is an alternative to exome-array. .

3050F

Supporting DNA variant interpretation: The global variome shared LOVDatabases. *J. T. den Dunnen*^{1,2,3}, *J.A.L.H. Lopez Hernandez*¹, *I.F.A.C. Fokkema*¹.

1) Human Genetics, Leiden University Medical Center, Leiden, Netherlands; 2) Clinical Genetics, Leiden University Medical Center, Leiden, Netherlands; 3) Global Variome, Newcastle upon Tyne, United Kingdom.

The simplest and cheapest answer to the possible consequences of a variant found in a genome is history. Has the variant been seen before and, when yes, where and what were the associated consequences in relation to the health of the individual? The Global Variome shared LOVDatabases build on a >20 year history of collecting and sharing such material. The databases contain a wealth of information on Individuals (patients), Phenotypes (disease/traits), Screenings and the Variants found, which can be queried in many ways. While the databases do contain information from published literature, most information has not yet been published but was submitted directly (functionality incl. automated data submission). When performing DNA diagnostics it is essential to check LOVD, preventing easily available information to be missed. Facilities supported by LOVD include variant queries (through the website or the API) across all public database installations, containing information from >3.6 million unique variants linked to >450,000 individuals. For genes with active curators (incl. BRCA, colon cancer, CFTR, etc.) specific "Summary records" resume all supporting information towards the conclusion whether the variant is or is not associated with a disease (pathogenicity). "Classification records" show the evaluation (classification) regarding variant pathogenicity as shared by diagnostic labs. Records labelled "In vitro (cloned)" show the results of assays performed to analyse the functional consequences of variants. Where available LOVD shows the parental origin of a variant linked to haplotype information. The RNA field, not available in most other databases, shows whether the consequences of a variant have been analysed on RNA level and what these consequences were. This adds essential evidence to the predicted consequences at the protein level, critical information for the diagnosis in diseases like Duchenne/Becker muscular dystrophy. Based on some examples we will show how the LOVD functionality can be used best to retrieve variant or phenotype information. The Global Variome shared LOVD is a community driven initiative operating under the auspices of Global Variome, a UK charity. Per gene, gene variant home pages link to other resources and data views in the major genome browsers. The data is shared with several public repositories incl. UCSC, EBI and ClinVar, and copies can be download from the gene home pages. Urls: databases.LOVD.nl/shared, www.LOVD.nl/3.0/search.

3051W

New approaches in the cystic fibrosis diagnosis: use of high-throughput sequencing technologies and *in silico* tools to identify and predict pathogenic variants. S.V.N. Pereira¹, F.A.L. Marson^{1,2,3}, J.D. Ribeiro^{2,3}, C.S. Bertuzzo¹.

1) Department of Medical Genetics, School of Medical Sciences, University of Campinas, Campinas, São Paulo, Brazil; 2) Department of Pediatrics, School of Medical Sciences, University of Campinas, Campinas, São Paulo, Brazil; 3) Laboratory of Pulmonary Physiology, Center for Pediatrics Investigation, School of Medical Sciences, University of Campinas, Campinas, São Paulo, Brazil.

Introduction: Cystic fibrosis (CF [MIM 219700]) is caused by *CFTR* mutations (~400 pathogenic). Allelic heterogeneity challenges the molecular diagnosis and precision medicine approaches in CF. **Purpose:** To identify *CFTR* variants by high-throughput sequencing (HTS) and predict the pathogenicity of the new variants by *in silico* tools. **Methods:** 169 CF patients had genomic DNA submitted to a Targeted Gene Sequencing with custom panel. Results were obtained from Basespace® and the alignment from TruSeq® Amplicon by Smith-Waterman algorithm. The variants annotation was made in VariantStudio® and confirmed in the Integrative Genomics Viewer. American College of Medical Genetics and Genomics and the Association for Molecular Pathology recommendations (2015) were applied to deduce the pathogenicity using six *in silico* tools: MutationTaster, MutPred-2, MutPred-LOF, MutPred-Splice, PolyPhen-2, Human Splice Finder. **Results:** A total of 62 variants were identified in 169 CF patients (3 patients had 3 variants). The most frequent alleles were: c.1521_1523delCTT, p.Phe508del (n=192; 56.30%), c.1624G>T p.Gly542X (n=26; 7.62%), c.3909C>G, p.Asn1303Lys (n=11; 3.23%), c.3484C>T, p.Arg1162X and c.1000C>T, p.Arg334Trp (both n=9; 2.64%). Among identified mutations: 41 were pathogenic, according to the literature [classified as: (I) n=23 (56.09%); (II) n=6 (14.63%); (III) n=1 (2.43%); (IV) n=6 (14.63%); (IV and V) n=1 (2.43%); (VI) n=4 (9.75%)]; 14 had uncertain significance (9 being pathogenicity in all predictors and 5 being discordant); and 7 new mutations were identified, evaluated and, based on the theoretical type of change + prediction analysis, we suggested the classification of c.(580+1_581-1)_(2615+1_2616-1)dup, c.1936G>T, p.Gly646X and c.3557_3557delA, p.Pro1186Leufs as class I. There was concordance of the predictors as "likely pathogenic" for c.*1233T>A, c.2804T>A, p.Leu935Gln, and c.4281C>T, p.Ile1427=. Also, c.974A>T, p.Tyr325Phe presented 1 discordant result among the predictors. Compared with the CFF data, among the mutations described, 7 are not in the registry and, among the potentially pathogenic ones, 4 were described. **Conclusion:** HTS plays a major progress in CF molecular diagnosis and was effective to detect rare and new variants. The use of *in silico* tools are an important step to classify the pathogenicity. HTS and *in silico* analysis can identify *CFTR* mutations and to give the opportunity to include the precision medicine into daily practice in a near future.

3052T

Advances in targeted NGS: Towards robust ultralow mutation detection in small and large DNA panels. E. Lader, R. Samara, Z. Wu, J. Ning, Y. Wang, J. Dicarlo, Q. Peng, D. Kim, C. Xu, Y. Wang. QIAGEN, Frederick, MD.

The quest for higher sensitivity and lower false positive rates are constant motivators in clinical NGS development. While computational optimization can be made to improve one or the other of these objectives, in general improving one metric requires compromise in the other. One tactic currently in use is to optimize detection for a narrowly focused panel, sometimes only a dozen or so clinically important sites, at the expense of (and essentially ignoring) any false positives elsewhere in the read set. We have performed a systematic evaluation of the errors caused by library construction and sequencing methodology with the goal of identifying and minimizing physical and computational errors in our targeted NGS libraries. The combination of these advances allow us to reliably call mutations at levels far below the typical 1% limit, not just for a small select group of sites, but over much larger regions, such as the complete CDS of clinically relevant genes. Critically, false positives are kept at exceptionally low levels through a combination of the enzymology used in library construction, the use of advanced molecular indices, single primer extension (SPE) technology, and the application of a computational model for enzyme induced errors.

3053F

The accuracy of computer-based diagnostic tools for the identification of concurrent genetic disorders. *D.Y. Park¹, R.R. Wadhwa¹, M.R. Natowicz².*

1) Cleveland Clinic Lerner College of Medicine, Case Western Reserve University, Cleveland, OH; 2) Pathology and Laboratory Medicine, and Genomic Medicine Institutes, Cleveland Clinic, Cleveland, OH.

The increasing use of next-generation sequencing, especially clinical whole exome sequencing, has revealed that individuals having two co-existing genetic conditions are not uncommon occurrences. This pilot study evaluates the efficacy of two methodologically distinct, publicly available computational differential diagnosis generating tools – FindZebra and SimulConsult - in identifying multiple genetic conditions in a single patient. Clinical query terms were generated for each of 15 monogenic disorders that were effective in resulting in the top ten list of differential diagnoses for each of the 15 monogenic conditions when entered into these bioinformatics tools. Then, the terms of over 125 pairings of these conditions were entered using each tool and the resulting list of diagnoses evaluated to determine how often both diagnoses of a pair were represented in that list. Neither tool was successful in identifying both members of a pair of conditions in greater than 40% of test cases. SimulConsult, when used in its 'incidence' mode, outperformed FindZebra and SimulConsult in its non-incidence mode in detecting pairs of monogenic conditions. Disorder detection sensitivity was not homogeneous within a tool, with each tool favoring the identification of a subset of genetic conditions. In view of recent whole exome sequencing data showing an unexpectedly high prevalence of co-existent monogenic conditions, the results from this pilot study highlight a need for the development of computational tools designed to effectively generate differential diagnoses with consideration of the possibility of co-existing conditions.

3054W

Experience with rapid exome sequencing of critically ill children: One year milestone. *K.M. Abbott, B. Sikkema-Raddatz, A.H. van der Hout, Y.J. Vos, H. Lemmink, J.B.G.M. Verheij, I.M. van Langen, M.J. de Groot, H.H.J. Drok, R. Kinds, G. Beunders, E. Zonneveld-Huijssoon, J.C. Herkert, F. Vansenne, J.S. Klein Wassink-Ruiter, E. Gerkes, K.J. van der Velde, M.A. Swertz, R.J. Sinke, W.S. Kerstjens-Frederikse, C.C. van Diemen.* Genetics, University Medical Center Groningen, Groningen, Groningen, Netherlands.

Introduction After successfully exploring whole-genome sequencing in critically ill newborns, we implemented rapid exome sequencing for similarly ill children (under the age of 16 years) in routine diagnostics. The advantages of switching to exome sequencing are: better integration into current diagnostic workflow, ability to simultaneously incorporate sequencing data of parents to help interpret and confirm results. Here we present the outcome after one year of implementation regarding diagnostic yield, incidental findings and turnaround time. **Materials and Methods** Forty-seven critically ill or deceased patients under the age of 16 years were included between April 2017 and April 2018. Trio (child-mother-father) exome sequencing was performed using Agilent SureSelect XT exome v6 capturing. After filtering using a virtual gene-panel of 3850 monogenic disease genes and inheritance-mode, we further selected variants based on population frequency, inheritance patterns and prioritization using in-house developed tools (GAVIN). The final variants were discussed in a multidisciplinary team. Pathogenic or likely pathogenic variants matching the patients' phenotype were reported. Variants in genes not matching the patient's phenotype but considered as actionable, were reported as incidental findings. **Results** Turnaround time was less than 14 days in most samples. Pathogenic or likely pathogenic mutations explaining the phenotypes were detected in 10 of the 47 children (21%) and involved the *CHD7*, *CRLF1*, *KCNQ2*, *KCNT1*, *MAP3K7*, *MTM1*, *PDHA1*, *PTPN11*, *SMARCA4* and *TBCK* genes. Incidental findings with clinical consequences were found and communicated in seven families. **Conclusions** Rapid exome sequencing is feasible and has a high yield (21%), allowing more critically ill children to be diagnosed. Trio-based analysis facilitates variant interpretation and confirmation, saving time which is crucial for decision management of both parents and health professionals.

3055T

Comparative quality of whole exome sequencing among commercial laboratories for patients in the Undiagnosed Diseases Network. D.R. Murdock, H. Smith, S. Chen, B. Dawson, L. Burrage, J. Rosenfeld, B. Lee, *Undiagnosed Diseases Network Members*. Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

The Undiagnosed Diseases Network (UDN) uses diagnostic tools including whole exome sequencing (WES) to solve complex medical cases. Many patients have now already undergone clinical WES through commercial laboratories prior to enrollment in the UDN. We aim to compare quality metrics among three major commercial labs by utilizing raw sequencing data available for these UDN patients. For patients accepted to the Baylor College of Medicine UDN clinical site, we obtained and reprocessed exome binary alignment map (BAM) files from three major commercial labs. A consensus exome design was used based on RefSeq gene transcripts since the proprietary exome designs from each lab were not available. We generated coverage and uniformity statistics for the exome, Online Mendelian Inheritance in Man (OMIM) disease genes, and American College of Medical Genetics Secondary Findings (SF) genes. We also analyzed the first coding exons, intron-exon boundaries, and coverage progressively deeper into introns. As no major commercial lab offers whole genome sequencing (WGS), we also analyzed data from the UDN WGS sequencing core. Importantly, we addressed the quality of the sequencing only, not clinical analysis pipeline or interpretation for clinical reporting. A total of 66 exomes sequenced between 2013 and 2017 were analyzed (Lab A: 39, Lab B: 9, Lab C: 18) along with 20 genomes. There was a small but significant difference ($p < 0.05$) in average coverage for the exome between labs (mean 122X, range 61-204X). This difference was greater ($p < 0.0001$) for OMIM disease and ACMG SF genes, indicating an enrichment for these clinically-relevant genes in some exome designs. Uniformity of coverage differed; some commercial labs provided better homogeneity of sequencing overall or in specific genes or regions (e.g., first exons). There was also a trend for better overall performance in more recent years. WGS provided lower average coverage (~50X) as expected but superior uniformity across all genes and regions, which may result in fewer false negative calls compared to WES at minimum 20X coverage. Clinical WES is an important tool in the diagnosis of patients with complex genetic cases. Our analysis suggests there are notable differences in the sequencing quality provided by different commercial labs, though this analysis does not address the potential impact on clinical utility. At the same time, WGS should be increasingly considered in the context of gene discovery and investigation.

3056F

A diagnostic odyssey ended via detection of CNV from whole-exome sequencing data. M. Wang¹, Y. Zou², F. Geng¹, J. Bai¹, N. Li¹, J. Guo¹, S. Shams³, Y. Jiang⁴. 1) Be Creative Lab, Beijing, China; 2) Clinical Cytogenetic Laboratory & Clinical Molecular Laboratory, University of Maryland, Baltimore, MD, USA; 3) BioDiscovery, Inc., El Segundo, CA, USA; 4) Peking Union Medical College Hospital, Beijing, China.

Detection of copy number variation (CNV) plays an important role in the diagnosis of patients with unexplained developmental delay/intellectual disability (DD/ID), autism spectrum disorders (ASD), and multiple congenital anomalies (MCA). To date, chromosomal microarray analysis (CMA) has been viewed as the "gold standard" for CNV detection and is recommended by ACMG as first line test for many constitutional abnormalities. However, the ability to detect CNVs using microarrays is limited by number of probes and non-uniform probe distributions. With the development of whole-exome sequencing (WES), such limitations are being overcome. We used WES to reliably call rare exonic CNVs using the MultiScale Reference (MSR) algorithm embedded in N.Clinical software. By using validation by molecular methods, we confirmed that small CNVs can be reliably called from whole-exome data. This report presents a 6-yr-old girl with intellectual disability, delayed language development, hypotonia, seizures, and autism spectrum disorders. The patient DNA was first analyzed using the Thermo Fisher CytoScan™ 750K SNP microarray which resulted in no unusual findings. The patient and her family were then screened using WES for any small structural variants (e.g. SNVs, In/Dels, etc.). Again, no obvious variant was detected that could explain the phenotype. The WES data was then processed through the MSR algorithm to detect CNVs of arbitrary size. The approach uses a dynamic binning algorithm that can detect small variants in areas of high coverage and larger variants in areas of lower coverage. This means small CNVs overlapping exons should be detected with good confidence. Using this method, we observed a rare deletion in the 12.288Kb range (Xq28(153,283,851-153,296,138)x1) overlapping a small portion of *MECP2*, all of *MIR718* and most of *IRAK1* genes. The left precise fracture point of this deletion is located in the 4th exon of *MECP2*, causing C-terminal domain (CTD) inactivation. The NxClinical software identified this as a de novo variant and the discordant mappings of the pair-end reads allowed us to visually detect the exact base pair location of the deletion. We used qPCR methods to confirm this small deletion. *MECP2*-related disorders in females include classic Rett syndrome, variant Rett syndrome, and mild learning disabilities which match the patient phenotype closely.

3057W

Up-to-date literature mining increases the clinical utility of whole exome sequencing. W. He, J.S. Salvo, F. Vetrini, R. Xiao, P. Liu, X. Wang, L. Meng, Y. Yang, F. Xia. Baylor Genetics Laboratory, Houston, TX.

Genetic changes underlie the majority of rare human diseases, especially early onset and severe Mendelian inherited developmental disorders. However, diagnosis of such disorders pose a big challenge for clinicians, largely due to the inherent genetic and phenotypical heterogeneity. The utilization of whole exome sequencing (WES) or whole genome sequence (WGS) has significantly expanded our ability to detect disease-causing mutations by testing all 22,000+ protein-coding genes. Using these techniques, many genes previously unknown to be the root cause of inherited developmental disorders have been discovered and published in the last decade. However, in clinical lab settings, the validity of genotype-phenotype correlation relies heavily on information from databases such as the Online Mendelian Inheritance in Man (OMIM), an online catalog of human genes and genetic disorders and Human Gene Mutation Database (HGMD). Yet information in these databases takes time to curate and doesn't include the most recently published literature. To keep up with newly published disease genes, we have developed a strategy of combining WES data with a list of recently discovered gene phenotype associations during case analysis so that any newly published disease genes will not be missed (So far we have accrued 1280 papers covering 1080 new disease genes in the list). Furthermore, we are able to reanalyze previous undiagnosed cases using these new gene disease associations, providing improved diagnoses years after initial testing. As we continue to accumulate more WES cases, access to the latest available gene disease associations has provided valuable patient insight, and will allow us to improve our diagnostic rate even further.

3058T

Double or multiple hits of CNV could be etiological factors for genetic disorders. K. Wittig^{1,2}, X. Wang², H. Wang², Y.M. Kim², L. Su^{2,3}, S. Lu^{2,4}, Z. Wen^{2,5}, W. Li², D. Taylor², X. Wang², S. Li². 1) Department of Pathology, University of Oklahoma Health Sciences Center, Oklahoma City, OK; 2) Department of Pediatrics, University of Oklahoma Health Sciences Center, Oklahoma City, OK; 3) Department of Pediatrics, Yanbian Medical University, Jilin, China; 4) Internal Medicine, Lishui Medical College, Zhejiang, China; 5) Internal Medicine, First Hospital of Jilin University, Jilin, China.

Much is still unknown of copy number variants (CNVs) detected in clinical diagnostic settings with each CNV being unique in size, genomic content, and location in the genome. CNVs are present throughout everyone's genome although many are not causal of genetic disorders with the majority of these being familial. Currently, some pediatric patients with non-syndromic genetic disorders see multiple specialists and undergo multiple different tests, including array comparative genomic hybridization (aCGH) and whole exome sequencing, with less than 40% receiving a causal diagnosis though there are also other factors to consider such as epigenetics, imprinting, and the environment. Previously, Girirajan et al. suggested a two-hit model regarding phenotypic heterogeneity – a second hit compounds with the first to produce a more severe phenotype. Often a variant found in a child also found in a parental sample is evidence for calling it benign. Here, we propose double or multiple CNV hits could be etiological factors for genetic disorders and a compound of two or more parental variants could in part be causative in the child by acting together in physiological networks to alter normal function. To prove this hypothesis, we conducted a pilot study. Our in-house database was mined for aCGH samples from 2015-2018 in two groups: pre-natal amniocentesis (n=29) and older children ages 7-13 years old (n=42). Pre-natal samples are referred to our lab due to advanced maternal age and suspected Down syndrome. A negative finding results in a normal report. Samples with a pathogenic finding or no clinical background provided were excluded. Testing was performed on Agilent 2x400K arrays V.1.0 and analyzed with Agilent CytoGenomics software. All CNV calls were made from the raw data with no size restrictions of three consistent probes within genomic content with a log ratio greater than ± 0.45 and confirmed by two reviewers. From our two groups, we observe older children having a significantly higher amount total CNVs ($p=0.01$) than the pre-natal samples with no significant sex or age bias. This data provides support that there may be complex genetic interactions responsible for the phenotypes presented in these children with an increased number of genetic events. Altered cellular characteristics may include multi-systemic signaling pathways, flank sequence integrity, transcription factories, chromosome territories, and mitotic pairing.

3059F

How to measure loss of chromosome Y (LOY). *M. Danielsson^{1,2}, J. Halvardsson^{1,2}, J. Mattisson^{1,2}, B. Torabi Moghadam^{1,2}, H. Davies^{1,2}, J. Dumanski^{1,2}, L. Forsberg^{1,2,3}.* 1) Immunology, Genetics and Pathology, IGP, Uppsala University, Uppsala, Sweden; 2) Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 3) Beijer Laboratory of Genome Research.

Recent discoveries show that mosaic loss of chromosome Y (LOY) in leukocytes is associated with increased risks of cancer, Alzheimer's and cardiovascular disease. More than 15% of men older than 70 show some degree of LOY and these men survive on average only half as long as men without LOY. But how is LOY measured today and what would be the best method for the future? Genome wide SNP-Arrays are most commonly used in LOY-studies, but other methods have also been used: whole genome sequencing (WGS), gene targeted qPCR or droplet digital PCR (ddPCR). We have compared, evaluated and made the most of SNP-Arrays, WGS and ddPCR targeting a homologous gene located on the X and Y chromosome for LOY-studies. The cheap and easy ddPCR method gives very similar results as the expensive and precise WGS, when using whole blood for detecting LOY. However, the sensitivity of all these methods are limited, leaving low frequencies of LOY undetected. These methods require high amounts of DNA input from hundreds of thousands of cells, but LOY is a binary event on the single cell level. The clonal expansion of rare LOY cells could have important clinical implications. Therefore, we aim to combine LOY and cell type detection in a clinically applicable single cell assay. The challenge of identifying LOY-cells could be approached by targeting LOY-specific upregulated surface proteins. Amplifying low signals with rolling circle amplification increases the number of targets to choose from and high throughput detection improves the sensitivity of the method. Using this method in the clinic would benefit both individual patients as well as the health care system and the society at large.

3060W

Application of ACMG criteria to classify variants in the human gene mutation database. *H.Q. Qu¹, X. Wang¹, L. Tian¹, Q. Li^{2,3}, H. Hakonarson^{4,5}.*

1) The Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA; 2) Faculty of Medicine, Memorial University of Newfoundland(MUN), Canada; 3) Princess Margaret Cancer Centre Research Institute, University Health Network(UHN), University of Toronto, Canada; 4) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA; 5) Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, 19104, USA.

Precision medicine based on personal genomics information has become the direction of future medicine. To apply genetic information in medicine, a critical but complicated step is to interpret the functional effects of the genetic variants. The American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) developed a set of standards and guidelines for this topic. The InterVar software enables automated interpretation of variants according to these guidelines, and in this study, it allows for a quantitative assessment of contradictory evidence in variant classification. The Human Gene Mutation Database (HGMD) professional dataset 2014 release was assessed using InterVar. This earlier version of HGMD dataset before the publication of the ACMG guidelines was used as a relatively independent database from the impact of ACMG. Among 139,541 HGMD variants, 42.93% were classified as pathogenic or likely pathogenic, and 6.07% were classified as benign or likely benign. Contradictory evidence was identified in both groups of variations. Evidence for benign classification, BP6 (0.69%, reputable source recently reports variant as benign), BS2 (0.36%, observed in a healthy adult individual), and BP1 (0.23%, missense variant in a gene for which primarily truncating variants are known to cause disease), was identified among variants classified as pathogenic. For likely pathogenic variations, BP1 (3.45%) and BS2 (0.35%) were identified. For the most commonly observed contradictory BP1, this type of pathogenic variants are commonly reported in the literature. Although initial interest in the analysis of some Mendelian diseases focused on finding truncated variants, missense variants are more commonly found to be disease causing. Pathogenic evidence is also observed among benign variations. PM1 (well-established functional domain) is commonly observed among variants classified as benign (24.65%), while PM2 (absent from controls) and PM1 are commonly identified among variants classified as likely benign (49.55% and 42.98%, respectively). Therefore, PM2 and PM1 should not be used as an effective evidence for pathogenicity unless supported by additional data. Taken together, these observations may not only help researchers and clinicians to better understand the coexistence of contradictory evidence in variant classification, but suggests there may be need for potential revisions and improvements of the current ACMG guidelines.

3061T

How to be atypically consistent: Clinical classification of atypical canonical splice variants. J.D. Murdoch, Y. Xue. Fulgent Genetics, Atlanta, GA.

A significant proportion (1-2%, PMID:11058137) of canonical ± 1 -2 splice sites are not in the prevailing GT donor-AG acceptor classes. Mutations at GC donor sites are reported in ichthyosis, spastic paraplegia, etc. (PMID:28403545, 24704789). ACMG guidelines (PMID:25741868) state that canonical ± 1 -2 splice variants are very strong evidence (PVS1) when null variants are a known mechanism of pathogenicity. However, they do not cover atypical canonical splice site variants. Current predictive tools for atypical canonical splice site variants also struggle with non-GT-AG splice variants - in Alamut, only 1 of 5 algorithms gives non-GT-AG predictions. Given short conserved sequences around sites, a genome-wide search for possible atypical splice sites is prohibitive. These shortcomings are a notable gap in current interpretation of potentially actionable splicing variants; e.g. the same clinical lab has called two ± 1 variants in the *SCN8A* gene VUS and P - despite the P variant being atypical and the VUS being typical (and predicted to abolish all splicing), even while stating for the VUS that loss of function is not a known disease mechanism (at odds with HGMD and ClinVar for LoF *SCN8A* variants, some from that lab). Other labs just call all atypical splice variants VUS. To better report these variants while consistently applying proper criteria, we propose utilizing all available tools: clinical literature, allele frequency databases, reviewed mutation databases, and previously reported internal data. Lacking predictive models, the interpretive framework is necessarily more restrictive than for typical canonical splice variants. We suggest a prudent approach of treating previously unreported atypical canonical site variants as VUS, given no further information or inconsistent allele frequencies, and previously reported variants in cases as LP, given no inconsistent allele frequencies or conflicting literature. Notably, one family, the voltage-gated sodium channel A genes, may constitute a possible special subclass. Half (15/30) of all atypical splice sites encountered in database/literature searches occurred in this family, and of 12 in ClinVar, 11 were reported P/LP, 8 based on literature. Of 83 ± 1 -2 splice site SNV entries in ClinVar for this family, 77 were P/LP and only 1 was B/LB (even as a typical site - again in *SCN8A*). We propose that ± 1 -2 disruptions to splice sites in this family, typical and atypical, may be treated as LP even without prior reporting.

3062F

Capture of sample identification SNPs in NEBNext Direct target enrichment panels. A.B. Emerman¹, K.M. Patel¹, S.K. Bowman¹, S.M. Adams¹, B.S. Desmond¹, J.S. Dunn¹, A. Barry², T. Davis², S.E. Corbett¹, C.D. Elfe¹, E. Mauceli¹, C.L. Hendrickson¹. 1) Directed Genomics, Ipswich, MA; 2) New England Biolabs, Ipswich, MA.

Target-enrichment for next-generation sequencing is a powerful approach that focuses sequencing data on selected genomic regions of interest and is widely used for genetic screening and clinical diagnostics. When used in the clinic, robust tracking of DNA samples is critical to maintain sample and data integrity and to prevent false reporting of results to a patient. A reliable method to track patient samples from sample preparation through data analysis and reporting is to monitor single nucleotide polymorphisms (SNPs) in the patient's DNA. Work from others used whole exome sequencing data from diverse populations to identify 24 SNPs that have a high discriminatory power among individuals. Here we demonstrate that these 24 SNPs can be targeted with hybridization-based target enrichment panels that can be spiked into a primary panel or can be used as a stand-alone assay. Baits targeting the 24 sample identification SNPs were designed for use with the NEBNext Direct target enrichment protocol. After an initial screening of the baits to assess performance, the concentrations of the baits were adjusted to result in uniform coverage of the SNPs, with a final normalized coverage of the targets between 0.4-1.7. The resulting panel had greater than 95% of the reads mapping to the targeted regions. The ability of the panel to distinguish sample identity was validated on 16 HapMap samples from Coriell Institute. Additionally, by titrating the SNP panel into a larger, cancer-focused panel, we could control the reads dedicated to the hereditary SNPs such that we obtained sufficient coverage of the SNPs for sample identification while maintaining the deep coverage of the primary targets that is necessary for low-level variant detection. This work shows that baits targeting sample identification SNPs can be incorporated into NEBNext Direct target enrichment panels or may be used as a secondary screen to confirm sample identity and maintain data integrity. As the NEBNext Direct technology converts samples into sequencer-ready libraries within one day, this method offers a fast and convenient approach to track patient identity in high-throughput sequencing workflows.

541T

Acceptability of incorporating genetics in risk prediction of cognitive impairment in hematologic cancer survivors treated with blood or marrow transplantation (BMT). *N. Sharafeldin, E. Matsuda, P. Singh, K. Battles, K. Triebel, V.G. Wadley, S. Bhatia.* University of Alabama at Birmingham, Birmingham, AL.

Background: Cognitive impairment is prevalent post-BMT with inter-individual variability in risk suggesting a role for genetic susceptibility. Gene-based personalized interventions carry the potential to induce changes in outcome compared to conventional interventions, however, clear evidence is lacking on patient acceptability of communication of their individual genetic risk and whether that would lead to enhanced intervention acceptability. **Methods:** This is a preliminary report from the Cognitive Training and Attitudes towards Genetics (cTAG) study on a survey of BMT survivors assessing (i) knowledge and understanding of genetics using a 16-item survey (Jallinoja and Aro, 1999), (ii) attitudes towards receiving personal genetic information, and (iii) acceptability of incorporating genetics in risk prediction to inform cognitive interventions in a hypothetical scenario. **Results:** To date 50 BMT survivors have completed the survey: mean±SD age was 50±14.4y; 80% self-identified as white and 20% as black; 62% were males; and 44% had a college degree or higher. The median total score on the 16-item knowledge survey was 62% (IQR: 50-75%); higher education was associated with increased knowledge of genetics ($p=0.04$). Participants tended to score higher on questions pertaining to inherited causes of disease (median score=62.5%) compared to factual questions on genes, chromosomes, and cells (median score=56%). The higher score on inheritance and disease related genetic knowledge was associated with higher education and white race ($p=0.005$ and $p=0.05$, respectively). Fifty percent considered receiving tests of genetic testing before BMT somewhat or extremely important, 26% were neutral, and 24% considered it not important. Participants presented with a hypothetical scenario of a high risk of cognitive problems based on genetic testing were more likely to consider cognitive interventions (61.5%) as compared with those presented with a scenario of a low risk of cognitive problems based on genetic testing (38%). Overall, 86% of participants were confident they understood the questions, and 80% considered it not hard to use genetic information to make intervention decisions. **Conclusion:** These preliminary findings show potential utility of incorporating genetics in predicting risk of cognitive impairment among BMT recipients. Findings from this survey will inform the development of targeted interventions for those identified to be at high risk of cognitive impairment.

542F

Rates of HPV infection in patients of health institutions in the district of David, province of Chiriqui, Panama. *O.I. Batista^{1,2}.* 1) Human Genetics, Centro Gendiagnostik, S.A., David, Chiriqui, Panama; 2) Centro Especializado de Genética/UNACHI, David, Chiriqui, Panama.

Infection with the human papillomavirus (HPV) is a public health problem worldwide. These viruses are a heterogeneous group, which have been associated with the etiology of a number of benign, precancerous and cancerous lesions on the skin, mucosa, and respiratory tract. Although several factors may contribute to the development of cancer of the uterus, epidemiological studies have suggested that the definitive event responsible for uterine cancer is the infection of the cervix by some types of HPV. The objective of this study was to determine the prevalence of HPV infection in patients attending health institutions in the district of David, Province of Chiriqui, Panama. A total of 102 samples were collected by obstetricians, gynecologists, or urologists using swabs. Patient samples consisted of 45 cervico-vaginal samples from women who visited public healthcare institutions. The second group consisted of 43 cervico-vaginal samples and 14 urethral mucosa samples from women and men, respectively, collected in private healthcare facilities. Both sexes showed characteristic signs of sexually transmitted diseases, including genital warts. DNA was extracted from the swabs and was amplified using PCR with the degenerate oligonucleotide primers MY09 and MY011, primers that target the L1 region of the HPV genome. The prevalence of HPV infection in the entire population sample was 66%. 65% of women tested positive for HPV whereas 71% of men tested positive. The prevalence of HPV infection among patients at public or private healthcare institutions was 84% and 51%, respectively. An additional screening test for HPV 16 and 18, the main causative agents of cervical cancer and performed on most of the positive patients in group 2, showed that one patient presented infection by type 16. These results indicate the high prevalence of HPV in the region and the need to offer and improve genetic diagnostics, possibly by genotyping and / or DNA sequencing, using capillary electrophoresis, which are very sensitive technologies and with greater discriminatory power. Besides a methodology based in hybridization will be also performed for comparative analysis. The application of multiple methodologies would allow the unambiguous typing of the viruses present in our environment, a better monitoring of the risk of cancer of the uterus and a more targeted attention to both sexes, frequently afflicted by this infectious disease.

543T

How Canadian general practitioners perceive their roles and information needs in a context of genetic susceptibility to breast and ovarian cancer in a rural area. *M. Bezeau¹, R. Drouin^{2,3}, C. Bouffard¹.* 1) Université de Sherbrooke, Sherbrooke, Québec, Canada; 2) Université Laval, Québec, Québec, Canada; 3) Université du Québec à Montréal, Montréal, Québec, Canada.

Many genetic tests make it possible to identify people who are at higher risk of developing breast or ovarian cancer and to estimate their risks of presenting one. They allow doctors to propose preventive approaches that can significantly reduce the mortality rate among people at risk. However, some measures, such as prophylactic mastectomy, are not universally accepted because of their major psychological and physical repercussions. Despite differences of opinion, there is a consensus on the importance of informing patients following the identification of a mutation in a gene associated with breast and ovarian cancer so that they can make informed decisions about their health. It is even more important that general practitioners (GPs) provide adequate and equitable medical care regardless of their practice setting. Yet, the literature indicates many gaps regarding: a) organization and delivery of services; b) levels of knowledge and training of health professionals; c) the type of information they need regarding genetic testing and clinical management in case of increased risk. Finally, the literature shows that the management of patients with a *BRCA1/BRCA2* mutation is far from optimal and may even make access to preventive interventions inequitable. Research interest: 1) What are the perceptions of Canadian GPs regarding the role they believe they should play in the follow-up of patients with a *BRCA1 / BRCA2* mutation? 2) What are their needs for support, resources and training on genetic testing and patient management? Methodology: An exploratory qualitative approach has generated in-depth data on a little-explored topic among the people directly involved. For the first phase of this project, we conducted interviews with 13 GPs in Bas-Saint-Laurent (BSL), a rural region of Canada with some of the highest rates of *BRCA1/BRCA2*-related breast cancer. Results: Participants feel that they do not have enough genetic training to adequately monitor their patients. GPs who wish to become more involved have different training and support needs from those who prefer to leave this follow-up to specialists. Some GPs report that genetics is a hard-to-reach specialty for patients living in remote areas of major centers. Conclusion: The needs of GPs vary depending on the role they want to play. Training and support strategies must evolve with these needs, which are also changing with the arrival of new tests and genes panels.

544F

The other genes: A qualitative look into experiences of patients with a genetic risk for breast cancer in genes other than *BRCA 1/2*. *K. Clift¹, S. Macklin², S. Hines³.* 1) Center for Individualized Medicine, Mayo Clinic, Jacksonville, FL., USA; 2) Department of Clinical Genomics, Mayo Clinic, Jacksonville, FL., USA; 3) Center for Breast Health, Mayo Clinic, Jacksonville, FL., USA.

Since the 1990s, genetic testing has been offered to individuals at high risk of carrying a pathogenic variant in the *BRCA1/2* genes. Strong evidence has been gathered on how pathogenic variants in these two genes affect health and thorough, well-researched management guidelines exist. However, pathogenic changes in *BRCA1/2* do not account for all cases of hereditary breast cancer, and additional genes have since been associated with increased cancer risk. Many of these more newly discovered genes are now included on hereditary cancer syndrome gene panels. Exact cancer risks are not available for many of these genes, and many of these genes appear to convey a moderately increased risk for breast cancer compared with the high risk associated with *BRCA1/2*. Also, while association with increased breast cancer risk may be apparent, often, it is not entirely clear which other organs may have increased risk. This ambiguity can make it difficult to decide whether interventions, like increased surveillance, prophylactic surgery and/or pharmacologic risk reductions, should be initiated and when. We are conducting qualitative interviews to explore the personal and healthcare experiences of men and women with pathogenic variants in these "other" genes in order to compare their experiences to those with *BRCA1/2* variants. So far we have found that information available to patients is lacking and patients often find themselves educating their physicians about the meaning of their results. Some participants have stated that they feel they have been "brushed off" by their healthcare professionals because these genes are even less well understood outside of the genetics community. Despite recent rapid growth in the genetic industry, public knowledge and physician knowledge remain lacking. Future research will be required to fully understand the significance and proper management of these more recently discovered hereditary breast cancer genes. In the meantime, to improve patient care, further evaluation of patient and provider education needs should be at the forefront of the conversation. While this study focuses on genes related to hereditary cancer, this topic will be important across numerous indications.

545T**Increased breast cancer screening and prevention in *ATM*, *CHEK2*, and *PALB2* mutation carriers identified by hereditary cancer panel testing.**

J.O. Culver¹, C. Hong¹, C.N. Ricker¹, D. Sturgeon¹, P. Levonian², R. Koff², C. Rowe-Teeter³, K. Lara-Otero³, K.E. Kingham², N.M. Chun², A.P. Lebensohn², K. Lowstuter¹, K. Partynski¹, M.A. Mills², I. Petrovich², A. Hartman², B. Allen², K. Brown³, K.J. McDonnell¹, U. Ladabaum², J.M. Ford², S.B. Gruber¹, A.W. Kurian², G.E. Idos¹. 1) USC Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA; 2) Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA; 3) Myriad Genetics, Salt Lake City, UT.

Since the widespread use of hereditary cancer panels, three breast cancer susceptibility genes, *ATM*, *CHEK2*, and *PALB2*, have emerged as clinically useful for risk assessment and screening. The National Comprehensive Cancer Network recommends annual breast MRI screening for female *PALB2* mutation carriers over age 30 and female *ATM* and *CHEK2* mutation carriers over age 40. The aim of this analysis was to determine whether identifying mutations in these genes has influenced clinical decisions among mutation carriers or testing decisions among their relatives. Two thousand patients had a 25- or 28-gene panel test as part of a prospective study. Patients completed a baseline questionnaire to determine pre-test screening practices and were sent 3-, 6-, and 12-month follow-up questionnaires to determine post-test screening behaviors and whether relatives had testing. Chart reviews were used to determine screening behavior for non-respondents. Study participants had a median age of 51 years and were 81% female, 73% with cancer history, 35% non-native English speaking, and 30% with a high school education or less. The distribution of race/ethnicity was 40% White (non-Hispanic), 39% Hispanic, 12% Asian, 4% Black, 5% Other. Among the 2000 tested, 41 (2.1%) carried a mutation in *ATM* (N=15), *CHEK2* (n=17), or *PALB2* (n=9). Responses to at least one follow-up survey were received for 34/41 patients (83%). Among the 21 females eligible for breast MRI screening after testing, 3/21 (14%) had planned annual breast MRI prior to genetic testing. By twelve months after receiving genetic test results, 15/21 (71%) indicated that genetic evaluation influenced breast screening or surgery decisions; 15/21 (71%) had undergone breast MRI and 6/21 (29%) had undergone risk-reducing mastectomy (RRM). A total of 17/21 (81%) had undergone an MRI and/or RRM. Five of six participants who underwent RRM had a unilateral diagnosis of breast cancer or DCIS. Among the 31 respondents to family genetic testing questions, 26/31 (84%) had encouraged family members to undergo testing, and 14/31 (45%) reported that a family member had undergone testing. Findings indicate that for most female mutation carriers eligible for screening and risk reduction, the identification of *ATM*, *CHEK2*, and *PALB2* mutations led to an increase in preventive behaviors within 12 months. Future analysis in this and other populations will help elucidate whether patients continue to comply with screening recommendations.

546F**Ovarian and breast cancer genetic risk assessment in a community education program setting. K. Dhillon, H. Malakoff, K. Driscoll. Rivkin Center for Ovarian Cancer, Seattle, WA.**

Individuals with mutations in *BRCA1*, *BRCA2* and other DNA repair genes are at higher risk for ovarian, breast and other cancers, and can take advantage of prophylactic surgery and other preventative interventions. Increased awareness and access to genetic testing has led to increased numbers of individuals who do genetic testing for cancer susceptibility genes. This increase has predominately been in individuals without family or personal history of either cancer, while testing is underused by those who may be at high risk and be carriers of cancer-associated mutations. Community health education is an impactful approach to reach high risk individuals. Here we report the progress and impact of the first year of the Rivkin Center's CanCan education program—a joint ovarian and breast cancer education program with a focus on the genetic and family history risks for both cancers. Program goals are to help identify high risk individuals and to educate diverse populations on risk factors, signs, symptoms, prevention options, and screening guidelines for both cancers. A trained instructor and an ovarian or breast cancer survivor teach the one-hour free, interactive workshops. Attendees receive a two-step risk assessment worksheet to (1) document their family cancer history, and (2) complete a risk assessment checklist to determine if they should seek genetic counseling. Workshops take place in diverse community settings, including low income housing developments, religious orgs, corporate offices, and hospitals in Washington State, and in the campus setting for younger women at all PAC-12 universities. Focus on outreach to the Ashkenazi Jewish, minority, immigrant, and LGBTQ communities ensures that the program reaches high risk and underserved communities. Since June 2017, CanCan has conducted 190 workshops, with over 10,000 individuals educated. Post-workshop surveys returned by 563 attendees show that 95% will talk to their family about health history, 99% will seek a healthcare provider for a medical concern, 91% had an excellent or good understanding of the signs, symptoms, and screening guidelines for both cancers, and 96% will share what they learned with others. In summary, the first year of the CanCan ovarian and breast education program has been very successful, with attendees reporting increased awareness and intent to assess their risk. Future goals of the program are to increase program reach in the community and through digital education.

547T

A pipeline for identification and confirmation of potentially actionable germline mutations in tumor only genomic sequencing. *D. Farengo Clark¹, J. Powers¹, K.L. Nathanson², D.B. Lieberman³, J.J.D. Morrisette³, K.N. Maxwell⁴, S.M. Domchek⁴.* 1) Department of Hematology-Oncology, University of Pennsylvania, Philadelphia, PA; 2) Division of Translational Medicine and Human Genetics, Department of Medicine, Perelman School of Medicine at the University of Pennsylvania; 3) Department of Pathology and Laboratory Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia; 4) Division of Hematology-Oncology, Department of Medicine, Perelman School of Medicine at the University of Pennsylvania.

Background: A published concern of next-generation sequencing for tumor DNA profiling (tNGS) is the identification of potential germline findings. For both the patient and their family, it is critical to develop referral and testing pathways to address potentially actionable germline findings. We describe a novel algorithm for referral of patients for germline testing after tNGS and results of initial implementation. **Methods:** A multidisciplinary team was assembled to determine: (1) criteria to analyze somatic variants for germline actionability; (2) pathogenic variants (PV) triggering "automatic referrals" for germline testing; and (3) a protocol for clinical genetics referral. Data analyzed included gene-specific germline: somatic mutation ratio, genotype-phenotype associations and clinical utility. **Results:** 1695 patients underwent tNGS from 2016-2018 with 111/1695 identified by automatic referral criteria. The criteria to refer germline confirmation included: (1) all PV in *BRCA1/2*, *BRIP1*, *PALB2*, *RAD51C/D*, and mismatch repair genes (75%); (2) PV in patients with classic phenotype for associated cancer susceptibility syndrome (e.g. *TP53* in personal/family suggesting Li Fraumeni syndrome) (23%) and (3) known germline founder PV (2%). The oncologist was asked to discuss findings with the patient prior to the genetic counselor (GC) making contact. The GC facilitated germline testing and reporting (if patient agreeable). Forty-eight (43%) patients underwent germline testing and 28/48 (58%) had confirmed germline PV. Forty-six percent of confirmed germline PV were in genes associated with the tumor type. All *BRCA1/2* PV were referred and 72% (13/18) were confirmed as germline. Eight patients with germline *BRCA1/2* PV had classic HBOC associated cancers [pancreatic (6), prostate (2), breast (1)], while four were identified in tumors not classically associated with HBOC (glioblastoma (2), gallbladder and colon). No germline testing was performed in 63/111 (56%) of patients due to lack of referral (39%), poor patient health or death (24%), and patient refusal (30%). **Conclusions:** 6.5% of patients had somatic findings triggering referral. Of these, 48 (43%) underwent genetic testing, 28 were positive, of which, 10/28 (36%) may not have been referred by personal or family history of cancer alone. Continued research is critically needed to address the challenges of implementation of this algorithm and capture of potentially actionable germline findings found by tNGS.

548F

Effects of nicotinamide on primary cervical cancer cells (fibroblasts): A new therapeutic modality. *R.N. Hassan, H.L. Luo.* Genetics, Sun Yat-Sen University, Guangzhou, Guangdong, China.

Background: Cervical cancer is a spectrum from pre-malignant condition to aggressive disease. Unfortunately, drug resistance is inevitable in this cancer. Alternative therapeutic strategies are urgently needed to improve outcomes and to reduce the cost of treatment. Nicotinamide is a known amide of niacin that plays major role in critical cellular processes. Whether nicotinamide can be used as an anti-cancer agent is still unclear. **Methods:** Fibroblasts derived from primary cervical cancer patients were treated with nicotinamide and the cell viability were assessed using the CCK-8 assay. We used flow cytometry to measure the reactive oxygen species (ROS) and glutathione peroxidation. Cellular morphology were studied post-treatment by DAPI staining. Annexin V assay was conducted to assess apoptosis. We conducted western blot to demonstrate caspase cleavage. Finally, quantitative reverse transcriptase polymerase chain reaction was used to measure redox gene expression in the treated cervical cancer cells. All experiments were performed in triplicates and a p-value of <0.05 were considered significant. **Results:** Nicotinamide was able to inhibit the cellular proliferation of cancer cell fibroblasts in a dose dependent and time-dependent manner. It decreased the GSH content in the cells, while increasing the ROS. Induction of apoptosis was observed in nicotinamide treated cells. **Significance:** We demonstrated that nicotinamide could induce apoptosis and gross inhibition of cervical cancer cells through increased oxidative stress. This new evidence suggests that nicotinamide is a promising anti-cancer agent. **Key words:** nicotinamide, cervical cancer, fibroblast, apoptosis.

549T

The influence of Malay cultural beliefs on breast cancer screening and genetic testing: A focus group study. *T. Shaw¹, D. Ishak¹, D. Lie², S. Me-non¹, E. Courtney¹, S-T. Li¹, J. Ngeow^{1,3,4,5}.* 1) Cancer Genetics Service, Division of Medical Oncology, National Cancer Centre Singapore, 169610; 2) Office of Clinical Sciences, Duke-NUS Medical School, Singapore 169856; 3) Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore 308232; 4) Oncology Academic Clinical Program, Duke-NUS Medical School, Singapore 169857; 5) Institute of Molecular and Cell Biology, Agency for Science Technology and Research (A*Star), Singapore 138673.

Objective: Malays comprise an Asian cultural group reported to have low breast cancer screening uptake rates and poor cancer outcomes. Little is known about Malay cultural factors influencing beliefs and practice of cancer screening and genetic testing. Our study aims to explore health beliefs of Malay women around breast cancer screening and genetic testing. **Methods:** We conducted focus groups among healthy English-speaking Malay women in Singapore, aged 40 to 69 years, using a structured guide developed through literature review, expertise input and participant refinement. Thematic analysis was conducted to extract dominant themes representing key motivators and barriers to screening and genetic testing. We used grounded theory to interpret results and derive a framework of understanding, with implications for improving uptake of services. **Results:** Five focus groups (four to six participants per group) comprising 27 women were conducted to theme saturation. Major themes were 1) spiritual and religious beliefs act as barriers towards uptake of screening and genetic testing, 2) preference for traditional medicine competes with Western medicine recommendations, 3) family and community influence health-related decisions, complexed by differences in intergenerational beliefs creating contrasting attitudes toward screening and prevention. **Conclusion:** Decisions to participate in breast cancer screening and genetic testing are mediated by cultural, traditional, spiritual/religious and intergenerational beliefs. Strategies to increase uptake should include acknowledgement and integration of these beliefs into counseling and education and collaboration with key influential Malay stakeholders and leaders.

550F

Cervical screening for cancer and other cytologies: Role of repair genes P53 and hMLH1 methylation status. *S. Thumojur¹, YR. Ahuja¹, M. Vasavi^{1,2}.* 1) Genetics and Molecular Medicine, Vasavi Medical and Research Centre, Hyderabad, India; 2) Genetics, Hansgene Cancer Foundation, Hyderabad, India.

Introduction: Cervical Cancer is preventable. The main obstacle for the success of screening efforts in women is the stigma about the screening procedure both in rural and urban population. Tissue changes, fueled by DNA damage and repair errors, alter the genome to produce instability that drives tumorigenesis. hMLH1 repair gene is known to be hypermethylated in a variety of tumors. A common polymorphism of P53 (Pro to Arg) at codon 72, affects the susceptibility of P53 to E6-mediated degradation in vivo. The association of this polymorphism with different cytologies and cervical cancer cases was studied. **Methods:** Creating awareness, training para-medical staff, involving local voluntary bodies for conducting Pap testing and HPV subtyping in health-camps gave us optimal leads to manage these patients. A total of 315 screened subjects and 148 histologically confirmed cervical cancers were recruited for the hMLH1 methylation study. A total of 482 subjects, out of which 121 normal pap cytology, 213 abnormal cytology and 148 cervix cancers were recruited for P53 genetic variant study by PCR analysis (rs1042522) P53 (exon4) primers. **Results:** A total of 530 women (general population) were screened; 1.88% showed high risk HPV subtypes. Initial poor compliance was improved drastically after we changed our approach with pre-camp, general awareness talks and one-to-one counselling of women on importance of testing to avoid future risk for cancer. hMLH1 promoter methylation was significantly higher in inflammation, infection, RCC, ASCUS, SCC cases whereas atrophy, LSIL and surprisingly HSIL did not show any significant change compared to controls. The P53 (Pro to Arg) residue at codon 72 significantly showed an association of the variant amino acid Arginine with carcinoma of uterine cervix in comparison to cases with normal cervical cytology. **Conclusion:** Conducting national screening programs along with counselling for awareness can reduce the burden of this cancer in developing countries. There is a need to screen women for high risk HPV sub-types for successful application of any future vaccinations programs. Promoter methylation of hMLH1 gene may be associated with increased risk of pathology in the cervical tissue; this could also be enhanced by Hr-HPV infection. Genotyping of P53 (Pro to Arg) at codon 2 in patients with abnormal pap cytology may be useful in identifying individuals at a high risk of worsening of cervical health and in developing cancer.

551T

Digital direct engagement is an effective method for communicating updated test results. S. Topper, C.L. Neben, C. Lai, C. Heydrich, A.Y. Zhou. Color Genomics, Burlingame, CA., Select a Country.

The duty to recontact patients with new information that may meaningfully alter medical care was first established in 1999. The issue has grown in urgency since then, and ACMG, ASHG, and ESHG have all issued updated statements re-examining and re-affirming this duty. Despite the ethical framework and the increasing urgency, little has been published about the methods or effectiveness of recontacting. Color Genomics, a clinical laboratory, utilizes electronic communications to maintain active interactions with clients, regardless of their physical location. Here we present data on the effectiveness of providing updated information to 417 clients who received genetic testing ordered by their personal healthcare provider (traditional) or ordered by an open-access, independent healthcare provider (independent). We first assessed whether clients viewed their genetic testing reports after they were made available. A total of 83.7% (170) in the traditional cohort (72.9% within 7 days) accessed their results compared to 96.7% (202) of the independent cohort (93.3% within 7 days). The decreased uptake in the traditional cohort could be attributed to clients directly discussing reports with their personal healthcare providers. At a subsequent date, all of these individuals (and their providers) were notified by email of a revised report based on new genetic or clinical information including variant reclassification, new health history, or updated risk. When looking at the access rates for revised results, a total of 45.4% (93) of the traditional cohort accessed their updated results (43.4% within 7 days) compared to 69.8% (148) in the independent cohort (67.5% within 7 days). We next examined whether the type of new information in the revised reports affected access rates. Clients with updated screening guidelines or recommendations (due to updates in professional society recommendations) accessed revised reports most frequently (87.5%, 14), followed by new health history (56.1%, 125) and variant reclassification (57.3%, 102). In conclusion, to our knowledge, this is the first report on the effectiveness of providing updated information in a clinical laboratory. Taken together, the data presented here demonstrate that digital direct communication mediated by a clinical laboratory is effective for communicating updated test results to clients.

552F

Genetic testing for breast cancer survivors. J. Wiggins¹, A. George^{1,2}, Z. Kemp^{1,2}. 1) Cancer Genetics Unit, Royal Marsden Hospital, Sutton, United Kingdom; 2) Institute of Cancer Research, 15 Cotswold Road, Sutton, United Kingdom.

Introduction Knowledge of germline *BRCA1/2* status increasingly influences the management of early breast cancer. However, many breast cancer survivors have not benefited from genetic testing because it may not have been offered, testing criteria may have subsequently broadened or family history may have emerged since initial treatment. Others may not have felt ready to consider testing at diagnosis. The identification of *BRCA1/2* mutations remains important during survivorship both in terms of management of future cancer risk and cascade testing amongst family members. It is anticipated that there will be 2 million breast cancer survivors in the UK by 2040. In the US ACS/ASCO breast cancer survivorship care guidelines (2015) recommend doctors revisit genetic assessment for breast cancer survivors. There are no such recommendations in the UK. Here we describe a collaboration between the genetic and breast cancer teams at the Royal Marsden Hospital to identify patients eligible for genetic review prior to discharge from the breast service. **Method** Early breast cancer patients are routinely reviewed at a virtual MDT (VMDT) 5 years post diagnosis to determine long term cancer surveillance and endocrine requirements. A Senior Genetic Counsellor joined the VMDT to highlight patients eligible for genetic referral. Genetics appointments were offered in writing. **Results** Between May 2015 and December 2016 935 patients were discussed at the VMDT. One third (309/935) were eligible for genetic assessment +/- testing based on their personal +/- reported family history. Less than half of eligible patients had been referred to genetics in the period of breast cancer follow up. Of the patients offered genetic assessment one third declined the appointment or did not attend. Of the 117 patients who attended the genetics clinic, 88 were offered genetic testing and 9 *BRCA* mutation carriers were identified (10.2% detection rate). When taking into account those who had been offered testing before, 3.4% (32/935) of the follow up population were identified as mutation carriers compared with 2.4% (23/935) prior to this intervention. **Conclusion** A significant number of *BRCA1/2* mutation carriers are not identified during their breast cancer treatment pathway. Breast cancer survivors and their families can benefit from routine genetic assessment in the same way issues such as endocrine treatment and surveillance are currently considered.

553T

Findings of a universal screening program for Lynch syndrome in an integrated healthcare system. J.E. Hunter, L. Jenkins, E. Shuster, J. Zepp, T. Kauffman, S. Gille, K.A.B. Goddard. Center for Health Research, Kaiser Permanente Northwest, Portland, OR.

Lynch syndrome (LS) is associated with an increased risk of cancer, most commonly colorectal (CRC) and endometrial (EC) cancers. A diagnosis of LS has a significant impact on clinical management of patients and their at-risk family members. Universal tumor screening (UTS) of all patients newly diagnosed with CRC and EC is recommended to increase identification of cases of LS and reduce morbidity and mortality compared to selective screening based on risk factors such as age and family history. Currently, there is limited data regarding outcomes of screening programs in community healthcare settings, including uptake of follow-up testing in patients who screen positive for LS. Kaiser Permanente Northwest (KPNW) is an integrated healthcare system that serves northwest Oregon and southwest Washington. KPNW implemented a UTS program for LS among cases of CRC in January 2016 and EC in November 2016. Screening includes immunohistochemistry (IHC) of the four mismatch repair proteins associated with LS (MLH1, MSH2, MSH6, and PMS2) followed by reflex to testing for the *BRAF* V600E variant (CRC only) and *MLH1* hypermethylation (CRC and EC), when appropriate, to rule out likely sporadic cases. As of March 2018, 278 cases of CRC and 70 cases of EC had been screened for LS. Overall, ~15% of patients had an abnormal IHC, consistent with the percentage reported in prior studies. Eleven patients (9 CRC, 2 EC) received an abnormal screen result in the absence of signs of sporadic cancer and were offered genetic counseling and genetic testing. Of these, 10 patients (8 CRC, 2 EC) underwent genetic testing. Six (5 CRC, 1 EC) patients were found to have a pathogenic or likely pathogenic variant in one of the LS genes: 1 *MLH1*, 1 *MSH2*, 1 *MSH6*, and 3 *PMS2*. Of those with a diagnosis of LS, age at onset of cancer ranged from 32 to 81 years. Two additional patients with CRC received results for a variant of uncertain significance (VUS): one patient with a VUS in *MLH1* and a second patient with 2 VUSs in *MLH1* and *MSH6*. Genetic testing was negative in three patients. The typical time following surgery for a patient to be contacted by the genetics department for follow-up was roughly 2 weeks. However, 91% of patients who received an abnormal screening result for LS followed up with genetic counseling and testing, indicating that the timing of receiving tumor testing results (i.e., soon after surgery for CRC) is not a significant barrier for follow-up.

554F

Implementing personalized cancer medicine into primary care: Challenges and solutions. J.C. Carroll^{1,2}, T. Makuwaza^{1,2}, D.P. Manca³, N. Sopcak³, M.A. O'Brien⁴, J.A. Permaul⁵, R. Heisey^{1,4}, R. Urquhart^{5,6}, C. Sawka⁷, S. Pruthi⁸, N. Schneider⁹, E. Grunfeld^{1,10} On behalf of CanIMPACT. 1) Department of Family and Community Medicine, University of Toronto, Toronto, Ontario, Canada; 2) Ray D Wolfe Department of Family Medicine, Sinai Health System, Toronto, Ontario, Canada; 3) Department of Family Medicine, University of Alberta, Edmonton, Alberta, Canada; 4) Department of Family and Community Medicine, Women's College Hospital, Toronto, Ontario, Canada; 5) Beatrice Hunter Cancer Research Institute, Halifax, Nova Scotia, Canada; 6) Department of Surgery, Dalhousie University, Halifax, Nova Scotia, Canada; 7) Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, Ontario, Canada; 8) General Internal Medicine, Mayo Clinic, Rochester, Minnesota, US; 9) Independent, St. Walburg, Saskatchewan, Canada; 10) Ontario Institute for Cancer Research, Toronto, Ontario, Canada.

Objective: To realize the benefits of personalized medicine (PM), it must be integrated into primary care (PC). Previous work identified increased knowledge, connection to genetics specialists and point-of-care (POC) resources as important for successful implementation. Our objective was to explore an implementation strategy using a PM cancer toolkit. **Design:** A mixed methods implementation study including focus groups (FG) and pre/post questionnaires. **Setting and Participants:** PC providers in Canadian team practices (Ontario 2, Alberta 1). **Intervention:** After incorporating PC feedback, we developed a toolkit including POC risk assessment and management tools for hereditary breast and colorectal cancers, introduced at FG1. Using the Consolidated Framework for Implementation Research, we explored implementation strategies with practices through FGs at 2 and 4 months, with evaluation at 6 months. **Methods:** We used semi-structured, framework-informed interview guides and descriptive thematic analysis of transcripts by site and across time. Pre/post questionnaires assessed confidence in PM competencies using Fisher's Exact tests. **Findings:** 19 PC providers participated. Implementation strategies included integrating POC tools into the Electronic Health Record (EHR), modifying workflow to collect family history (FH), and bookmarking the GECKO genetics education website (www.geneticseducation.ca). Participants described how having few patients with significant cancer FH made it challenging to keep PM top of mind and see the need for practice change. They recognized FH as key to PM and wanted electronic tools for collection and interpretation; additional PM POC EHR tools with routinized reminders; and seamless fit to clinic flow. Local champions were valued for advocating use of PM tools. At 6 months, there was a significant increase in PC providers' confidence regarding hereditary cancer (pre 12%, post 54%, $p=0.02$), deciding who to offer genetics referral (pre 29%, post 75% $p=0.025$), and knowing where to obtain PM information (pre 0%, post 50%, $p=0.005$). **Conclusions:** PM continues to be an implementation challenge in PC. Successful strategies must address both knowledge gaps and practice issues.

555T

Maternally inherited *SDHD*: Management challenges. *J. Vengoechea, C. Tallo, D. Davis.* Department of Human Genetics, Emory University, Atlanta, GA.

Paragangliomas and pheochromocytomas (PGL/PCC) are rare tumors that develop in neuroendocrine tissue, can secrete hormones leading to high morbidity and may be malignant. Approximately 40% of PGL/PCC cases are hereditary and about 30% of these are caused by a paternally inherited mutation in *SDHD*. Maternally inherited *SDHD*-related PGL/PCC has been reported (Yeap et al, 2009; Bayley et al, 2014; Burnichon et al, 2017), which is plausible due to two independent somatic events: 1) a mitotic recombination resulting in two recombined copies of chromosome 11 and 2) loss of the recombined chromosome containing a paternal 11q23 region and maternal 11p15 region. We describe a case of maternally inherited *SDHD*-related PGL/PCC and highlight the management and genetic counseling challenges for these families. The proband is a 53 year-old woman who was diagnosed with bilateral carotid body tumors and a right glomus vagale tumor, the right sided tumors were removed and confirmed to be paragangliomas. She is undergoing radiation therapy for the left sided tumor. Her family history was notable for a sister who died at 39 from metastatic PGL and a daughter diagnosed with a PCC at age 29 who had tested negative for mutations in *RET*. Clinical testing revealed a pathogenic variant, c.242C>T in *SDHD*. The affected daughter and unaffected 36 year old daughter then tested positive for the variant. There is no consensus on the screening method and frequency for identification of PGL/PCC in presymptomatic individuals with a predisposing germline mutation. We have developed a biochemical and imaging screening protocol, the latter can be costly and time consuming, and often reveals incidental findings that require extensive follow-up. The decision of initiating this protocol in a case of maternally derived *SDHD* should not be taken lightly. In this family, we were able to demonstrate that a maternally inherited *SDHD* variant was penetrant. There is yet to be a consensus about whether this can only be caused by a somatic translocation event. Until there is further clarification on the mechanisms behind penetrance of a maternally inherited *SDHD*-related PGL/PCC, we ordered baseline CT, MRI and MIBG tests. Genetic counseling about how this represented an exception to the general rule of parent-of-origin effect in *SDHD* proved to be challenging. A better understanding of the causes for maternal transmission of *SDHD* will help provide appropriate genetic counseling and management recommendations.

556F

Incidental identification of a moderate penetrant germline *RET* mutation in a patient with advanced epithelial ovarian cancer: A significant benefit of a multidisciplinary molecular tumor board. *T.A. DeMarco¹, R. Moore², K. Matthijssen¹, A. Schott¹, D. Alaeddin¹, T. Cannon³, L. Knopp³, N. Hauser¹, J. Deeken¹.* 1) Inova Translational Medicine Institute, Fairfax, VA; 2) Taussig Cancer Center, Cleveland, OH; 3) Inova Schar Cancer Center, Fairfax, VA.

The primary purpose of a molecular tumor board (MTB) is to convene a group of experts to review patients with refractory cancers and to make potential recommendations for targeted therapy based on their somatic tumor profiling. Another potential benefit of an MTB is the identification of somatic tumor results which may have potential germline implications for patients. We present the case of an Asian female diagnosed with advanced metastatic epithelial ovarian cancer at the age of 65. In July of 2016 the patient underwent a multi-gene pan-cancer genetic testing panel composed of 32 genes and no mutations were identified. In November of 2016 somatic tumor profiling was performed indicating several positive alterations, including the identification of the *RET* p.V804M mutation. A genetic counselor present at MTB noted the *RET* mutation as possibly being germline. Since the patient's previous panel testing did not include the *RET* gene, the counselor recommended that the patient receive additional genetic counseling with the option of germline genetic testing for this particular mutation. The patient underwent additional single-site germline genetic testing and her results were positive for the *RET* p.V804M mutation. Mutations in the *RET* gene are associated with MEN2A and an increased risk for the development of medullary thyroid cancer, hyperparathyroidism and also pheochromocytoma. The patient's personal and family history was non-contributory for any of these particular findings; and given that p.V804M is a moderate penetrant *RET* mutation, this is not uncommon. However, this particular *RET* mutation may also increase the risk for the development of papillary thyroid cancer. Interestingly, the patient's daughter had a history of papillary thyroid cancer; but, she tested negative for the familial *RET* mutation. The identification of the *RET* mutation provided information about further cancer risks for the patient and the opportunity for cascade testing in her relatives. Given the lack of suggestive personal and/or family history, had this patient not been presented at MTB attended by a genetic counselor, it is much less likely that the *RET* germline mutation would have been identified. This case illustrates the importance of a multidisciplinary MTB utilizing genetic counselors to identify and discuss somatic tumor findings that may have significant germline implications for patients and their relatives.

557T

BRCA1 and BRCA2 variant reclassifications from Mount Sinai Hospital, Toronto. J. Lerner-Ellis^{1,3}, C. Mighton^{1,2}, M. Wang^{1,3}, S. Shickh^{1,2}, N. Watkins^{1,3}, K.R. Zakoor², A. Wong³, Y. Bombard^{1,2}. 1) University of Toronto, Toronto, ON; 2) Li Ka Shing Knowledge Institute, St. Michael's Hospital, Toronto, ON; 3) Laboratory Medicine and Pathobiology, Sinai Health System, Toronto, ON.

Background: Accurate interpretations of genetic variants are essential for diagnosis, management and counselling of patients with inherited disorders. Periodic variant reassessment can ensure that classifications reflect current evidence and best practices. Even variants in well-characterized genes such as *BRCA1* and *BRCA2* are prone to inter-laboratory discordance and may be reclassified over time as new evidence is generated. The Advanced Molecular Diagnostics Laboratory (AMDL) provides *BRCA1/2* testing for patients who meet provincial eligibility criteria. Here we report on *BRCA1/2* variant reclassification between 2012 and 2017. **Methods:** Variants were assessed using a standardized variant assessment tool based on ACMG guidelines and tracked in an in-house database. Variants were classified to one of five levels, based on ACMG criteria: benign, likely benign, variant of uncertain significance (VUS), likely pathogenic, and pathogenic. Variants were shared through the Canadian Open Genetics Repository and submitted to ClinVar. **Results:** Between January 1, 2012 and August 18, 2017, 1213 *BRCA1/2* variants were reported. In total, 33.1% (402/1213) were reassessed and 12.5% (152/1213) reclassified. Of the 152 reclassified variants, 63.2% were reclassified to benign, 21.0% to likely benign, 9.9% to variant of unknown significance, 2.0% to likely pathogenic, and 3.9% to pathogenic. In 17.1% of cases, the variant was reclassified more than once, and 8.6% of variants were reclassified back to their initial classification through multiple reclassification events. **Conclusions:** Variants may be reclassified over time, highlighting the importance of periodic reassessment to help ensure that patients receive the most appropriate care based on their genetic test results. It is also important that laboratories engage in data sharing, to advance collective knowledge of variant pathogenicity. In some cases, variant reclassification could impact patients' management. For instance, reclassification of a *BRCA1/2* variant to pathogenic could inform surgeries, increased screening, and predictive testing for family members. Variant reclassification presents ethical and practical challenges related to patient recontact, and suggests a need for research and professional guidance on recontacting patients.

558F

BRCA in China: Two common founder mutations of BRCA1 in Chinese population. A. Li, Y. Zi, T. Zhou, Y. Wu, Y. Deng, J. Luo, Y. Shi. TopGene Tech Co., Ltd. Guangzhou, Guangdong Province, P.R.China.

Tumour suppressor genes, *BRCA1* and *BRCA2*, confer a high risk of breast cancer, ovarian cancer and associated hereditary breast and ovarian cancer syndrome (HBOC). Its genetic testing has thrived on next-generation sequencing (NGS) and strong indication for precision medical management. Several recurrent mutations were considered to result from founder effects in previous Chinese *BRCA* researches. To overcome geographical restriction and limited number of carriers (which have previously caused difficulties in verifying presumed founder mutations), we screened a large Chinese nationwide cohort of 5537 samples comprising 630 carriers, in which 196 women (31.1%) and 2 men (0.3%) were diagnosed with breast cancer, 327 (51.9%) with ovarian cancer, 63 (5%) with breast and ovarian cancer, and 36 (5.7%) without cancer. Thirty hot-spot mutations were identified with multiple occurrences (occurred in ≥ 4 unrelated patients). *BRCA1*: c.5470_5477delATTGGGCA was identified from 59 individuals out of 50 unrelated families, and *BRCA1*: c.4801A>T from 7 individuals out of 6 independent families. More importantly, our preliminary bioinformatic analysis indicated that the estimated age of c.5470_5477delATTGGGCA is 125 generations (with the help of haplotype reconstruction: case vs. control, p-value = 0.01). For unrelated carriers, their family members (without mutations) and normal Chinese control, further comparison of SNPs and STRs (within linkage disequilibrium region) was performed. The results supported that the occurrences of c.5470_5477delATTGGGCA and c.4801A>T have resulted from the same ancestry, whereas *BRCA1*: c.981_982delAT, *BRCA1*: c.3770_3771delAG and *BRCA2*: c.3109C>T is likely to be caused by independent mutational events. The identification of founder mutations would greatly reduce the cost for genetic screening of HBOC in China. Furthermore, to address the lack of public and comprehensive *BRCA1/2* database in China, we have built an online website for researchers to acquire Chinese spectrum of *BRCA1/2* mutations, linking to our developed computational tool for variant interpretation using artificial intelligent technology (submitted). We also observed enrichment at exon 10 of *BRCA1* (density=0.5, p<0.05) and exon 13 of *BRCA2* (density=0.03, p<0.05) in ovarian cancer. For breast cancer, concentrated regions were observed in exon 10 of *BRCA1* (density=0.3, p<0.05) and exon 23 of *BRCA2* (density=0.03, p<0.05).

559T

Identification of pharmacogenetic markers in cancer supportive care. Z. Gu, M. Wei. Admera Health, South Plainfield, NJ.

Background: Precision medicine has become increasingly critical in cancer supportive care, with implications for treatment selection, medication dosing, and side effect prediction. The presence of clinically actionable germline variants brought the hope that accurate prediction of drug efficacy and toxicity would be in the future. This will greatly improve the comprehensive oncology care, where supportive care treatment options can be subjective and personalized. **Methods:** The changes in drug effect and toxicity caused by genetic variations were characterized in 4287 patients submitted for clinical pharmacogenetics testing using a gene panel including key pharmacokinetic and pharmacodynamic genes. For each patient, the need to select an alternate medication and adjust medication dose was determined for the treatment of pain (*CYP2D6*, *CYP1A2*, *OPRM1*), nausea/vomiting (*CYP2D6*), antifungal prophylaxis (*CYP2C19*) and depression (*CYP2D6*, *CYP2C19*). It is recognized that there are many other factors playing in the variations of the response and toxicity to these classes of medications. **Results:** Of the 4287 patients, 964 (22%) would have a recommendation to choose alternatives of oxycodone, codeine, or tramadol therapy and 3394 patients (79%) for a non-standard dosage. A change in certain antiemetics such as ondansetron would be suggested for 171 patients (4%). If voriconazole antifungal prophylaxis was clinically required, 1228 (29%) would need a dose increase, while 133 (3%) should be switched to a different regimen. 2727 patients (64%) should avoid or adjust dosage on some of the commonly used agents for depression due to reduced efficacy or heightened risk of toxicity. **Conclusions:** Pharmacogenetics markers associated with reduced drug efficacy and higher rate of ADRs were commonly observed for medications used in cancer supportive care. Identification of these markers would help clinicians achieve optimal treatment outcome with manageable time and cost. This provides great opportunities to apply precision medicine principles to cancer patient management.

560F

Precision medicine in clinical cardiovascular care: Implementation of a multidisciplinary cardiogenetics clinic. L.D. Hellwig^{1,2,3}, C.E. Turner^{2,3}, A. Mains², A. Krokosky², J. May², D. Shanta², C. Dobson², M. Haigney^{2,3}. 1) Henry M Jackson Foundation for the Advancement of Military Medicine, Bethesda, MD; 2) Uniformed Services University of the Health Sciences, Bethesda, MD; 3) Walter Reed National Military Medical Center, Bethesda, MD.

Background: Cardiac diagnosis and follow-up for families can take great coordination across multiple specialties. The Walter Reed Cardiogenetics Clinic was designed as a multidisciplinary approach that centers on elevated quality of personalized and family-centered care. The Cardiogenetics Clinic includes providers from Adult Cardiology, Pediatric Cardiology, and Genetics, allowing optimization of proper diagnosis, medical management, and follow up care using combined expertise in one clinic visit, as opposed to separate visits for each specialty. We hypothesized that this multidisciplinary approach would increase efficiency of evaluation of Service Members and their families and unify medical management plans for the family. **Methods:** In a single Cardiogenetics visit, patients and their family members receive a thorough cardiac evaluation, physical examination, genetic counseling, and coordinated genetic testing (if indicated and patient-elected). The impact of this clinic was determined by calculating the total number of separate specialty visits the clinic saved since it began on 09 March 2017 and through qualitative assessments by providers. **Results:** Between 01 March 2017 and 1 Feb 2018, a total of 97 patients were evaluated in Cardiogenetics Clinic. Prior to the Cardiogenetics Clinic, these 53 family units would have had to make 188 separate trips spanning Pediatric Cardiology, Adult Cardiology, and Genetics. This indicates a savings of 72% of all visits by employing a multi-disciplinary approach and seeing the family units together in one visit. In addition, cross-disciplinary communication in patient care and clinical decision making has qualitatively improved the quality of care. **Conclusions:** The Cardiogenetics Clinic significantly reduced the number of separate visits each family needed to attend for their medical management. In addition to the time saved for families, this multidisciplinary approach allows for a variety of experts to provide elevated quality of care for these families. Intentions to further this multi-disciplinary approach to precision-medicine and family-centered care include future efforts to expand the current referral network, incorporate telemedicine initiatives to broaden access to this service, institute new research endeavors, and initiate extramural collaborations with other national leaders in the field.

561T

Clinical outcomes of the TTR V122I mutation in a brother-sister pair. *J. Liu, J.Y. Liu, A. Sara, J. Fan, J. Wang.* UCLA Ronald Reagan Medical Center, Los Angeles, CA, USA.

Approximately 4% of the African-American population possess a valine-to-isoleucine substitution within the transthyretin protein that results in a tendency for a normally tetrameric protein to dissociate into misfolded, monomeric subunits. These misfolded proteins can then accumulate pathologically and cause an autosomal dominant amyloid cardiomyopathy. Homozygous patients are infrequently documented in case reports, and though there are larger studies among heterozygous patients, there is a lack of studies or reports comparing disease within a family. Here, we report a 65-year-old African-American woman with a history of recurrent supraventricular tachycardia requiring radiofrequency ablation and recurrent chest pain and elevated troponin despite no evidence of coronary artery disease. Our patient was found to be homozygous for the TTR V122I mutation with evidence of infiltrative process on cardiac MRI, diagnosed shortly after her 61-year-old brother died of related cardiac amyloidosis while awaiting orthotopic heart transplantation. Here, we compare the clinical course and imaging data for the aforementioned brother-sister pair in the context of the amyloidogenic V122I transthyretin gene variant. Through this familial report, we aim to highlight variations in expression within a family and emphasize the importance of genetic testing of families at risk for this specific variant within the African-American community.

562F

Type 2 diabetes: More likely in eveningness chronotype than morningness chronotype. *Q. Naz^{1,2,3,4}, N. Verma², K. Usman⁴, A.A. Mahdi².* 1) Physiology, Biochemistry & Medicine, King George's Medical University, Lucknow, India; 2) Dept. of Physiology, King George's Medical University, Lucknow, India; 3) Dept. of Biochemistry, King George's Medical University, Lucknow, India; 4) Dept. of Medicine, King George's Medical University, Lucknow, India.

Background: Humans living in modern industrialized societies have easier access to light and are often forced to be awake by social norms against their preferred times independent of the master circadian rhythm. These environmental stimuli eventually lead to the disruption of the circadian system. Thus, both genetic variations in clock genes and environmental elements affect the distribution of chronotype. **Aim:** The aim of this study is to investigate the mRNA expression of circadian clock genes (related to metabolic homeostasis) RevErb alpha and ROR alpha in Type 2 diabetic T2DM patients with Morningness, Eveningness and mixed chronotype. **Methodology:** A total of 135 subjects (morningness n=55, mixed n=55 & eveningness n=25 age 18 to 16 years were recruited in Clinical OPD of General Medicine, KGMU. The subjects we enrolled were mainly treated their diabetes & life style related disease including diabetes & other metabolic disorders. 1 µg of total RNA was reverse transcribed into cDNA using Mol Bio™ reverse transcriptase. RT-PCR analysis was performed by the AB Applied Biosystems in 96 well plates using the SYBR GREEN. Primers were designed with the help of NCBI. GAPDH was used as an internal control. The relative quantification of mRNA expression was calculated as a ratio of target gene to GAPDH. **Results:** When we compared these 3 groups, Significant Different was found in various parameters like FBG (P = 0.01) Postprandial (P = 0.03) HbA1c (P = 0.001) TG (P = 0.001), Total Cholesterol (P = 0.01) & VLDL (P = 0.005). Serum leptin (15.01 ± 14.70) also reduced in eveningness when compared to morningness type, which may increase appetite & reduce energy expenditure leading in turn to the development of obesity & deterioration of metabolism. It also shows the complete inversion of the cortisol level (171.55± 77.70). ROR α expression was reduced in eveningness (p= 0.071) and mixed chronotype (p= 0.025) when compared with morningness chronotype (p= 0.091). Also comparison of REV ERB α expression was reduced in eveningness (p= 0.039) & mixed chronotype (0.047) when compared with morningness type or early eaters. **Conclusion:** Clock gene (REV ERB α & ROR α) mRNA expression is significantly altered in morningness and mixed chronotype as compared to morningness type in T2DM Patients. RT-PCR determination of this specific clock gene may be helpful for earlier detection of T2DM in patients with altered feeding habit & unbalanced lifestyle.

563T

Identification of pathogenic variants associated with primary immunodeficiency (PID) in Jordanian patients using next generation sequencing (NGS). M.M. Abu-Halaweh¹, R. Alzyoud², M. Abu-Shukair³, KH. Al-Qaoud³, J.R. Lemke⁴, R. Abou Jamra⁴, H. Tawamie⁴. 1) Biotechnology and Genetics Engineering, Philadelphia university, Amman, Amman, Jordan; 2) Queen Rania Al Abdullah Hospital for Children, King Abdallah St. - Medical City, Amman, Jordan; 3) Department of Biological Sciences, Yarmouk University-Irbid 21163-Jordan; 4) Institute of Human Genetics, University Medical Center Leipzig, 04103 Leipzig, Germany.

Primary immunodeficiencies (PIDs) represent a heterogeneous group of monogenic disorders including more than 300 genetically defined diseases that influence the development and functions of innate and adaptive immunity. The extreme heterogeneous phenotypes of PIDs can make the diagnosis workup challenging and time consuming. Thus, accurate identification of causative variants and immunological pathways is crucial as it provides the opportunity for timely treatment to prevent life-threatening infections and irreversible organ damage. Whole exome sequencing (WES) has revolutionized the diagnosis of genetic diseases by providing an increasingly cost-efficient diagnostic tool. In this study we applied whole exome sequencing (WES) in patients from two families of Jordan descent. Homozygous candidate variants that may be relevant for the patients' phenotype in *IL21R* (c.563del, p.Leu188Argfs*43), *CD40LG* (c.418T>C, p.Trp140Arg) have been identified. The validation and segregation of the identified variants were performed by Sanger Sequencing. Our findings indicated that WES is a useful approach for the identification of pathogenic variants associated with PIDs. More studies are still necessary to confirm the genotype-phenotype correlation at the functional level.

564F

Anti-cyclic citrullinated peptide (CCP) in rheumatoid arthritis. Does it have any role in early onset of the disease? A. Niraula¹, S. Sapkota², K. Dahal³, S. Kunwar⁴, NP. Shah⁵. 1) Senior Resident, Biochemistry, B.P. Koirala Institute of Health Sciences, Dharan, 1, Nepal; 2) Medical Biochemist, Lumbini Zonal Hospital, Butwal, Nepal; 3) DM Resident (Nephrology), National Academy of Medical Sciences (NAMS), Bir Hospital, Kathmandu, Nepal; 4) Lecturer, Department of Biochemistry, Modern Technical College, Sanepa, Lalitpur; 5) Medical Technologist, Nepal Public Health Laboratory, Teku, Kathmandu, Nepal.

Background: Rheumatoid arthritis (RA) is the most frequent systemic autoimmune disease affecting approximately 1% of the human population. Rheumatoid arthritis is manifested with the chronic inflammation of small joints, becoming gradually even more destructive in the course of time. The cause for the disease is attributed to various environmental, microbial and genetic factors. The definite diagnosis is based on clinical features and serological tests like Rheumatoid factor (RF), erythrocyte sedimentation rate (ESR) and C-reactive protein. Although, no any marker has been in practice to assess the course of the disease from beginning. Markers like Anti-CCP has been shown to assess the disease at an early period and have shown to have high sensitivity and specificity. **Material and Methods:** This is a hospital based comparative cross-sectional study conducted in patients with RA and healthy controls. Patient was recruited from study Star Hospital Private Limited, Lalitpur, Nepal. Fifty patients with Rheumatoid Arthritis who met at least four criteria and 25 healthy patients who served as control were included in the study. It was done with an aim to predict the prognosis of RA at an early period by Anti-CCP comparing it with the traditional markers of RA like CRP, ESR and RA factor. **Results:** The findings from the study depict that mean age of the patients with RA in the present study is 39.88 ± 9.07 . Among the study participants majority were female (68%). Biochemical analysis shows median Anti-CCP level 21.05 (6.35, 30.58), Mean ESR level to be 20.88 ± 14.44 and median CRP level 1.20 (0.43, 3.02) and 46% positivity for Rheumatoid factor in RA patients respectively. On comparing Anti-CCP level between cases and control, a significantly increased concentration of Anti-CCP level was anticipated in patients (27.87 ± 14.10 versus 4.44 ± 2.18). Receiver Operating Characteristics (ROC) curve demonstrates the diagnostic efficacy of Anti-CCP for the diagnosis of Rheumatoid Arthritis. The curve depicts that at the cut-off value for Anti-CCP of 8.12 sensitivity of 92.4% and specificity 96.2% was obtained. **Conclusion:** Thus, Anti-CCP can be considered as a novel marker to predict the prognosis and treatment outcome in patients with Rheumatoid Arthritis.

565T

A rare case of familial hyperoxaluria and male infertility. *R. Frikha¹, MA. Msedd²*. 1) Department of Medical genetics , University teaching hospital of Sfax, Sfax, Tunisia; 2) Department of Urology , University teaching hospital of Sfax, Sfax, Tunisia.

Introduction: Hyperoxaluria is an autosomal-recessive disorder of endogenous oxalate synthesis characterized by an accumulation of calcium oxalate primarily in the kidney, and leads to nephrocalcinosis and end-stage renal disease. A wide spectrum of associated anomalies may be present and there are three forms of primary hyperoxaluria in which the underlying defects and clinical onset and severity have been identified. **Case:** It's about a 48-years-old Tunisian man presented with primary infertility since 11 years for who oligo-astheno-azoospermia was related to testicular microlithiasis. Evaluation of primary infertility was performed by assessment of personal and familial medical history, physical exam, ultrasound exam, hormonal and infectious testing. **Results:** The patient had a personal history of recurrent urolithiasis and familial history of similar recurrent lithiasis in his mother (urolithiasis) and brothers (nephrolithiasis) who was infertile. Lithiasis was composed of calcium oxalate monohydrate. Physical exam was normal, FSH level was high and testicular ultrasound showed varicocele as well as bilateral hypotrophy. Never infection was detected. **Conclusion:** To our knowledge; this is the first case of familial hyperoxaluria; especially type III a rare, late and middle entity with a mitochondrial transmission, which is associated with male infertility. .

566F

Effects of oral eliglustat on bone in adults with Gaucher disease type 1: Long-term results from four clinical trials. *J. Charrow¹, T.M. Cox², E. Lukina³, P. Mistry⁴, T. Marinakis⁵, M. Foster⁶, S.J.M. Gaemers⁶, M.J. Peterschmitt⁶*. 1) Northwestern University Feinberg School of Medicine, Ann and Robert H. Lurie Children's Hospital of Chicago, Chicago, IL; 2) University of Cambridge, Addenbrooke's Hospital, Cambridge, UK; 3) National Research Center for Hematology, Moscow, Russia; 4) Yale University, New Haven, CT; 5) General Hospital of Athens "G. Gennimatas", Athens, Greece; 6) Sanofi Genzyme, Cambridge, MA.

Background: Gaucher disease type 1 (GD1) is a multi-system disorder resulting from deficient activity of lysosomal acid β -glucosidase. Patients with GD1 are at risk of debilitating bone complications that impair quality of life. Eliglustat, an oral substrate-reduction therapy, is a first-line treatment for adults with GD1 who have extensive, intermediate, or poor CYP2D6 metabolizer phenotypes (>90% of patients). Unlike enzyme replacement therapy (ERT), the historic standard of care for GD1, which augments acid β -glucosidase activity in mononuclear phagocytes and is administered by intravenous infusions, eliglustat reduces glucosylceramide accumulation by partially inhibiting glucosylceramide synthase. **Methods:** We analyzed bone mineral density determinations in 4 clinical trials sponsored by Sanofi Genzyme that involved 393 GD1 patients treated with eliglustat for 4–8 years: Phase 2/NCT00358150, ENGAGE/NCT00891202, ENCORE/NCT00943111, EDGE/NCT01074944. Healthy reference scores are >-1 for T-scores and >-2 for Z-scores. **Results:** In treatment-naïve patients (Phase 2, N=26; ENGAGE, N=40), mean spine T-scores moved from the osteopenic range (>-2.5 to $\leq$ -1) at baseline to the healthy reference range after 2 years of eliglustat with continued improvement thereafter. Mean $\pm$ SEM spine T-score increased from -1.55 $\pm$ 0.28 to -0.59 $\pm$ 0.34 (n=14; mean increase from baseline: 0.96 $\pm$ 0.32), after 8 years in Phase 2, and increased to -0.53 $\pm$ 0.27 in ENGAGE patients with bone data at 4.5 years (n=9; mean increase from baseline: 0.53 $\pm$ 0.16). Femur T-scores improved over time in both trials, as did spine and femur Z-scores. In the ENCORE trial of patients stabilized after a mean of 10 years of ERT (N=157), T- and Z-scores (spine and femur) remained in the reference range for up to 4 years; least square mean spine Z-scores improved by 0.29 ($P<0.0001$). In EDGE (N=170, mostly ERT switch), patients stabilized on eliglustat during a 6- to 18-month lead-in period maintained normal T- and Z-scores (spine and femur). Eliglustat was well-tolerated throughout. Most adverse events were mild/moderate (97%) and considered unrelated to eliglustat (86%); 2% of patients withdrew due to adverse events considered drug-related. **Conclusion:** Long-term eliglustat treatment induced continued improvement in bone parameters in treatment-naïve patients; these parameters remained stable and within healthy reference ranges in trials involving patients who had been previously treated.

567T

Living with xeroderma pigmentosum in Guatemala: Yulmacap village experience. C.L. Carranza^{1,2}, N. Marroquín¹, N. Escobar¹, M. Herrera¹, C. Osorio¹, M. Guerra¹, V. Zamora¹, J. Ortiz¹, L. Alvarez^{1,2}. 1) Institute for the Research in Human Genetic and Metabolic Diseases -INVEGEM-, Guatemala; 2) Istmo University, Guatemala.

Xeroderma Pigmentosum (XP) is a rare autosomal recessive disorder characterized by DNA repair defects that cause photophobia, sunlight-induced cancers, and neurodegeneration. Many of the genes related to Xeroderma Pigmentosum are part of a DNA-repair process known as nucleotide excision repair (NER), and are named as *XPA*, *XPB*, *XPC*, *XPD*, *XPE*, among others. Yulmacap is a village located in Guatemala containing many patients with XP. This is due to the high grade of consanguinity found in this community. The mutation found in these people was first described by Cleaver J in 2006, and is a stop codon in exon 8 in *XPC* gene. It's has been difficult to explain to these people the effects of sun exposure in the affected children. So the majority of the children have severe non-melanoma skin cancer. Their tumors are advanced, with necrosis and infection and they don't use any sun protection. The main goal for the present study was to improve the quality of life of the affected children and to prevent the transmission of the XP disease to another people. We performed a molecular testing for the mutation detection to establish the affected children and the XP family carriers. For doing this, we extracted blood samples and DNA extraction to the affected families. And then performed a polymerase reaction chain followed by enzyme restriction for the mutation detection. We found 9 children affected with the disease and 27 people that are carrier of the mutation. This village is very isolated from the majority of Guatemalan population, they only speak a Mayan language and their education level is extremely low. We developed a special suit that is made from special clothes that prevent the UV light exposure. And we designed the architectural plans for the construction of a day care building, which prevents that UV light penetrates the walls and windows. On the other hand, our team gave a genetic counselling to the mutation carriers, explaining the risk of disease transmission. In conclusion, several institutions collaborate to improve the quality of life of children living with XP. The activities that we performed were the genetic characterization of *XPC* mutation and genetic counselling, the special suits for affected children, architectural planning of the day care center and the performing skin tumor surgeries.

568F

Academic performance variability amount adults with Asperger's syndrome: A clinical cohort study. P. DeSanto. American University of Antigua, Manhattan, NY.

Background: Asperger's syndrome in adulthood is frequently associated with difficulties in social interaction, however few studies have explored the possible association between academic performance variability among those diagnosed. Our goal was to assess this prevalence in a clinical cohort of patients in the US. Method: A clinical retrospective cohort study was performed in patients diagnosed with Asperger's syndrome by analyzing survey data collected at an outpatient clinic between April 1st, 2011 and June 1st, 2018. Patients successfully completed a clinical assessment questionnaire assessing education level, current career, employment status and specific hardships faced during their education along with other self-reported measures of traits. We compared the levels of education along with academic performance during times of schooling with the general population and other clinical groups. Findings: 872 adults (587 men and 285 women) were diagnosed with Asperger's syndrome in the specific study period. All patients selected for this study had attended a learning institution at one point in their lives. 260 respondents have at least a 4th grade education level while 612 respondents received a high school diploma, 284 of those respondents has received a bachelor's degree. 328 respondents received a Master's degree or higher. 375 respondents had received a grade point average of over 3.5/4.0 scale, while 170 had an average between 3.0-3.5 and the remaining 67 were below 3.0. 97% of patients complained of anxiety and increased difficulty surrounding school exams and projects. 80% reported performing below average on multiple exams and disciplines during their education. 35% reported difficulty performing projects in groups or with others. 20% reported difficulty with attendance and punctuality. Compared with the general population, people with Asperger's syndrome are more likely to have higher income careers with a specific focus more commonly around technology, science and medicine. Conclusion: Our finding lend support to reports of increased variable academic performance among patients with Asperger's syndrome. Our study has found that most patients with Asperger's have higher grade point averages, and advanced degrees despite a large number reporting academic difficulties and hardships during their education.

569T

A rare case of Jacobsen syndrome with global diffuse hypomyelination of brain. G. Mainali, A. Kumar, G. Raymond. Penn State Hershey Medical Center, Hershey, PA.

Introduction: Jacobsen syndrome (JS) is a genetic disorder caused by terminal deletion of the long arm of chromosome 11. First report by Jacobsen et al in 1973, more than 200 people have been described. Along with neurocognitive delay, patients may demonstrate malformations involving gastrointestinal, genitourinary, cardiovascular, platelet dysfunction with bleeding and central nervous systems. The deletion ranges from 7 to 20 Mb with a breakpoint at 11q23.3 in 70–80% of cases and often includes HEPACAM associated with megalencephalic leukoencephalopathy with subcortical cysts type 2 (MLC2). Around 65% of JS have a structural abnormality of the brain, issues with myelination have only been rarely reported. We describe a child with severe diffuse hypomyelination mimicking a leukodystrophy. **Case presentation:** An 8 month old presented with developmental delay and hypotonia. His medical issues included hypoplastic left heart syndrome, undescended testes, neonatal thrombocytopenia and unusual facies. Neurological exam showed, truncal as well as appendicular hypotonia with intact reflexes. Chromosome microarray was sent which showed deletion 11q23-11q24, consistent with JS. Head MRI done initially at 18 months showed severe diffuse hypomyelination of cerebral as well as deep white matter felt to be consistent with leukodystrophy. Basic metabolic work up was unremarkable. Repeat MRI at 3yr showed less extensive lesions consistent with hypomyelination. **Discussion:** The estimated incidence of Jacobsen syndrome is 1 in 100,000 newborns, with a 2:1 female-to-male ratio. Affected patients have a wide phenotypic spectrum, but ~65% of JS have some CNS structural abnormality including enlarged ventricles; brain atrophy; agenesis of corpus callosum; or pachygyria. While delayed myelination has been reported, there are only a few reports of neuroimages. Our patient had diffuse hypomyelination and was initially suspected of having a progressive destructive process of myelin, but improved with time. Awareness of the MRI presentation of JS is crucial to prevent unnecessary evaluation in this condition. .

570F

Identifying and counseling patients amenable to mutation-specific therapies in Duchenne muscular dystrophy: Knowledge of resources will fuel genetic counselors' impact. A. Ferlini¹, E. O'Rourke², M. Pastore³, A. Martin⁴.

1) University of Ferrara, Ferrara, Ferrara, Italy; 2) Sarepta Therapeutics, Inc., Cambridge, MA; 3) Nationwide Children's Hospital, Columbus, OH; 4) Parent Project Muscular Dystrophy, Hackensack, NJ.

Introduction: Mutation-specific therapies for Duchenne muscular dystrophy (DMD) are in clinical trials worldwide. More than 60% of DMD patients have DMD gene mutations amenable to two current approaches: stop codon read-through and single exon skipping. **Methods:** We identified gaps in DMD genotyping and examined potential solutions for genetic counselors to help overcome barriers in determining patient eligibility for mutation-specific therapies. **Results:** Access to genetic testing is one barrier. A survey of 27 physicians who managed more than 2,000 DMD patients showed substantial variability in genotyped patients (~25%-100%). In Europe, multiplex ligation-dependent probe amplification is widely adopted, while availability of sequencing procedures is nonhomogenous, especially in Eastern countries. Awareness of financially viable genetic testing options is essential (eg, International DMD, University of Ferrara [Europe]; Decode Duchenne [USA, Canada]). Accurate, unequivocal mutation interpretation presents a second barrier. While nonsense mutations are amenable to stop codon read-through, other DMD deletion mutations vary in amenability to exon skipping, depending on the exons involved. To direct patients to appropriate therapy, clinicians must discern if a mutation is appropriate for exon skipping and amenable to a particular exon-skipping therapeutic. This pinpoints the importance of correct genotyping and genetic counselor involvement in mutation interpretation. **Conclusions:** Access to interpretation tools and guidance is paramount for genetic counselors to advise both patients and their providers. We will review the DMD gene exon map, Leiden DMD database, and Decode Duchenne mutation-specific resource as tools to help genetic counselors determine DMD patients' eligibility for exon skipping trials, with the goal of increasing the number of genotyped patients.

571T

Return of genetic results: Social and cultural challenges in an African population. G. Landouré^{1,2,3}, S. Coulibaly^{2,4}, H.A. Sango², L. Cissé¹, S.P. Coulibaly^{2,4}, T. Coulibaly^{1,2}, A. Bocoum¹, M.E. Dembélé¹, S. Diarra², A. Maiga², A. Taméga², A. Yalcouyé², S.H. Diallo^{2,5}, S. Diallo⁵, K.G. Meilleur⁶, C.O. Guinto^{1,2}, *The H3Africa Consortium*. 1) Service de Neurologie, CHU du Point G, Bamako, Mali; 2) Faculté de Médecine et d'Odontostomatologie, USTTB, Bamako, Mali; 3) Neurogenetics Branch, NINDS, NIH, Bethesda, MD; 4) Service de Psychiatrie, CHU du Point G, Bamako, Mali; 5) Service de Neurologie, CHU de Gabriel Touré, Bamako, Mali; 6) Neuromuscular Symptoms Unit, NINR, NIH, Bethesda, MD.

Introduction: Hereditary neurological disorders (HND) are incurable and debilitating. Low literacy in a population leads to stigmatization of patients and families with genetic disease. While genetic medicine has become standard practice in developed countries, Africa has been exposed to it more recently. One key element of genetic studies is returning results to patients, especially in diseases with strong evidence of causation. In Africa, family and community leaders have an important influence in the decision-making over others in all aspects of their lives. This may negatively impact research and clinical care, if not well managed. **Objective:** To gather difficulties encountered during genetic counseling in an African setting and to share this experience, with the ultimate goal of developing recommendations or guidelines for scientists working on populations of similar cultural backgrounds. **Methods:** Patients with HND and relatives went through a consent process. The goals of the study were stated with a special emphasis on the voluntariness of participation. They were informed that results could show incurable disease and were asked if they wanted to receive the results back. Genetic counseling sessions were recorded. **Results:** Findings from 44 families were available, and 27 were returned. Three patients and the father of a minor had emotional shock upon hearing the result and the implications. Only two patients, a retired teacher and a businessman, showed denial. While the first sought care with other physicians, the second just did not believe the genetic defect was the cause of his disease. In six cases, patients were minors and only one parent was present during genetic counseling. Three patients were mentally inapt adults and were accompanied by relatives other than living spouses. Feedback of findings with extended family members was done in three cases, upon requests by patients and a head of the family with no reluctance. Two heads of family have asked the return of results be postponed due to ongoing issues in the family. The remaining families were not reachable. **Conclusion:** Returning results is an important aspect of the medical practice, especially in genetic medicine where findings have implications for patients and family members. There is no optimal way to ensure standardized return of result process across populations. It is advisable to take into account cultural specificities without undermining the voluntariness and individual preferences.

572F

Parental postzygotic de novo mutation causes different clinical manifestations of COL4A1-related disorder in one Chinese family. J. Wang, X.D. Wang. Cipher Gene, Beijing, China.

Mutations in COL4A1 inherited in an autosomal dominant manner cause a wide spectrum of disorders, including porencephaly, schizencephaly, small vessel disease, etc. The proportion of cases caused by a de novo pathogenic variant is estimated to be at least 27%. We studied one family in which the proband (a girl, 8 months old) had a clinical diagnosis of global development delay and microcephaly with brain MRI characteristics (bilateral forehead, top, occipital lobe hemorrhage, right lateral ventricle with old lesions), and her mother had seizures. A de novo pathogenic heterozygote variant in COL4A1 (c.2843G>A, p.Gly948Asp) in proband was initially identified by routine trio-whole exome sequencing (WES). MRI findings confirmed the COL4A1-related small vessel disease. However, the result could not explain mother's clinical symptom. Further investigation found that mother with seizures was a very low-frequency mosaic (Gly948Asp, 9.1%), which might be gonosomal mosaicism (a combination of germ-line and somatic mosaicism). Clinically, small-vessel brain disease manifests as infantile hemiparesis, seizures, single or recurrent hemorrhagic stroke, ischemic stroke, and isolated migraine with aura. Mother with low-ratio mosaic variant suffered from moderate clinical characteristics of seizures. Parental mosaicism of the mutation increases the risk of recurrence and should be fully considered during genetic counseling.

573T

Epigenetic, serological and self-report examinations of adolescent substance use demonstrates that sex abuse results in the initiation and acceleration of smoking: Implications for public policy and clinical practice. R.A. Philibert¹, M. Gerrard², K. Vercande¹, E. Papworth¹, R. Gibbons², M. Fleischli², A. Becker¹, C. Wadler¹, S. McElroy¹, W. Philibert¹, M. Grieve¹, H. El Manaway¹, K. Dawes¹, N. Beyer¹, A. Todorov², J. Long¹, J. Mills¹, S. Kuperman¹, A. Andersen¹. 1) University of Iowa, Iowa City, IA; 2) University of Connecticut, Storrs, CT; 3) Washington University of St. Louis, St. Louis, MO.

There is a strong relationship between a history of childhood sexual abuse and the subsequent presence of depression and substance use in young adults. However, questions exist about the primacy of the relation of sexual abuse to the onset of drug use and depression. Specifically, “*does substance use initiation precede or follow abuse?*” This is an important question because it affects how clinicians and policy makers approach substance use prevention and how clinicians approach the treatment of victims of sexual abuse. In part, this uncertainty about the primacy of these two disorders is secondary to the lack of informative longitudinal cohorts with objective measures of substance use and depression contemporaneous with the timing of the abuse. In this longitudinal study of 448 high school sophomores, we analyse the relationship between self-report of unwanted sexual contact (USC), depression, and substance consumption using self-reported, serologic, and newly developed DNA methylation assays of smoking and marijuana use. We show that USC is both common (9%) and is strongly associated with current depressive symptoms and self-reported initiation of smoking, drinking, and cannabis use. USC is also moderately associated with serological measures of cannabis use, and strongly associated with quantitative epigenetic assessments of smoking. Finally, consistent with the consumption measures, we demonstrate that students disclosing USC also report strong willingness to smoke cigarettes and use marijuana, and have favourable images of smokers and marijuana users, all strong predictors of future cigarette and cannabis use during adolescence. But they do not report differential inclination to consume alcohol, nor perceive those peers who consume alcohol favourably. We conclude that USC among American high school students is frequent, and is strongly associated with early initiation of cigarette and marijuana use that appears to result as a consequence of sexual abuse. We further conclude that: 1) detection of smoking or predisposition to smoke in adolescents by objective or subjective means could be an indicator of prior sexual abuse and 2) that victims of sexual abuse should be prioritized for substance use prevention and treatment interventions.

574F

Psychiatric pharmacogenomics: Panacea or predicament? E. Huang^{1,2}, J.K. Wagner¹, R. Pulk². 1) Lafayette College, Easton, PA; 2) Geisinger, Danville, PA.

Introduction: Pharmacogenomics (PGx), which involves the utilization of one's genotype to predict drug response, is an attractive panacea for the field of psychiatry, in which most medications used to treat mental illness are ineffective or have low adherence rates due to adverse side-effects. However, psychiatric PGx testing is controversial, in part because the mechanisms through which genetic factors influence treatment response are not well-understood. Examining tests marketed by private PGx testing companies can provide valuable insights on public perceptions and expectations of psychiatric PGx. **Methods:** Relying on online information from corporate websites, the psychiatric gene and/or medications panels offered by six pharmacogenomic companies were examined. Advertising strategy used (i.e., targeting clinicians, jointly targeting patients and clinicians, and targeting consumers) was noted. For each gene on all publicly available psychiatric gene panels, PharmGKB clinical annotations on variants associated with psychiatric drugs were reviewed. The psychiatric drugs for which each of these genes had clinical annotations were recorded and compared against each company's stated panel of medications used to make their drug recommendations. Each company's panel of medications was also compared against the list of psychiatric drugs with FDA labels that include pharmacogenomic information. **Results:** Less than half of medications for which these pharmacogenomic companies provide recommendations have the potential to be strongly influenced by the genetic information collected. Over a third of the drugs on each of the medications panels published online did not overlap with the list of psychiatric drugs with FDA labels that include pharmacogenomic information. **Conclusion:** Our findings underscore concerns that current pharmacogenomic screening panels are not yet adequately substantiated. Future research on how patient-consumers and clinicians are interpreting test reports can help inform the design of effective educational tools for dispelling confusion generated by test results. Health systems' offering of PGx screening (coverage with evidence development) for all patients might help set a “gold standard” for how these tests ought to be handled and interpreted. If precision health is the goal, mental health parity must be part of the approach. We must work together to ensure that the field of psychiatry shares in the advances of pharmacogenomics.

575T

Utilization and impact of a psychiatric pharmacogenetic panel in the Department of Veterans Affairs. L.E. Hull¹, S. Tuteja², C. Chanfreau-Cof-finier³, O. Efimova³, G. Coronado³, D.W. Oslin^{2,4}, S.L. DuVall², J.A. Lynch². 1) VA Boston Healthcare System, Boston, MA; 2) University of Pennsylvania, Philadelphia, PA; 3) VA Salt Lake City, Salt Lake City, UT; 4) Philadelphia VA Medical Center, Philadelphia, PA.

Background: In 2014, the Department of Veterans Affairs (VA) and, subsequently, Medicare started covering the GeneSight[®] Psychotropic Panel (Assurex Health, Mason, OH), a pharmacogenetic (PGx) panel used to guide psychotropic medication selection. This test assesses potential gene-drug interactions for up to 55 medications and categorizes these medications in color bins: Green ("Use as Directed"), Yellow ("Moderate Gene-Drug Interaction"), or Red ("Significant Gene-Drug Interaction"). In 2017, the VA initiated the Precision Medicine in Mental Health Care (PRIME) study to assess the utility of pharmacogenetics in mental health treatment. This analysis summarizes test use outside of the prospective study. **Objectives:** 1. Identify patient and provider characteristics associated with test orders 2. Describe the impact of the PGx panel on providers' prescribing behaviors **Material and Methods:** Using the VA's electronic health record and patient-level laboratory test data, we conducted a retrospective cohort study. We described demographic and clinical characteristics of patients tested. We assessed whether providers who ordered tests made a medication change based on the results. We analyzed how the PGx panel reported specific genetic variants and potential for adverse reactions. **Results:** There were 221 Veterans tested between 10/6/14 and 2/1/2018. We excluded 40 Veterans from analyses due to inadequate data. Our final study population included 181 Veterans with a mean age of 45.9 years. They were: 84% male, 75% White, and had a mean of 2 mental health diagnoses and 2 psychotropic medication prescriptions at baseline. The majority of providers (25/38) ordered the panel once. After testing, there was a net increase in the number of prescriptions that were classified in the patient's Green bin (+37) and a net decrease in the number of medications prescribed that were classified as Yellow (-25) or Red (-15). Desvenlafaxine was always in the Green bin and was associated with the largest increase in prescriptions (+14). The test report contained data about several genetic variants, although the implications of this information were not always clearly stated. **Conclusions:** A handful of early-adopter VA providers have used this panel to guide prescribing decisions. We are assessing whether this test has resulted in improvements in patient centered outcomes. The reports contain information about several genetic variants that could inform other aspects of care and should be studied.

576F

What do the mothers of patients with ATR-X syndrome consider their carrier risk? Y. Sato¹, T. Wada¹, M. Matsukawa¹, S. Honda¹, M. Torisima², S. Kosugi¹. 1) Medical Ethics and Medical Genetics, Kyoto University Graduate School, Kyoto, Japan; 2) Clinical Genetics Unit, Kyoto University Hospital.

[Background and Objective] Among hereditary diseases, X-linked diseases develop in males, whereas their mothers and female siblings are healthy, but can be carriers. For this reason, it is suggested that X-linked diseases place a greater psychological burden than other genetic disorders. However, studies on the psychological, cultural, and social problems of female carriers of X-linked diseases are rarely performed in Japan. This study aimed to explore the psychological, cultural, and social backgrounds of mothers who carry X-linked diseases. **[Method]** This was an exploratory qualitative study based on a self-administered questionnaire. We asked the mothers of patients with alpha thalassemia X-linked intellectual disability (ATR-X) syndrome to complete a questionnaire in cooperation with ATR-X syndrome Network Japan between November and December 2017. The frequency distributions of the responses to multiple-choice questions were examined. In addition, free description responses were sectioned by meaning, and then inductive content analysis and theoretical generation were performed. **[Results]** Of the 25 patients who responded, 16 had daughters. Among them, 5 had been confirmed to be carriers by carrier testing. Regarding the diagnosis itself, most of the respondents had recovered from the uneasiness they felt before the diagnosis was made and were able to understand their children's disease. In addition, it became possible for the subjects to connect with families with the same disorder and increase their sense of security and vision. On the other hand, some subjects experienced anxiety/dissatisfaction with the fact that the disease is rare and anxiety/guilt due to it being hereditary. As for carrier risks, when the subjects learned about them many felt guilty and became worried/uneasy about their daughters' carrier risk. Moreover, some respondents became reluctant about future pregnancies. Some subjects stated, "I hated my parents". However, there were also respondents who said that, "I did not think anything". **[Conclusion]** This an exploratory qualitative study focused the mothers of patients with ATR-X syndrome. Many respondents found that obtaining a diagnosis allowed them to plan for the future with their children. On the other hand, the respondents learned that their children had a rare disease and genetic disorder, which caused them various psychological consequences. .

577T

DMD Open-access Variant Explorer (DOVE): A scalable, open-access, web-based tool to aid in clinical interpretation of genetic variants in the DMD gene. M. Bailey, N. Miller. BioMarin Pharmaceutical Inc., Novato, CA.

Purpose: Duchenne muscular dystrophy (Duchenne) is caused by pathogenic variants in the *DMD* gene. Antisense oligonucleotides (AONs) are one emerging precision-medicine treatment for Duchenne. *DMD* molecular genetic testing results guide precision-therapy molecular eligibility, requiring health-care providers to perform analyses currently uncommon in clinical laboratory and medical practices. Clear *DMD* variant notation and interpretation are key components of clinical care with the availability of precision medicine. **Methods:** The *DMD* Open-access Variant Explorer (DOVE) is a web-based aid for *DMD* variant interpretation which additionally reports variant-specific predicted molecular eligibility for therapy. DOVE was developed in Python and adapted to the Django framework, integrates existing open-access tools, and operates independently from any previous variant report/classification. **Results:** DOVE (www.dnd.nl/DOVE) interprets colloquial and HGMD inputs of *DMD* variants to output HGMD variant nomenclature, theoretical molecular eligibility for therapy, and any predicted deleterious molecular consequences of therapy. DOVE relies on holistic *in silico* prediction of molecular eligibility for therapy in lieu of reference to an empirically-defined, "variant-eligible" list. Examples illustrate the advantage and necessity for holistic variant interpretation. **Conclusions:** DOVE may prove useful for variant interpretation both at patient-level and in large-scale programs such as newborn screening and has broad application in concept to molecular genetic test result interpretation.

578F

Regulation of mitochondrial replacement therapy: A literature review analysis of where Canada, the US, the UK, and Mexico stand. F. Noohi^{1,2}, A. Caulfield², Y. Joly^{1,2}. 1) Department of Human Genetics, McGill University, Montreal, Quebec, Canada; 2) Centre of Genomics and Policy, Montreal, Quebec, Canada.

Mitochondrial Replacement Therapy (MRT) is a new type of *in vitro* fertilization that aims to prevent the transmission of mitochondrial diseases (maternal transmission) by replacing mutated mitochondrial DNA in unfertilized oocytes or zygotes with normal mitochondria from a healthy donor. Since as a result of MRT, permanent changes are made to the germline that would be transmitted through generations, this controversially so-called "three-parent IVF" can be considered as a means of genetic modification. Besides the UK, which became the first country to approve MRT in 2015, only a few countries have addressed this controversial technique through public policy. Whether MRT should be considered genetic modification has been the center of controversy in debating the legitimacy of MRT among relevant stakeholders. In order to explore how the interdependence of ethics, policy, and public views can address questions that arise from new reproductive technologies and analyze the influence of socio-cultural factors upon stakeholders' motives for using MRT, we conducted a systematic literature review of relevant academic sources which characterized the socio-ethical issues and scientific advancements of MRT according to Oxford's Bodleian Libraries review guidelines. Next, we reviewed and assessed pertinent legislation, policies, and ethical guidelines, in both the research and clinical contexts, in Canada, the US, the UK, and Mexico whose approaches in governing the controversial technique ranged from strict bans with serious consequences to an absence of policy. The major themes identified in our review included, concerns over safety and efficacy, slippery slope of germline modification, issues of identity, risk for future generations, equity of access, available resources, and beneficence. Next, the interdisciplinary knowledge produced as a result of this research will be used to analyze perceptions and expectations of relevant stakeholders (clinicians, policymakers, couples with affected children, at risk couples who wish to have biological children, and egg donors—potential so-called third parents—with regard to MRT in the aforementioned countries.

579T

Genetic counseling for 397 deaf-mute couples directed by genetic testing. M. Han, P. Dai. Department of Otolaryngology-Head and Neck Surgery, Chinese PLA General Hospital, Beijing, 100853, China.

Objective: To analyze the molecular pathogenesis of deaf-mute couples by genetic testing and provide genetic counseling with different etiology based upon results of genetic testing. **Methods:** 397 deaf-mute couples joined in this study from 2005 to 2017. Each subject is with moderate to profound hearing loss. Genomic and mitochondrial DNA of each subject was extracted from whole blood. Genetic testing including *GJB2*, *SLC26A4* and *mtDNA 12S rRNA* was offered to each subject. And then the genetic counseling was followed. **Results:** There were 119 deaf couples (30.0%, 119/397) with high recurrence risk from 25% to 100%. It is 100% in 88 deaf couples, as each pair were caused by the same deaf genes (*GJB2* or *SLC26A4*); 18 couples had already had deaf offsprings before taking genetic testing; 2 couple got pregnant by means of artificial insemination with the sperm from the local Human Sperm Bank and had normal hearing babes confirmed by prenatal diagnosis; 2 couples got pregnancy and gave deaf babies to birth who were taken cochlear implantation and received the satisfied rehabilitation result. In 26 deaf couples, the recurrence risk is 50%: 20 of them were the one caused by homozygous/compound *SLC26A4* or *GJB2* mutations and the other with a single *SLC26A4* or *GJB2* mutation; 4 of them were the one diagnosed with Warfield's syndrome; 2 of them were the one carried a single *GJB2* mutation which is dominant inheritance. The recurrence risk of 5 couples was 25%: 3 of them were both *SLC26A4* carriers and the other 2 were both *GJB2* carriers. 8 of these 31 couples had normal hearing babies assisted by prenatal diagnosis. In 3 couples, the wives were *mtDNA* A1555G mutation carriers; the instruction for them is that their offsprings must strictly avoid the administration of aminoglycoside antibiotics. The rest of 278 couples (70.0%, 278/397) were with low recurrence risk (1~3%). 54 couples were caused by different genes; 144 couples were the one with hereditary deafness but the other one with the unknown cause; and 80 couples all had unknown cause. 67 of 278 deaf couples had normal hearing babies; 2 couples got the deaf offsprings and then 1 of them was confirmed to be compound MYO7A mutations by NGS. **Conclusion:** There is a high recurrence risk if a deaf-mute family. Genetic testing plays an important predicting role in genetic counseling for deaf-mute couples. **Key words:** Deafness; Genetic diseases; Genetic testing; Genes; Genetic counseling; Diagnosis.

580F

Clinical and epidemiological description of congenital anomalies associated with visual and auditory deficit in Bogotá Colombia from 2001 to 2016. K. Sarmiento¹, A. Ibáñez-Morantes², F. Suárez-Obando², J. Ángel², P. Ayala-Ramírez², N. Gelvez², G. López², J.C. Prieto², C. Rodríguez³, M.L. Tamayo², N. Tamayo², I. Zarante². 1) Departamento ciencias fisiológicas, Facultad de Medicina. Pontificia Universidad Javeriana, Bogotá, Colombia; 2) Instituto de Genética Humana. Facultad de Medicina. Pontificia Universidad Javeriana, Bogotá, Colombia; 3) Secretaría de Salud Distrital. Bogotá, Colombia.

Introduction: Worldwide, it is estimated that 32.4 million people are blind and approximately 360 million suffer from hearing loss. The congenital anomalies that are associated with deafness and blindness affect globally the health of patients. In Colombia, the Institute of Human Genetics at the Pontificia Universidad Javeriana, with the support of the Colombian Administrative Department of Science, Technology and Innovation (Colciencias -contract 44842-656-2015 / 724-2015), develops the program AIVA (Comprehensive Care for Families with Orphan Diseases with Visual and Hearing Anomalies) that focuses on the study of these anomalies. **Objective:** To describe the epidemiology of congenital visual and hearing anomalies in the city of Bogotá between 2001 and 2016. **Methodology:** A retrospective analysis was carried out based on the ECLAMC design (Estudio Colaborativo Latinoamericano de Malformaciones Congénitas) of cases and controls (case-control ratio of 2: 1). The information was acquired from the Congenital Anomalies Surveillance and Monitoring Program of Bogotá. Newborns with auditory and/or visual abnormalities were selected and classified according to their clinical presentation in 6 groups: exclusively Audiological, exclusively visual, multiple birth defects (nonsyndromic), syndromic, teratogenic effect and mixed clinical presentation of audiological and visual compromise. **Results:** 2,541 newborns presented congenital visual and audiological anomalies detectable at birth, of which 52% were male. A prevalence rate was calculated according to each group: exclusively audiological 3.33/10,000; exclusively visual 0.06/10,000; multiple birth defects 3.73/10,000; syndromic 5.42/10,000; teratogenic 0.27/10,000 and mixed audiological and visual compromise 0.21/10,000. In comparison to the control group, significant differences were found associated with risk factors as a lower newborn's weight for all the groups of anomalies ($p < 0.005$). The gestational age was significantly lower in the groups of audiological, multiple birth defects, syndromic and teratogenic ($p < 0.005$). The maternal age was significantly higher in the mothers of the newborns with audiological, multiple birth defects and syndromic anomalies ($p < 0.005$). **Conclusions:** Congenital auditory and visual abnormalities have a high prevalence and generate disability in Colombia. Better medical intervention in primary and tertiary prevention is required to reduce disability and improvement of rehabilitation.

581T

Psychosocial impacts of Mendelian eye conditions: A systematic literature review. C.S. DAmanda, L.A. Huryn, A.E. Turriff. National Institutes of Health/National Eye Institute, Bethesda, MD.

Purpose: To review systematically psychological and social impacts of living with a Mendelian eye condition on affected individuals and their family members. To improve understanding of the needs of these patients and their family members to improve clinical care and identify research gaps. **Methods:** A systematic review of material published through March 2018 was conducted using CINAHL, Embase, PsychInfo, PubMed, Scopus, and Web of Science to include peer-reviewed, original research reporting psychological and social impacts of Mendelian eye conditions on affected individuals and/or their family members. Two researchers screened studies for eligibility and recorded details of eligible studies in a data extraction table. Identified psychosocial impacts were iteratively coded using a thematic style analysis. **Results:** Of the 7120 studies identified in our search, 89 were eligible for inclusion. Retinitis Pigmentosa (n=26, 29%), Retinoblastoma (n=21, 24%), and Albinism (n=16, 18%) were the most frequent diagnoses evaluated. Most studies were quantitative (n=66, 74%) and had a cross-sectional design (n=78, 88%). While studies were conducted across 7 global geographic regions, the majority occurred in the European Union (n=33, 37%) and North America (n=21, 24%). Included studies were found to address impacts along 9 themes: mental health (n=44, 49%), relationships (n=37, 42%), quality of life (n=22, 25%), societal contributions (n=22, 25%), independence (n=16, 18%), coping (n=14, 16%), identity (n=14, 16%), stigma (n=14, 16%), and disclosure (n=7, 8%). There is evidence for higher rates of depression and anxiety, and lower quality of life across conditions and geographic regions. **Conclusions:** There is evidence suggesting Mendelian eye conditions negatively impact several domains of life, for both affected individuals and their family members. Future research should continue to explore understudied conditions and employ more rigorous study design. This research knowledge may be incorporated into intervention studies aimed at improving well-being in this population, patient management, clinical trials, and development of patient-centered outcomes.

582F

Are we ready for rapid genomic testing in acute paediatric care?

Attitudes of Australian health professionals. Z. Stark^{1,2,3}, A. Nisselle^{1,2,3}, B. McClaren^{1,2,3}, F. Lynch^{1,2,3}, S. Best⁴, J.C. Long⁴, M. Martyn^{2,5}, C. Patel⁶, L.J. Schlapbach^{7,8,9,10}, C. Barnett¹¹, C. Theda^{2,3,12}, J. Pinner¹³, M.E. Dinger^{14,15,16}, S. Lunke^{2,3}, C.L. Gaff^{1,2,5,17}. 1) Australian Genomics Health Alliance, Melbourne, VIC, Australia; 2) University of Melbourne, Melbourne, Australia; 3) Murdoch Children's Research Institute, Melbourne, Australia; 4) Australian Institute of Health Innovation, Macquarie University, Sydney, Australia; 5) Melbourne Genomics Health Alliance, Melbourne, Australia; 6) Genetic Health Queensland, Royal Brisbane and Women's Hospital, Brisbane, Australia; 7) Paediatric Critical Care Research Group, Mater Research Institute, The University of Queensland, Brisbane, Australia; 8) Paediatric Intensive Care Unit, Lady Client Children's Hospital, Brisbane, Australia; 9) Faculty of Medicine, The University of Queensland, Brisbane, Australia; 10) Department of Paediatrics, Bern University Hospital, Inselspital, University of Bern, Bern, Switzerland; 11) Paediatric and Reproductive Genetics Unit, Women and Children's Hospital, Adelaide, Australia; 12) Neonatal Services, Royal Women's Hospital, Melbourne, Australia; 13) Sydney Children's Hospital, Sydney, Australia; 14) Graven Institute of Medical Research, Sydney, Australia; 15) Genome.One, Sydney, Australia; 16) St. Vincent's Clinical School, University of New South Wales, Sydney, Australia; 17) Walter and Eliza Hall Institute, Melbourne, Australia.

Aim To investigate the attitudes of healthcare professionals towards the application of genomic testing in neonatal and paediatric intensive care units (NICU/PICU). **Methods** Before initiating a multi-centre implementation study of rapid genomic testing, we surveyed 289 participating intensive care physicians and genetics professionals, including clinical geneticists, genetic counsellors, and laboratory genomic scientists about perceived usefulness of results, preferences for service delivery models, and implementation readiness. We examined the confidence of intensivists with microarray and genomic testing, and conducted semi-structured interviews. **Results** The overall response rate was 58%, 46% for intensivists (79/170) and 76% (90/119) for genetics professionals. Intensivists reported moderate confidence with using microarray tests and lower confidence with genomic tests. The majority of intensivists (76%), clinical geneticists (87%) and genetic counsellors (81%) favoured a clinical genetics-led service delivery model of genomic testing in the NICU/PICU. Perceived utility of genomic results was lower in the intensivist group compared to the genetics professionals group (20% v 50%, p<0.001). While support for implementation was high, the main concerns expressed by genetics professionals related to implementation climate and organizational readiness. **Conclusion** Successful large-scale implementation of genomic testing in NICU/PICU will require the development of interdisciplinary service delivery models that support adoption of genomics by the intensive care workforce while ensuring safety, sustainability and efficiency.

583T

Evaluation of the clinical utility of proactive genetic screening for medically actionable conditions. J. Oldziej, P. Aghajanian, Y. Niu, C. Neckelmann, H. Gao, S. Strom. Fulgent Genetics, Temple City, CA.

With the introduction of the Precision Medicine Initiative in 2015, preventive medicine is becoming more commonplace within today's healthcare system. Consumers have an increased awareness of the importance of personalized health management and are taking a more proactive approach in their own individual disease prevention, diagnosis, and treatment plans, specifically within the field of Consumer Genomics. Fulgent Genetics' ACMG Secondary Findings Next Generation Sequencing (NGS) panel analyzes 59 genes outlined by the American College of Medical Genetics ("ACMG SF v2.0"). This physician-ordered test is designed to provide anyone the ability to test for genetic predispositions for common health concerns, including cancer- and cardiac-related issues. A total of 849 self-referred patients completed Fulgent's ACMG Secondary Findings NGS panel testing over the course of 5 months' time (November 2017-May 2018). Only findings determined to be clinically significant and actionable by the laboratory's clinical team were reported. Neither "Variants of Unknown Significance" nor single heterozygous variants in the recessive genes were reported. Of the total patients tested, 31 individuals (or 3.65%) had a clinically significant finding. The total number of patients with a cancer related finding was 11 (or 1.30%) and the total number of patients with a cardiac related finding was 20 (or 2.35%). There were no individuals identified as carrying any non-cancer or non-cardiac related finding within these 59 genes. No individuals tested were found to carry more than 1 variant. This study supports and justifies the utilization of population-wide genetic screening of these well-studied 59 genes. The Consumer Genomics, proactive approach to medicine will empower individuals in the prevention of their predisposition to develop certain common genetic diseases and will ultimately save healthcare time and money, as well as improve and elongate lives. We believe laboratories carry the responsibility of ensuring clear communication of the capabilities and limitations of this type of testing, as well as the meaning of results both positive and negative.

584F

Harmonizing outcomes for genomic medicine: Comparison of eMERGE outcomes to ClinGen outcome/intervention pairs. M.S. Williams¹, W.K. Chung², A. Fedotov³, K. Kiryluk⁴, C. Weng⁵, J.J. Connolly⁶, M. Harr⁷, H. Hakonarson^{8,9}, K.A. Leppig⁸, E.B. Larson⁸, G.P. Jarvik¹⁰, D.L. Veenstra¹¹, C. Hoell¹², M.E. Smith¹³, I.A. Holm¹³, J.F. Peterson¹⁴, J.L. Williams¹. 1) Genomic Medicine Institute, Geisinger, Danville, PA; 2) Columbia University, Departments of Pediatrics and Medicine, New York, NY, USA; 3) Irving Institute for Clinical and Translational Research, Columbia University, New York, NY; 4) Department of Medicine, Division of Nephrology, Columbia University, New York, NY; 5) Department of Biomedical Informatics, Columbia University, New York, NY; 6) Children's Hospital of Philadelphia, Philadelphia, PA; 7) Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 8) Genetic Services, Kaiser Permanente of Washington, Seattle, WA USA; 9) Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA; 10) University of Washington, Departments of Medicine (Medical Genetics) and Genome Sciences, Seattle, WA, USA; 11) University of Washington, Department Pharmacy, Seattle, WA,; 12) Northwestern University, Center for Genetic Medicine, Chicago IL USA; 13) Division of Genetics and Genomics, Boston Children's Hospital, and Department of Pediatrics, Harvard Medical School, Boston MA USA; 14) Vanderbilt University Nashville TN USA.

Introduction: As genomic medicine moves from research to the clinic, evidence of the impact of genomic medicine interventions on health outcomes is needed. Standardized outcome measures to evaluate the impact of interventions targeted to specific genetic conditions are lacking. **Methods:** The electronic Medical Records and Genomics (eMERGE) Outcomes working group (OWG) developed measures to collect information on outcomes following the return of genomic results to participants for several genetic disorders. These outcomes were compared to outcome intervention pairs for the same genetic disorders developed independently by the Clinical Genome Resource's (ClinGen) Actionability working group (AWG). **Results:** A total of 12 disorders were scored. Three disorders were included in eMERGE, but were not scored by ClinGen, as they didn't fit the AWG criteria. In general, there was concordance between the defined outcomes of the two groups. The ClinGen outcomes tended to be higher level and the AWG scored outcomes represented a subset of outcomes referenced in the accompanying AWG evidence review. eMERGE OWG outcomes were more detailed and discrete, facilitating collection of relevant information from electronic health records. Five gene-disorder pairs had equivalent definitions between the OWG and AWG facilitating comparison of outcomes. Only two of the outcome intervention pairs scored by AWG were not present in the eMERGE outcomes. Both represented health outcomes rather than process or intermediate outcomes. eMERGE did not include many health outcomes due to the short follow-up time for participants. The remaining four had differences in the gene-disorder pair definitions that impacted the comparison. The eMERGE OWG grouped disorders into broader categories compared to the AWG. This illustrates the tension between pragmatic decisions to reduce the burden to collect outcomes of interest, at the expense of capturing outcomes that are specific to individual disorders lumped within these overarching categories. **Conclusion:** This project demonstrates that common outcomes for genomic medicine interventions can be identified. Some key differences in the definition of the disorders could inform approaches to collect relevant outcomes. Further work is needed to standardize outcomes across genomic medicine implementation projects and make these publicly available to enhance dissemination and assist in evaluating the impact of precision public health.

585T

The life of a “poor metabolizer” — through the eyes of medical genetics summaries. *L.C. Dean, A.J. Malheiro, B.L. Kattman.* Medical Genetics and Human Variation, National Center for Biotechnology Information, Bethesda, MD.

An individual who is a poor metabolizer does not have adequate enzyme capacity to metabolize drug(s). Potential consequences include potentially severe adverse effects or toxicity (for active metabolites that are inactivated by the enzyme) or decreased or absent therapeutic effects (for pro-drugs that are pharmacologically activated by the enzyme). The enzyme CYP2D6 contributes to the metabolism of over a quarter of drugs prescribed in the USA today, including antidepressants, antiseizure drugs, antiarrhythmics, and more. Individuals who are carriers of two non-functional *CYP2D6* alleles have little or no enzyme activity and are defined as “CYP2D6 poor metabolizers”. Up to 10% of the US population fall in this category. NIH’s Medical Genetics Summaries (MGS, www.ncbi.nlm.nih.gov/books/NBK61999) provides practical information about genetic testing to guide drug therapy including translating alleles into functional consequence and describing therapeutic recommendations based on genotype — so that therapy can be optimized for each individual. Developed as one of NCBI’s precision medicine resources, each MGS article focuses on one drug, and the gene(s) that influences the drug metabolism and response. MGS (1) synthesizes pharmacogenomic information from authoritative sources, such as FDA, and practice guidelines from medical and professional societies; (2) describes how to interpret alleles to help clinicians understand a lab report; (3) links to the relevant orderable genetic tests at NIH’s Genetic Testing Registry (GTR), and (4) presents dosing recommendations based on genotype from authoritative sources. The majority of individuals who are CYP2D6 poor metabolizers are unaware until they are prescribed a drug that requires this metabolic pathway. Some drugs may be metabolized by an alternative pathway in which case this phenotype may have little or no consequence. However, for a growing number of drugs, CYP2D6 poor metabolizers will respond differently than a normal metabolizer. For example, they may get little or no pain relief from codeine or tramadol, have an increased risk of relapse of breast cancer if taking tamoxifen; and be more likely to experience side effects if taking a tricyclic antidepressant. This presentation will demonstrate how to use MGS in clinic to help ensure that appropriate genetic testing is considered and lab reports are interpreted correctly to optimize initial dosing and therapeutic management for each individual based on genotype.

586F

Clinical application of targeted whole exome sequencing (WES) at the Geneva University Hospital: An update. *E.B Hammar¹, S. Fokstuen¹, M. Guipponi^{1,2}, E. Ranza¹, A. Paoloni-Giacobino¹, K. Varvagiannis¹, L. D’Amato-Sizonenko¹, P. Makrythanasis^{2,4}, L. Quteineh¹, M.T. Carminho Amaro¹, M. Lidgren¹, F. Sloan-Béna¹, S. Gimelli¹, F.A. Santoni^{1,2}, S. Laurent¹, C. Gehrig¹, A. Vannier¹, T. Nospikel¹, J.-L. Blouin¹, M. Abramowicz^{1,2}, S.E. Antonarakis^{2,3}.* 1) Service of Genetic Medicine, University Hospitals of Geneva, Switzerland; 2) Department of Genetics and Development, University Medical Center (CMU), Geneva, Switzerland; 3) iGE3 Institute of Genetics and Genomics of Geneva, Geneva, Switzerland; 4) Biomedical Research Foundation of the Academy of Athens, Athens, Greece.

In 2012, we introduced in Geneva whole exome sequencing (WES) followed by targeted bioinformatics analysis of individual in silico gene panels for the diagnosis of Mendelian disorders and we have created a multidisciplinary working group, the Genome Clinic Task Force which meets once a week. During these meetings, clinical cases and molecular results are presented, the class of variant pathogenicity is debated and the final laboratory reports are critically discussed. Financial access to the analyses and ethical issues such as informed consent, disclosure of incidental findings and/or of VUS are also addressed. In Switzerland, reimbursement of diagnostic NGS tests is integrated in the public health insurance since January 2015, and our NGS approach was accredited by the governmental Swiss Accreditation Service (SAS) in early 2018. A cornucopia of diseases have been analyzed to date, including developmental delay, epilepsies, cardiac and renal diseases, sensory disabilities, and various genetic syndromes. To date, a total of 765 cases (235 with developmental delay (DD) and 530 with various other mendelian diseases (VDM)) have been analyzed. We found pathogenic variants (class 4 or 5) in 29% of patients with DD and in 32% with VDM; the average detection rate of (likely) causative variants was 31%. The decision to report VUS was made on an individual case basis. We regularly update our panels and invite colleagues from other specialties to discuss the results of their patients at our Genome Task Force. Our main aims are to 1) further develop the Task Force into an open access Genome board for the various medical specialties interested in genomics 2) implement reimbursement also for trio analysis especially for DD cases 3) introduce whole genome sequencing for specific cases 4) participate in a national platform for annotating local variants, in order to boost the development of personalized health care in Switzerland.

587T

A novel web-based application for searching pharmacogenetic information for migalastat amenability in Fabry disease. X. Wu¹, J. Masanoff¹, M. Mustafa¹, M. Lyles¹, V. Kessler¹, B. Stockham², B. Zipkin³. 1) Amicus Therapeutics, Inc., Cranbury, NJ; 2) Amicus Therapeutics UK Ltd, Buckinghamshire, UK; 3) Pharma Forward LLC, Madison, WI.

Fabry disease is a rare, X-linked disorder of α -galactosidase A (α -Gal A) deficiency caused by mutations in the gene encoding α -Gal A (*GLA*). Migalastat, an oral pharmacological chaperone, was designed to treat Fabry disease by binding to and stabilizing specific mutant forms of α -Gal A, the genotypes of which are referred to as *amenable* mutations. The amenability status of >1000 Fabry disease-associated *GLA* mutations has been determined using prespecified qualification rules and results from a good laboratory practice-validated in vitro assay. Given the large number of *GLA* mutations, it is challenging to verify that a mutation is amenable using conventional product labelling or other print materials. We developed a dynamic, web-based search tool (galafoldamenabilitytable.com) in collaboration with regulatory agencies to help healthcare professionals quickly and accurately determine the amenability of a specific mutation. The country/region-specific website (initially available in Canada, the European Union [EU], Japan, and South Korea) uses a mutation database that contains the amenable/nonamenable mutations approved in that country/region. The application classifies a queried mutation based on whether the search term matches an amenable mutation in the database ("amenable"), matches a nonamenable mutation in the database ("nonamenable"), does not match any mutation in the database ("not tested"), or does not conform to standard genetic mutation nomenclature ("not recognized"). A mutation confirmed as "not tested" may be submitted for amenability testing and the result added to the database. The easy-to-use interface is compatible with all web browsers and devices and displays in 27 languages. Between June 2016 and March 2018, the EU database was visited 4973 times by users from 28 countries, who submitted 6583 mutation searches. galafoldamenabilitytable.com is the first website hosting comprehensive pharmacogenetic information for a precision medicine and could serve as a product-labeling model for other precision medicines involving diverse genetic backgrounds.

588F

Growing the genetic counselor workforce: A successful local effort.

C.A. Campbell^{1,2}, T.K. Grider³, S. Feely², K. Shipley⁴, K. Schaa⁵, J. Robillard⁶, S. Singh⁷. 1) Iowa Institute of Human Genetics, University of Iowa, Iowa City, IA; 2) University of Iowa, Department of Internal Medicine, Iowa City, IA; 3) University of Iowa, Department of Neurology, Iowa City, IA; 4) University of Iowa, Holden Comprehensive Cancer Center, Iowa City, IA; 5) University of Iowa, Department of Obstetrics and Gynecology, Iowa City, IA; 6) University of Iowa Carver College of Medicine, Iowa City, IA; 7) University of Iowa Hospitals and Clinics, Iowa City, IA.

Many healthcare systems are trying to grow the genetic counseling workforce to meet the needs of their patients as genetic counselors are integral to the implementation and delivery of precision medicine. An increasing number of genetic counselors also work in industry, laboratories, and insurance companies. Genetic counselors specializing in a growing number of specialties such as internal medicine, nephrology, neurology, psychiatry and cancer can assist in the implementation of genomic medicine into routine clinical care by educating non-genetics providers and patients about genomics and genomic testing. As the roles of genetic counselors continue to diversify, it can be challenging to recruit and retain these healthcare professionals. Like many hospitals in the nation, the University of Iowa Hospitals and Clinics (UIHC) had a shortage of genetic counselors which impacted the ability to care for patients as well as the ability to recruit physicians. Over the last four years, we have worked to improve the environment for genetic counselors through a series of initiatives. Our hypothesis was that implementation of education programs to attract local students to the profession combined with an improved work environment for genetic counselors, would result in growth of the local genetic counselor workforce. A summer undergraduate internship program has seen 600% growth in applicants over the past five years, and two of the past interns have been hired by the University of Iowa Hospitals and Clinics demonstrating the success of this long-term recruitment and education initiative. Commitment to pursue and successfully obtain state licensure of genetic counselors was of critical importance to grow and maintain the genetic counselor workforce. These and other education, infrastructure, and workplace satisfaction initiatives have resulted in the successful recruitment of fourteen genetic counselors in the last three years, over tripling the total number of genetic counselors at the hospital. Over two-thirds of these counselors have been hired in specialties such as cancer, cardiology, nephrology, neurology, pediatric neurology, and women's health. In summary, through a series of institutional led initiatives and creating a supportive culture, UIHC has been able to grow the genetic counselor workforce in a time of high demand. We believe the growth of this workforce will improve patient access to precision medicine thereby improving quality of care for Iowans.

589T

Evaluation of perceptions and attitudes of medical school undergraduate students towards medical genetics at El Bosque University in Colombia. R. García-Robles, A.L. Montealegre Páez, G. Díaz Muñoz, S. Perdomo Velásquez. Instituto de Investigación en Nutrición, Genética y Metabolismo, Facultad de Medicina. Universidad El Bosque (Bogotá, Colombia).

Background: The increasing knowledge and great advances in medical genetics have encountered some needs and difficulties in terms of perception and attitudes of health professionals. The common perception is that this field involves exclusively some diseases or just research, and therefore it underestimates the importance of genetics in clinical practice. **Methods:** A virtual survey with 13 questions was applied to medical students attending 3rd-year (end of the basic cycle) and 6th-year (at the end of the clinical cycle) of medical training. Questions and response options were provided with Likert-type measurement scale (totally agree to totally disagree), and questions with a binomial answer option (YES / NO). **Results:** A total of 190 students were invited to participate, 118 accepted the invitation to respond the survey but only 112 (83 female and 29 male students) accepted to participate. 64.4% and 52.3% of 3rd and 6th-year students, respectively. 83 (74.1%) were female and 29 (25.9%) male students. The majority of answers show that students recognize the importance of genetics in medicine, but face some difficulties to acquire this knowledge. Students recognize the necessity of more accessible sources of genetic information with clear and simple terminology. **Discussion:** Our results are concordant with similar studies recognizing the relevance but also the need for more education in medical genetics. These aspects are of increasing importance due to the current developments in genomics and translational medicine.

590F

Participants as partners in research: Communicating a timeline to research participants for results with a “fuzzy ending”. N.S.Y. Liang¹, E. Lau¹, B.C. Lenahan², I. Jordan², K. Ohs², J.M. Friedman¹, P.H. Birch¹. 1) University of British Columbia, Vancouver, Canada; 2) Parent Partner, Vancouver, Canada.

Background The IMAGINE project uses whole genome sequencing and metabolomics to identify genetic causes of atypical cerebral palsy in affected children. Preliminary results take 6 months, during which time there is little contact with participating families. During this period, parents speak of powerlessness, anxiety, disengagement from the research, and concern that they may have been forgotten: *“There’s nothing like the black hole of healthcare-waiting we spend our life in.”* **Aim** Our goal was to provide family-friendly, automated, informative study updates that continue to engage parents in the research. We also introduced the concept of a “fuzzy ending” to reflect the lack of a firm diagnostic endpoint due to on-going analyses and increasing scientific knowledge over time. **Method** At families’ suggestions, we co-developed an online timeline to visualize progress through the study. When a sample moves from one stage to the next, e.g., from “DNA extraction”, to “sequencing begins,” families receive an email link to an updated timeline, with optional pop-up text explanations. Parents’ evaluated the timeline via a brief survey. **Results** Feedback has been favourable: *“It’s like an order tracking bar on FedEx”* and *“keeping everyone on both sides informed feels more like active participation.”* The timeline reminds parents to notify researchers if their child’s health changes, which may help researchers to further define the child’s phenotype. Many parents wish to receive preliminary results, as indicated by the timeline: *“We want to be treated as adults - with the option of knowing everything the researchers know.”* The timeline also conveys that some families receive a definitive genetic diagnosis through our study, whereas for others, the result is inconclusive or negative. With guidance from parents, we have chosen to use the term “fuzzy ending” to convey this uncertainty and to indicate the possibility of ongoing genomic re-analysis. Rather than creating anxiety, our data suggest that some parents see the “fuzzy ending” as positive, offering hope that medical advances may provide more answers in the future. Anecdotally, the visual timeline appears to be helpful for non-English speakers, who make up 30% of our cohort, prompting us to implement and test other language versions. **Conclusion** In this pilot study, the visual timeline appears to engage parents as partners in the research by addressing their desire for ongoing involvement.

591T

Evaluating reach of genetic services to medically underserved populations. *M. Lyon, D. Maiese, A. Keehn. NCC, ACMG, Bethesda, MD.*

Purpose In June 2017, the Health Resources Services Administration (HRSA) created the seven Regional Genetics Networks (RGNs), the National Genetics Education and Family Support Center (Family Center), and the National Coordinating Center for the Regional Genetics Networks (NCC) to improve access to quality genetic services for medically underserved populations. To understand the impact of this work, HRSA created five performance measures that examine the use of RGN educational materials, RGNs assistance in connecting patients with genetics services, and the use of telehealth modalities to provide genetic services. **Methods** Through an iterative process begun in August 2017, the RGNs, NCC, and HRSA crafted the five performance measures in Table 1.

Performance Measure	Example Indicators Collected (Zip code/medically underserved collected for every measure)
Number of providers, individuals, families that received education/training or resources developed by the RGNs	Resource analytics
Number of patients for whom the RGN program facilitated connections to a geneticist	RGN specified variable (race/ethnicity, public/private insurance)
Number of providers trained in an RGN supported event that focuses on telehealth modalities for genetics	Provider Specialty
Number of RGN sites that use telehealth modalities	Provider Specialty Modality: Teleconsultation or Telegenetics
Number of patients for whom telehealth modalities for genetics were used.	Insurance coverage RGN specified variable (race/ethnicity, public/private insurance)

Results Beginning in June 2018, the RGNs submitted data for each of the five performance measures. The results of the data presented will demonstrate how: 1) how many individuals, families, and providers were served by educational materials developed by the RGNs; 2) demonstrate how RGN activities (such as a toll-free phone line) can assist in connecting medically underserved patients with genetics providers; and 3) reveal how the use of telehealth modalities can increase access to genetic services for patients. This project is supported by the Health Resources and Services Administration (HRSA) of the U.S. Department of Health and Human Services (HHS) under Cooperative Agreement #UH9MC30770-02 from 6/2017-5/2020 for \$800,000 per award year. This information or content and conclusions are those of the author and should not be construed as the official position or policy of, nor should any endorsements be inferred by HRSA, HHS or the U.S. Government.

592F

Genome Odyssey: Theatre as an effective way to teach population genetics concepts to children. *A. Malaspinas, Genome Odyssey Consortium. University of Lausanne, Lausanne, Switzerland.*

Over the course of about eight months a team of roughly 30 professionals in the arts and science worked together to create a theatre play intended for children. The starting point of the play was a scientific publication about the peopling of Australia and the intention was to communicate some scientific results to children as young as 8 years old. The play covered population genetic concepts, was meant to raise awareness about Aboriginal Australians while also conveying the fact that science is fun and is meant for all. We surveyed about 350 children before and after the play asking them the same questions to assess how much they learned from it. We present the results in this talk and show that we were able to convey scientific concepts effectively and to have an impact on implicit gender-science stereotypes.

593T

Long-term impact of presymptomatic genetic testing for HD: Family narratives. J.M. Bollinger¹, K.M. Stutgen^{1,2}, A. McCague^{1,2}, R.L. Dvoskin¹, B. Shpritz³, J. Brandt^{3,4}, D.J.H. Mathews¹. 1) Berman Institute of Bioethics, Johns Hopkins University, Baltimore, MD; 2) Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD; 3) Department of Psychiatry and Behavioral Science, Johns Hopkins University, Baltimore, MD; 4) Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD.

In 1983, Huntington's disease (HD) became the first genetic disease mapped using DNA markers. From 1986-1998, the Johns Hopkins University Huntington's Disease Center offered presymptomatic genetic testing for HD in the context of a research study. Now, almost 30-years later, ongoing relationships with this original cohort have afforded us the unique opportunity to examine the long-term consequences of presymptomatic testing for HD. Prior studies on presymptomatic genetic testing have focused on the impact of testing on at-risk individuals over years, not decades, with little to no attention on the effect on family members. As the use of genomic sequencing in research and clinical care expands, more individuals are learning about genetic risks for conditions pre-symptomatically, in some cases incidentally, and with direct implications for their family members. Thus, it is imperative that we understand the implications of presymptomatic testing, for not only at-risk individuals, but also their families. We interviewed 40 at-risk probands who enrolled in the presymptomatic HD testing program at Hopkins between 1986 and 1998. Sixteen tested with expanded repeats for HD, 20 tested normal, and 4 declined testing. Interviewees were asked about their reasons for testing and the impact of receiving results (or not) on family relationships. 32/38 interviewees cited both personal and family reasons for testing (2 affected probands were unable to answer the question). The most often cited family reasons were reproductive decision-making (17), learning information for the sake of children/family (15), financial planning (14), decisions about relationships (7) and jobs (6), and planning long-term care (3). Personal reasons focused on easing anxiety and a need to know. Probands reported both positive and negative effects on their families of testing and/or learning their status. Positive effects included becoming closer to their family (5) and better understanding of family members (4). Negative effects included resenting family members' choices to be tested or not (7), feelings of survivor's guilt (5), and drifting away from family (4). Our data demonstrate the degree to which at-risk individuals focused on the impact of this testing not for themselves, but for their families. These data highlight a need to improve and develop new guidance for how we talk to and about family members in the context of genetic testing and research.

594F

The burden of biased variant datasets: Examining enrollment and candidate genes for non-European rare disease probands in the Duke Task Force for Neonatal Genomics. E.F. Bullis¹, E.M. Ghanaim¹, A. Sadeghpour², E.E. Davis², S.H. Katsanis¹. 1) Duke University, Durham, CA; 2) Center for Human Disease Modeling, Duke University Medical Center, Durham, NC, USA.

Disproportionate representation of racial and ethnic diversity in variant databases has clinical and social implications for individuals of non-European descent. This has prompted researchers to promote diversity among participants in genomic studies, and to improve engagement with economically disadvantaged communities. We examined how race, ethnicity and socioeconomic status (SES) influenced a trio-based whole exome sequencing (WES) research study of families with probands who present with unexplained medical conditions. Following bioinformatic filtering for rare variants, the research team reported anecdotally that analysis for some families of under-represented ancestries required more resources due to an increased number of candidate variants. To probe this, we collected data from the Duke Task Force for Neonatal Genomics, a project that intersects WES data with *in vivo* models to elucidate the genetic basis for phenotypes. We compared self-reported race/ethnicity to the number of candidate genes derived from the WES data for each proband. We then layered the data with median household income to gauge any additional correlation and compared these factors to families who declined participation. Among the 101 probands (73 White; 20 non-White; 8 mixed), non-White and mixed probands ($\bar{x}=14.15$; range=41) had an average of 6.1 more candidate genes ($p=0.00035$) than white probands ($\bar{x}=8.05$; range=31). Probands that self-identified as Latino were analyzed separately and then in combination with their reported race. Probands identifying as White and Latino ($n=7$) were found on average to have 5 more candidate genes than White non-Latinos ($n=61$) ($p=0.012$). By comparing the 211 enrolled families to the 239 families that declined, we found that families who declined participation had \$8,222 lower median household income ($p=0.015$). In addition, families with a higher number of candidate genes had lower median household incomes. Although the resources needed to analyze candidate genes for causality can range dramatically, more candidates broadly translate into a higher burden on the laboratory. Our analysis reinforces the perception that race, ethnicity, and SES are interconnected and can contribute to both the enrollment of families into genomic studies and the increased effort to analyze sequencing data. Future community endeavors will be necessary to alleviate such barriers for equitable access to actionable clinical results from genome sequencing.

595T

Gene editing: Emerging law and policy perspectives in Europe. A.C. de Paor. School of Law and Government, Dublin City University, Glasnevin, Dublin 9.

Scientific and technological innovation is becoming more refined, resulting in an increase in the availability and use of genetic testing, and other cutting edge genetic technologies, including gene editing. CRISPR/Cas 9, a gene-editing technology at the cutting edge of science and medicine, is gaining attention worldwide and may ultimately be used to eradicate serious disease. Recent developments in the area of gene editing signal significant advances and growing application of this new technology in the future. Although offering substantial opportunities for understanding health and medicine, the emergence of gene-editing highlights a minefield of ethical, legal and practical challenges in both clinical and commercial contexts, and in society in general. In addition to the safety concerns arising with these genetic techniques and the unknown medical implications of gene editing for future generations, there are many ethical and legal questions arising. One concern centres on the potential trend towards facilitating the selection of embryos based upon genetic desirability and human enhancement. Disability perspectives are particularly relevant in this area. Advances in the field of genetics provide a new lens to view, identify and now eradicate disability. With developing gene-editing technologies there is a concern that the technology will be used to screen out disability, difference and diversity in society. This provokes a range of human rights concerns for persons with (current and future) disabilities, and those genetically predisposed to disability, including discrimination and stigmatisation. Without caution and appropriate regulatory controls, such developments may operate to devalue the lives of persons with disabilities, and may lead to the widespread view that persons with disabilities are inferior, which may result in discrimination and other mistreatment. In consideration of the connection between genetics and disability, this paper will highlight the key ethical, legal and policy concerns raised as gene editing advances and the consequences for disability rights, the right to non-discrimination, and respect for diversity and difference. The research looks particularly at the United Nations Convention on the Rights of Persons with Disabilities and how it may be used to guide best practice and policy in this area in Europe.

596F

The duty to recontact in genomics: Is the law out of touch? B. Knoppers, A.M. Thorogood, S. Beland, M.H. Zawati. McGill University, Montreal, Quebec, Canada.

The evolving clinical implications of genomic data results present opportunities for invention and intervention, and raise ethical and legal questions, particularly with regards to the duty of physicians to provide medical follow-up. We compare case law and professional guidelines determining the content of the duty of care in the genetics context, as well as the contours of a physician's general ethical and legal duties to follow-up in different countries (Canada, U.S., UK, France, Australia, ...), highlighting differences and similarities between jurisdictions. Additionally, we review relevant literature and policy recommendations emanating from various organizations (American College of Medical Genetics, European Society of Human Genetics, ...). We find unsurprisingly, that there is currently no general legal duty for physicians to recontact patients in light of new clinical information. There may be specific situations however, where a legal duty to recontact arises. A comprehensive personalized medicine ecosystem is emerging aiming to improve the quality, availability, and currency of genomic data. This ecosystem will involve modification and expansion of the organizations and health professionals involved in health care delivery (to include for example genetics laboratories and databases), as well as automated genomic communication and interpretation tools, all of which make up an interpretive supply chain. Preemptive genomic sequencing, long term patient genomic data storage for reuse, in combination with Big Data collection and analysis, are creating new possibilities for continuous update. We anticipate that the dual evolution of the genomic knowledge base and individual patient data sets needed to capture the benefits of genomics will substantially expand the legal duty to follow-up. Moreover, the locus of this legal responsibility may also be shifting across a broader range of individuals in the interpretive supply chain (primary care physicians, genetics healthcare professionals, genetics laboratories, database managers, patients, ...). Evolving data management practices and recontact may also have implications regarding patients' privacy and right not to know, which we consider in our analysis.

597T

Preferences and incidence of secondary findings in a cohort of adult patients undergoing exome sequencing. S. Shickh^{1,2,11}, K. Zakoor¹, M. Gutierrez Salazar^{1,2}, J. Gu¹, N. Watkins^{1,3}, M. Care^{3,4}, S. Horsburgh⁴, O. Morar⁴, J. Murphy^{3,4}, M. Szybowska⁴, D.L. Nevay⁴, M. Aronson^{3,5}, S. Panchal^{3,6}, J. Stavropoulos⁷, C. Marshall¹, N. Abdul¹, S. Armel⁸, S. Khalouei¹, R. Godoy⁶, K. Semotiuk¹, C. Elser⁸, Y. Bombard^{2,11}, C. Mighton^{1,2,11}, D. Chitayat¹, R. Kim^{1,4,7,8,9}, J. So^{4,9}, H. Faghfoury^{4,9}, J. Silver^{4,9}, C. Morel^{4,9}, J. Lerner-Ellis^{1,10}. 1) Pathology and Laboratory Medicine, Mount Sinai Hospital, Sinai Health System, Toronto, ON; 2) Genomics and Health Policy Research, Li Ka Shing Knowledge Institute, St. Michael's Hospital, Toronto, ON; 3) Department of Molecular Genetics, University of Toronto, Toronto, ON; 4) Fred A. Litwin Family Centre in Genetic Medicine, Toronto General Hospital and Mount Sinai Hospital, Toronto, Canada; 5) Familial Gastrointestinal Cancer Registry, Mount Sinai Hospital, Sinai Health System, Toronto, ON; 6) Marvella Koffler Breast Centre, Mount Sinai Hospital, Sinai Health System, Toronto, ON; 7) Hospital for Sick Children, Toronto, ON; 8) Princess Margaret Cancer Center, Toronto, ON; 9) Department of Medicine, University of Toronto, Toronto, Canada; 10) Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON; 11) Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON.

Background and Purpose Exome sequencing (ES) is increasingly utilized in practice, given its ability to improve diagnostic yield, inform surveillance and provide treatment options at a cost comparable to standard genetic tests. However, ES presents challenges for patients and clinicians, including generating secondary findings (SF). The American College of Medical Genetics and Genomics (ACMG) recommends that at minimum, patients undergoing ES be offered the option to receive results for 59 medically actionable genes, which ACMG defined as genes associated with highly penetrant disorders with available medical intervention. Despite the recommendations, there continues to be variation in the return of SF across laboratories and clinics which may reflect provider and patient preferences. Studies evaluating patient preferences have generally found that participants are interested in receiving results beyond those that are medically actionable. We sought to survey adult patients undergoing ES and their relatives to evaluate preferences for receiving SF and the frequency of SF. **Methods** We surveyed 40 adults patients suspected to have genetic disorders, but had previous uninformative results and were undergoing ES as part of a study at Mount Sinai Hospital in Toronto, Ontario and 31 of their relatives via questionnaire. The questionnaire asked participants about their level of interest in receiving SF for diseases with varying penetrance levels and types of genetic variants/conditions. Patients' exome data was also analyzed to evaluate the frequency of medically actionable results (ACMG 59 genes plus actionability list developed by the Clinical Genome Resource) and carrier results. **Results** Overall, probands and relatives were interested in receiving SF, with a high level of interest for diseases with higher penetrance (63/71) and later onset diseases with available treatments (70/71). Most participants were also interested in receiving carrier results (66/71) and variants of uncertain significance (64/71). Relatives tended to have a higher preference for receiving SF than probands. The prevalence of medically actionable variants and carrier results identified in this cohort will be reported on as well. As ES becomes increasingly utilized in clinical practice, understanding patient preferences and the frequency of different types of SF is critical as they can help to inform resource needs and guidelines on the return of SF.

598F

Genetic risk: Whether, when, and how to tell adolescents. K.M. Stutgen^{1,2}, A. McCague^{1,2}, J.M. Bollinger¹, R.L. Dvoskin¹, B. Shpritz², J. Brandt^{3,4}, D.J.H. Mathews¹. 1) Berman Institute of Bioethics, Johns Hopkins University, Baltimore, MD; 2) Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 3) Department of Psychiatry and Behavioral Science, Johns Hopkins University School of Medicine, Baltimore, MD; 4) Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD.

Genetic test results are often relevant not only to the person tested, but also to her/his children. Questions of whether, when, and how to disclose genetic test results to children can be difficult for parents to navigate. A better understanding of how parents and children feel about genetic risk/status, how it is communicated, and how it influences well-being and family relationships is vital in order for genetic counselors and other health care professionals to provide the best guidance possible. In-depth semi-structured interviews were conducted with at-risk parent/child pairs from two populations: families at risk for HD and families with hereditary cancers. Children in this study were ages 15-17. Parents and children were interviewed separately and were asked about topics that included: the genetic information that was communicated, outcome of the communication, factors that led to positive and negative experiences, and advice for other parents and children. In-depth semi-structured interviews were also conducted with genetic counselors (GCs) who had prior experience communicating with children about genetic risk information at the request of parents, or advising parents on how to do so. To date, interviews have been conducted with 17 parent-child pairs affected by/at-risk for hereditary cancer, 12 parent-child pairs affected by/at-risk for HD, and 19 GCs. Results suggest that while parents feel hesitant to burden children with information and/or children will not be able to handle it, children believe they are able to sense that something is wrong in the family and wish parents would be more forthcoming and offer more information. Both existing data and our data suggest that parents should start communicating with children at a young age, but conversations should occur gradually over time and should involve age-appropriate language. Additionally, our data suggest that parents should take time to process his/her genetic test results and/or diagnoses before communicating with their child to avoid becoming overly emotional when talking with children. Finally, both parents and genetic counselors expressed a need for more resources for parents and providers on how to communicate with children about genetic risk information as well as child-specific resources on various genetic conditions. We hope these data will help assist both parents and health professionals when faced with needing to communicate this information to children.

599T

The factors related with attitudes toward informed assent for genomic research. Z. Yamagata¹, K. Muto². 1) University of Yamanashi, Yamanashi, Japan; 2) The University of Tokyo, Tokyo, Japan.

Introduction: The purpose of this study is to determine the factors related with attitudes of the general Japanese public toward genetic research on children and informed based on nationwide surveys conducted 2016. **Methods:** 2,000 people (age = 20–69) from the general Japanese population, were selected using a stratified two-phase sampling method. In a mail survey administered in 2016, the participants were surveyed regarding the following topics: (1) their attitudes toward genetic testing; (2) their perspective regarding informed assent; (3) their level of scientific literacy regarding genomics; and (4) their demographic information and socioeconomic status. **Results:** The response rate was 50.8% in 2016. Conducting genetic testing on children for disease susceptibilities was favored by 57.2% of participants. Regarding obtaining blood donations from children, 55.2% approved, 11.3% disapproved, and 34.8% were undecided. A higher proportion of participants with high genomic literacy levels approved of obtaining blood donations from children. The multiple logistic analysis odds ratio regarding genomic literacy was 1.40 (95% confidence interval = 1.08–1.52). Regarding whether seeking consent from children is appropriate, 52.0% of people answered that seeking consent from a child is acceptable if the child understands the details of the research. These responses were associated with participants with high genomic literacy. **Conclusions:** The results of this study suggest that people's genomic literacy is related to people's perspective on genomic research on children.

600F

A new model for genomics support of military healthcare providers:

Early findings from the MilSeq project. M.D. Maxwell^{1,2,3}, C.L. Blout^{1,2}, K.D. Christensen^{1,2}, C.L. Gardner³, J.B. Krier^{1,2}, M. De Castro⁴, R.C. Green^{1,2,5,6} for the MilSeq Project. 1) Brigham and Women's Hospital, Boston, MA; 2) Harvard Medical School, Boston, MA; 3) 59th Medical Wing, Joint Base San Antonio - Lackland Air Force Base, San Antonio, TX; 4) Air Force Medical Genetics Center and Molecular Diagnostic Laboratory, Keesler Air Force Base, Biloxi, MS; 5) Partners Personalized Medicine, Boston, MA; 6) The Broad Institute of MIT and Harvard, Cambridge, MA.

Introduction: As technological advancement, cost reduction and accessibility of genomic sequencing (GS) usher in an era of precision medicine, the demand for expert clinical interpretation exceeds the supply of trained genetic professionals. This shortage is intensified in the Military Health System (MHS), which employs 6 active-duty and 2 civilian geneticists, as well as a handful of genetic counselors (GCs) to provide services for over 1,300,000 service members. The MilSeq Project is a pilot study of GS in ostensibly healthy Airmen designed to explore the benefits and risks of GS implementation into the MHS. To address the professional shortage, we pilot a scalable model in which military healthcare providers (MHCPs) manage GS results with support from GCs and geneticists. **Methods:** MHCPs are briefly oriented, supervised and evaluated in their provision of genetic services by a Genome Resource Center (GRC) that can be locally or remotely staffed by GCs and geneticists. MHCPs are required to attend a 3-hour, GC-instructed training prior to disclosing results to Airmen. A GC is embedded in clinic for in-person consult and ongoing support, and MHCPs have access to a 24-hour virtual GRC. MHCP disclosures are audio recorded, transcribed and reviewed by the GRC for accuracy and appropriateness. Identified opportunities to improve MHCP practice are organized into 4 categories: (1) Knowledge limitation; (2) "Minor" adjustments for errors posing minimal risk to optimal care; (3) "Moderate" adjustments for errors posing some degree of suboptimal care; and (4) "Critical" adjustments for errors endangering patient safety. Feedback is provided to MHCPs after transcript review. **Results:** Of 7 transcripts reviewed to date, no critical errors have been identified. At least 1 opportunity for improvement has been identified in each of the other 3 categories. Qualitative review revealed emerging themes that would benefit from further education, including contextualizing within the pedigree, cascade testing and relative risk counseling. We will have 75 evaluated transcripts by the end of this phase of the MilSeq Project and additional data will be included in this presentation. **Conclusion:** In order for genomic services to be scalable in both the MHS and civilian healthcare, new service delivery models are needed. Here we present our experience with a model that provides for pre-disclosure training, real-time point-of-care consultation and post-disclosure evaluation and feedback.

601T

A literacy-focused “modified” genetic counseling approach for the return of exome sequencing results to underserved and ancestrally diverse patients. *L.M. Amendola¹, M. Gilmore², K.P. Anderson³, B. Biesecker⁴, C. Guerra⁵, T.L. Kauffman², R. Lee⁶, L. Riddle⁵, A.F. Rope⁷, B. Wilfond⁸, K.A.B. Goddard⁸, G. Joseph⁸.* 1) Division of Medical Genetics, University of Washington, Seattle WA 98195, USA; 2) Center for Health Research, Kaiser Permanente Northwest, Portland OR 97227, USA; 3) Denver Health Ambulatory Care Services, Denver Health, Denver CO 80204, USA; 4) Research Triangle Institute, International, Washington DC 20005, USA; 5) University of California, San Francisco Department of Anthropology, History & Social Medicine, & Helen Diller Family Comprehensive Cancer Center, San Francisco CA 94143, USA; 6) University of California, San Francisco Cancer Genetics and Prevention Program, San Francisco CA 94143, USA; 7) Department of Genetics, Kaiser Permanente Northwest, Portland OR 97227, USA; 8) Department of Pediatrics and Seattle Children’s Research Institute, University of Washington, Seattle, WA, USA.

Background: Early research efforts on the application of exome sequencing in clinical practice have inadequately addressed diverse or underserved populations. Research conducted with genetics patients who have limited English proficiency and/or health literacy suggests a mismatch between the information provided by genetic counselors (GCs) and information desired and meaningful to patients. As part of the Clinical Sequencing Evidence-Generating Research (CSER) consortium, the Cancer Health Risk Assessment Reaching Many (CHARM) study will enroll participants with a focus on traditionally underserved populations including individuals with socioeconomic and racial/ethnic diversity as well as individuals with low health literacy. We will evaluate the disclosure of exome sequencing results via traditional genetic counseling or a literacy-focused “modified” genetic counseling approach in the context of hereditary cancer genetic testing. **Methods:** A work group within the CHARM study was convened to define the two results disclosure approaches. Here we describe the essential elements of each and implementation plans.

Results: Traditional genetic counseling (usual care) will include genetics terminology, scientific language and discussion of genetic concepts, such as inheritance patterns and disease penetrance. Modified counseling will adapt to the literacy level of the individual and will employ evidence-based communication techniques for individuals of lower health literacy, such as the use of plain language and teach back, and a more directive approach. In both arms, we will prioritize conveying essential information and address participant specific psychosocial concerns and questions. GCs conducting modified counseling will be trained via didactic lectures and role play. To ensure fidelity to the intervention, recordings of results disclosure conversations will be assessed to identify variation and protocol adherence. GCs providing traditional counseling will remain naïve to the modified model throughout the study. **Discussion:** We hypothesize that the modified approach will be non-inferior to traditional genetic counseling and may be more effective in terms of satisfaction with counseling, participant engagement, perception of communication, family communication, and understanding of and adherence to recommended care. The trial results will inform effective methods for results disclosure for exome sequencing across diverse healthcare settings and populations.

602F

Communicating uncertainties in genomics: Results of a narrative literature review. *B.B. Biesecker¹, H. Peay², R. Paquin², B. Boyea², A. Wheeler², M.A. Lewis².* 1) Research Triangle Institute, International, Washington DC, United States; 2) Research Triangle Institute, International Research Triangle Park, NC, United States.

Given the frequency of unknown variants returned after genome sequencing, there is critical need for evidence to inform best practices in communicating uncertainties. A narrative literature review on communicating clinical uncertainty identified 149 abstracts that met inclusion criteria from a search of PubMed, Web of Science, PsycINFO, and Communication Source databases. Twenty-nine abstracts described studies in genomics, 33 in cancer, 23 in other specialties and 53 in illness risk generally. Lacking best practices in genomics, existing evidence was synthesized for use in guiding communication of uncertainty practices and to inform study design in genomics. Evidence endorses: A) Use of positive framing when engaging patients in discussions of uncertainty. Examples include conveying a commitment to care for the patient, partnering and providing reassurance s/he is not alone. B) Prior to testing, helping patients anticipate the potential for uncertain findings. After testing, providing clear communication about what is known and unknown, and avoiding use of vague descriptors. C) Providing reassurance to patients that information is not being withheld. D) Communicating provider-patient shared ownership of uncertainty. E) Encouraging and providing opportunities for patients to ask questions. F) Communicating with credible authorities to facilitate effective management of uncertainty using reappraisal and reframing. Existing evidence informs several hypotheses: A) How uncertainty is communicated will affect not only understanding and risk perceptions but also how patients manage the threat of uncertainty over time. B) Many factors, including personality traits and situational factors, will play a role. C) Maintaining hope in an uncertain situation will promote effective coping and adaptation. D) For many individuals, managing information is a means of coping with illness-related uncertainty. For those who maintain perceptions of uncertainty to preserve hope, they avoid follow up care and seeking information as learning details may threaten their hope. Those who perceive uncertainty as a danger work to reduce it by seeking information and follow up care. Implications: This literature review and data synthesis provide guidance in communicating uncertainties in genomics, while simultaneously suggesting hypotheses to be tested to generate evidence to inform best practices.

603T

An iterative process to develop a statewide network of genetic counselors in Connecticut. K. Sanghavi¹, M. Gyure², J. Brown², K. Adams¹. 1) The Jackson Laboratory for Genomic Medicine, Farmington, CT; 2) Institute for Systems Genomics and Dept. of Allied Health Sciences, University of Connecticut, Storrs, CT.

Introduction: The purpose of development of a Connecticut (CT) statewide network of genetic counselors (GCs) through an iterative process is to enhance the genetic counseling workforce in CT and identify areas in which the biomedical research community and GCs can be supported to address real or perceived gaps in representation of genetic counselors in state efforts such as Bioscience initiatives, as well as educational, research and networking opportunities. **Approaches:** An IRB-approved survey was administered digitally to the GCs providing genetic counseling services in CT to understand the landscape of genetic counseling services including the need of a statewide network of GCs in CT recognizing that CT provides licensure to GCs. Follow-up discussions with select respondents were scheduled to better understand the needs and inform the development of a statewide network of GCs in CT similar to other states such as FL, GA, PA, TX, and VA. Various stakeholders including the leadership of existing statewide associations such as Pennsylvania Association of Genetic Counselors were engaged in this iterative process. **Findings:** 34/36 (94%) respondents were interested in a statewide network of GCs in CT. Topic areas for consideration in the context of the content for statewide GC efforts included: Professional issues, CT Bioscience & GCs, educational and leadership needs, clinical/research rotation opportunities for students, innovative service delivery models, state genetics/genomics plan, and reimbursement for GCs. One respondent reported, "*Credentialing and billing have been an ongoing issue, even after obtaining licensure. I would appreciate peer input or a subcommittee in the statewide organization to address practice issues such as this.*" Initial follow-up discussions revealed an overall need to advocate the inclusion of GCs in statewide initiatives and the expansion of the number of GCs in CT. Next steps include interdisciplinary focus group discussions with different stakeholders including other state GC networks. **Conclusion:** The process of establishing a network with statewide partners and stakeholders provides an example of professional development opportunities for GCs to foster the sharing and dissemination of best practices within the biomedical and GC communities.

604F

Genetic counseling services in Washington state: A capacity assessment. N. Shridhar. Screening and Genetics Unit, Washington State Department of Health, Kent, WA.

In October 2017, the Journal of Genetic Counseling (JGC) published a paper (Hoskovec et al.) suggesting optimal genetic counselor (GC) levels to provide adequate access for genetic services in the United States. The paper posited two sets of GC to general population ratios (1:75,000 & 1:100,000), as satisfying the demand for GC's nationwide, and creating an equilibrium between demand and supply. At the Washington State Department of Health (DOH), the Screening and Genetics Unit (SGU) was already engaged in a patient access study evaluating wait times for access to genetic services. This study, a follow-up to a 2013 study, aimed to better understand the impact of the Affordable Care Act (ACA) implementation and Medicaid expansion on our state's capacity for genetic services. We did this by comparing patient wait times before and after the implementation of the two events we hypothesized would increase the need for genetic services statewide. Based on the JGC publication, we augmented our study to assess WA State capacity for genetic services as a function of both wait times (patient access to genetic services) and GC supply (number of GCs in WA providing direct patient care). To conduct this follow-up study in August 2017, the SGU contacted the same nine WA Regional Genetic Clinics across the state that were contacted in October 2013 for the first study. For both studies, patient access was measured using appointment wait times, in specific, the 'Third Next Available Appointment' methodology to accurately reflect the true patient load of each facility. To estimate WA GC capacity, we obtained the list of GC's currently licensed to provide services in our state, and determined if they were providing direct patient care. When we modeled the GC ratio for Washington State (per the JGC paper) to better understand our demand and supply scenario, we found that that we had a direct patient care GC to general population ratio that was in between the two scenarios (1:75000 & 1:100,000) posited by the paper for adequate supply of GC. And yet, we found that patient wait times for appointments to genetic services in Washington State have doubled between 2013 and 2017, with an average wait time of over 2 months. In this presentation, we will provide data that suggests that the JGC paper's low end estimate of 1:100,000 is too low a supply model to fit the current demand for GC in WA.

605T

Return of data on genetic and lifestyle risks to participants of Estonian Genome Center.

N. Tonisson^{1,2}, M. Palover^{1,3}, L. Leitsalu¹, K. Krebs^{1,3}, K. Läll^{1,5}, M. Kals^{1,5}, T. Nikopensius¹, A. Reigo¹, H. Alavere¹, K. Fischer^{1,5}, L. Azin Milani^{1,6}, R. Mägi¹, S. Reisberg¹, A. Allik¹, B. Kolk¹, K. Metsalu¹, M. Puusepp¹, K. Mikkel¹, M.L. Tammesoo¹, E. Pintaar¹, T. Temberg¹, T. Esko^{1,4}, A. Metspalu^{1,3}.

1) Estonian Genome Center / Institute of Genomics, University of Tartu, Tartu, Estonia; 2) Dept. of Clinical Genetics, United Laboratories, Tartu University Hospital, Tartu, Estonia; 3) Dept. of Biotechnology, Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia; 4) Broad Institute of Harvard and MIT, Cambridge, MA, USA; 5) Institute of Mathematics and Statistics, University of Tartu, Tartu, Estonia; 6) Department of Medical Sciences, Molecular Medicine, University of Uppsala, Sweden; 7) Institute of Computer Science, University of Tartu, Tartu, Estonia.

Introduction: Estonian Genome Center / Institute of Genomics, University of Tartu, Estonia (EGCUT) holds genotype and health data of more than 52,000 individuals and is currently undergoing a large nationwide recruitment. From the end of 2017, EGCUT participants can come to counselling sessions on individual genetic risk predictions. Our standard reports include data on common diseases, high and moderate monogenic risk factors, carrier status, and pharmacogenetics. **Methods:** To date, nearly 5,000 biobank participants have been either whole genome or whole exome sequenced; 47,000 biobank participants have been genotyped with Illumina Global Screening Array. The reporting system includes a digital semi-automated report assembly system and participant portal for signing an informed consent for return of results and updating information. This portal is also used for setting the counselling appointment. The genetic risk assessments on common disorders are based on polygenic risk scores. Actionable genetic variants reported are selected, according to ACMG guidelines. Participants' immediate - and long-term feedback is continuously surveyed. **Results:** by May 2018, 914 participants have expressed interest through logging in to patient portal and 578 have attended genetic counselling sessions. **Conclusion:** The EGCUT initiative of return of results to population biobank participants provides insight on how healthy population responds to personalised genetic risk information. The long-term aim of the project is to promote introduction of personalised and translational genomics medicine in Estonia.

606F

Understanding Fabry disease symptoms and impact on daily lives: A qualitative study.

A. Hamed¹, N. Greene¹, C. Gwaltney². 1) Sanofi Genzyme, Cambridge, MA United States; 2) Gwaltney Consulting, Westerly, RI, United States.

Background: Fabry disease (FD) is a multisystemic, X-linked, genetic lysosomal storage disorder resulting in accumulation of globotriaosylceramide and other α -GalA substrates leading to irreversible tissue damage and impaired function of multiple organs. Clinical manifestations of FD are heterogeneous and may vary by age of onset, rate of progression, and severity. **Objectives:** The objective of this qualitative study is to gain an in-depth understanding of the most salient symptoms experienced by the FD patients and the impact on their daily lives. **Methods:** One-on-one concept elicitation interviews (CEI) were conducted with adult FD patients. Patients were recruited from US clinical sites and interviewed either in person or via telephone. A semi-structured CEI guide was used to obtain information on patient experiences of FD-related symptoms and impacts and to identify the language patients used to express the concepts associated with their condition. Interviews were transcribed and coded in order to identify themes in the data. The significance of FD-related symptom and impact was assessed by the frequency of patients' mention of each FD-specific concept. **Results:** A total of 37 FD patients were interviewed at six clinical site locations. Interviewed patients ranged in age from 23 to 74 years with a median age of 45.7 years. Twenty-two (60%) were women, 25 (68%) were married, and 32 (87%) were White. For concept saturation, 37 transcripts were categorized chronologically into 4 sub-groups. Complete concepts saturation of salient symptoms and impacts was achieved by the completion of the third group of interviews. Neuropathic symptoms, temperature intolerance, tiredness/fatigue, hearing/vision impairment, and gastrointestinal symptoms were most frequently mentioned by patients. The symptoms most-frequently reported spontaneously were tiredness, fatigue, and burning pain (each reported spontaneously by over 35% of the patients) and followed by heat intolerance and lack of sweating (each reported spontaneously by 32% of the patients). Work/school limitations, exercise limitations, social activity limitations, depression/memory problems, worry, fear, and difficulty with household responsibilities, were the most frequently reported daily impacts due to FD. **Conclusions:** FD patients reported a variety of symptoms and impairments impacting their daily lives. Reducing these symptoms is an important goal of FD treatment. (This work was funded by Sanofi Genzyme).

607T**Using real-world pharmacogenomics knowledge resources to guide clinical decision-making.** *R. Hart^{1,2}, G. Jarvik¹, L. Garrison, Jr.³, B. Devine^{2,3}.*

1) Division of Medical Genetics, University of Washington, Seattle, WA; 2) Institute for Public Health Genetics, University of Washington, Seattle, WA; 3) The Comparative Health Outcomes, Policy, and Economics Institute, School of Pharmacy, University of Washington, Seattle, WA.

Pharmacogenomic (PGx) testing can help to optimize drug therapy selection. Despite the availability of FDA Guidance and select expert-developed guidelines, however, clinical utility evidence gaps limit the uptake of PGx testing in clinical settings. Commercial PGx knowledge resource solutions may increase adoption. We investigated real-world data from a commercial PGx knowledge resource by characterizing the Clinical Pharmacogenomics Implementation Consortium (CPIC) level of evidence and sought to identify predictive factors for the clinical actionability made at the time of result interpretation. For 99,727 patients who underwent PGx testing, 90% of patients were tested for six or fewer drugs; 41% were male. Predictors that were statistically significant differed across the four clinical subgroups that were evaluated (pain, cardiology, psychiatry, gastroenterology). Twenty-eight of the 33 PGx gene-drug pairs, characterized by CPIC level of evidence lower than B, resulted in a warning or critical actionability recommendation (versus informational recommendation). This indicates a possible change in drug selection, and highlights a discrepancy in level of evidence determinations between public and commercial knowledge resources. To our knowledge, this is the largest retrospective study conducted to date on patients undergoing PGx testing in a broad set of clinical settings. Our findings provide a robust characterization of real-world PGx testing utilization data and demonstrate evidence of clinical actionability from a commercial PGx knowledge resource. Healthcare systems may wish to consider commercial knowledge resources to spur appropriate PGx test adoption. These solutions can adapt quicker than can centralized authoritative resources to synthesize clinical actionability data and inform time-sensitive clinical guidance. They often can be customized to the population of each Learning Health System, rather than rely on publicly available, standardized sets of PGx knowledge resources.

608F**Pharmacy-supported return of pharmacogenomic test results: Preliminary data.** *E.T. Matey, J. Giri, J. Wright, J. Anderson, L. Oyen, W.T. Nicholson.* Mayo Clinic, ROCHESTER, MN.

Background: Mayo Clinic is conducting pre-emptive pharmacogenomic (PGx) testing for 10,000 biobank participants as part of a research study. Test results for 13 genes are in the process of being returned to the electronic health record (EHR). Pharmacy leadership advocated for pharmacist involvement in providing an eConsult (electronic review of PGx results for each patient's current medications with reference to drug-gene interactions). Primary care providers (PCPs) may need guidance to manage current medications based on drug-gene interactions. Hence, this eConsult process was established to support and encourage adoption by PCPs who did not order this test. Additionally, clinical decision support alerts are available to providers regarding major drug-gene interactions for future prescriptions. Methods: In addition to the PGx test results for the 13 genes, a supplemental report for over 300 medications based on PGx results was added to each patient's EHR. These medications were categorized as semi-urgent and clinically actionable. Semi-urgent: Potential medications to cause serious harm based on PGx results. Expected time for completion of a semi-urgent eConsult is 48 hours from time of assignment. Clinically actionable: Potential medications to cause an adverse drug reaction or have reduced efficacy. Expected time for completion of a clinically actionable eConsult is 5 business days from time of assignment. A pharmacist eConsult was provided only when patients had current medications that were identified as either semi-urgent or clinically actionable. Results: To date, 1519 PGx results have been returned to the EHR. Out of these, 426 patients had an eConsult completed. One percent (5/426) of these eConsults were semi-urgent and 99% (421/426) were clinically actionable. The average time per eConsult was 22 minutes. The eConsult time ranged from 4 to 75 minutes. Seven pharmacists consisting of PGx and medication therapy management pharmacists were involved in the completion of 426 eConsults. A total of 9114 minutes (152 hours) of pharmacist time was spent over a 3 month period (February through May). Future Direction: Evaluate pharmacists' time involved in the completion of the eConsults to assist pharmacy leadership with future resource allocation for Mayo Enterprise PGx implementation. Assess future adoption of PGx by clinicians and pharmacists across Mayo Enterprise and provide necessary tools such as education to support successful implementation.

609T

Knowledge and attitudes on pharmacogenetics among pediatricians in America and Japan. S. Rahawi^{1,2}, H. Naik¹, K. Blake³, A. Owusu-Obeng⁴, R.M. Wasserman², Y. Seki^{1,2}, V.L. Funanage³, K. Oishi¹, S.A. Scott^{1,2}. 1) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Sema4, a Mount Sinai venture, Stamford, CT; 3) Nemours Children's Health System, Wilmington, DE, Jacksonville and Orlando, FL; 4) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai; and Pharmacy Department, The Mount Sinai Hospital, New York, NY.

Accumulating evidence supporting the clinical validity and/or utility of pharmacogenetic testing has prompted recent genotype-directed clinical practice recommendations for selected gene/drug pairs. However, common barriers to implementing pharmacogenetics into clinical practice include inconsistent evidence, inadequate reimbursement, and a lack of provider education. Previous studies have identified a general lack of professional knowledge on pharmacogenetics; however, this has not been tested among pediatricians. Importantly, ~17% of U.S. emergency department visits are due to pediatric adverse drug reactions, underscoring the potential utility of clinical pharmacogenetic testing for children. Consequently, we conducted an anonymous online survey to assess pediatrician knowledge of pharmacogenetics, preferences on educational tool design, and compared the responses of American (n=210) and Japanese (n=72) pediatricians. Over 50% of respondents had >10 years of experience and were based in teaching hospitals, and over 75% had prior undergraduate or graduate education in genetics. However, less than 10% felt they were 'very' or 'extremely' familiar with pharmacogenetics. This was further demonstrated by their lack of knowledge on codeine pharmacogenetics, as only 15% of providers correctly identified *CYP2D6* ultrarapid metabolizers as having an increased risk for respiratory depression and death following use of codeine. Moreover, the vast majority of respondents (93%) were previously unaware of the Clinical Pharmacogenetics Implementation Consortium (CPIC). Despite being unfamiliar with pharmacogenetics, over 75% of respondents indicated that clinical pharmacogenetics will be valuable and that pediatricians should be capable of applying this type of testing to their practice. In addition, respondents indicated their preference for educational tools, which included CME/CE courses, and seminars on interpretation and therapeutic recommendations. Interestingly, despite differences in federal medication regulatory agencies, results were similar across American and Japanese respondents. Although these results indicate that knowledge of pharmacogenetics among pediatricians is generally low, their interest in pharmacogenetic testing and related education is very high, which will likely facilitate future clinical implementation.

610F

Building quality assurance measures into an effective evidence based precision medicine program. A. Schultz, J. Baye, N. Petry, R. Wilke, J. Reiner, A. Aifaoui. Sanford Imagenetics, Sioux Falls, SD.

April Schultz, Pharm D; Jordan Baye Pharm D, MA, BCPS; Natasha Petry Pharm D, BCACP; Russ Wilke MD PhD FACP; Jennifer Reiner PhD; Aissa Aifaoui Pharm D Genomic medicine, including pharmacogenomics, has become one of the most innovative mechanisms utilized by healthcare professionals to provide optimal clinical care for their patients. However, successful execution of a precision medicine program requires a multidisciplinary approach to ensure continuity of care from the bench to the bedside. Sanford Health's pharmacogenomics program, Imagenetics, has created a team inclusive of medical doctors, molecular laboratory directors, laboratory compliance specialists, and pharmacists to evaluate and validate laboratory and pharmacogenomics data prior to clinical utilization. Herein, we describe a model for quality assurance in the practice of precision medicine. This strategy involves monitoring of observed allele frequencies with expected frequencies published in public databases, validation of conversions between genotype/diplotype and metabolizer status, and review of the corresponding drug dosing guidelines. Inclusion of these validated data in the patient's medical record permits programmed decision support within the electronic medical record (EMR), ensuring valid clinical recommendations throughout the patient's lifetime.

611T

Pharmacogenetic testing in primary care practice: How to speed up the transfer of knowledge? K. Tremblay¹, M.E. Blackburn², C. Dubois-Bouchard², M.P. Frigon¹, S. Tardif¹. 1) Université de Montréal / ECOGENE-21, Chicoutimi, Canada; 2) ECOBES Recherche et Transfert / Cégep de Jonquière, Canada.

Introduction: Some pharmacogenetic tests (PGt) are recommended by health authorities when prescribing drugs that can cause adverse events or be ineffective according to specific gene variants. However, medical professionals such as primary care physicians or pharmacists do not currently apply PGt testing, even if the possible advantages, such as improved therapeutic dosage and response, are generally well-recognized. **Objectives:** In this original research, objectives are 1) to understand how medical professionals and patients stand with regard to PGt and their potential application in clinical practice and 2) to make recommendations to improve the transfer of knowledge of PGt in clinical practice. **Methods:** We have conducted six focus groups in the Saguenay–Lac-Saint-Jean region (Quebec, Canada): two groups composed of general practitioners (GPs) (n=25), two groups composed of pharmacists (n=11) and two groups composed of patients (n=30). Part of the interview was based on real-life situations with examples of commonly prescribed drugs. In addition, we had asked all participants about their knowledge and perception of genetics in medical practice and what the roles of different professionals in regards to PGt could be. **Results:** The data highlighted a willingness in regard to PGt application in all interviewed groups. Questioned about their needs to make possible the implementation of PGt in their practice, GPs and pharmacists suggested to provide training, to develop software to link all patient's information, to present more evidence on the clinical utility and to clarify laws about confidentiality to avoid genetic discrimination by insurance companies. In addition, GPs and pharmacists have expressed the limitation of implementation of PGt in the healthcare system mainly because of a lack of information about the accessibility of these tests, the delay of obtaining results, cost-benefits studies and ethical considerations. The patients were generally in agreement with the PGt and trusted the medical professionals. **Conclusion:** We have been able to highlight a set of recommendations that could better help apply PGt knowledge in clinical practice. .

612F

Pharmacist process for providing personalized eConsults to support large scale pre-emptive pharmacogenomic (PGx) testing. J. Wright, E. Matey, J. Giri, J. Anderson, L. Oyen, N. Nicholas, H. Marker. Mayo Clinic, Rochester, MN.

Background: Pre-emptive pharmacogenomic (PGx) testing was performed on 10,000 Mayo Clinic Biobank participants. PGx results are in the process of being returned to patients' electronic health record. To support Primary Care Providers (PCPs) in adopting PGx results, pharmacists across the Mayo Clinic Enterprise were engaged to complete eConsults (electronic review of PGx results for each patient's current medications with reference to drug-gene interactions) for patients with drug-gene interactions. **Methods:** A train-the-trainer workflow model was designed to engage pharmacists across the institution whose daily workflow may not include performing eConsults. For this, PGx pharmacists were engaged whose role includes, but is not limited to, providing PGx guidance through eConsults and face-to-face consultation. This facilitated identification and resolve of issues that arose with the processes. Once the workflow was established, a team of "early adopter" pharmacists were trained. This group included Medication Therapy Management (MTM) pharmacists who had expressed interest in PGx and had experience in the eConsult process. MTM pharmacists' role consists of providing eConsults and face-to-face consultations. The "early adopters" collaborated with the PGx pharmacists to identify potential issues and further refine the workflow. The rollout will continue with the PGx pharmacists and 'early adopter' pharmacists training other MTM pharmacists. These aforementioned categories of pharmacists will then become trainers for the supervisors of inpatient and outpatient pharmacists across the institution. Supervisors will assign eConsults to their staff pharmacists and serve as a resource for PGx process questions. PGx, 'early adopter', and MTM pharmacists will continue to serve as trainers and a resource for PGx clinical questions to all pharmacists across the institution. A step-by-step guide was developed for pharmacists to standardize eConsults. The completed eConsults were electronically sent to the patient's respective PCP. **Future direction:** This train-the-trainer educational model will support adoption, dissemination, and utilization of PGx in clinical practice. Pharmacists across Mayo Clinic will be equipped to serve as a resource for PGx-related questions in partnership with other members of the care team to individualize patient care.

613T

Disclosure of reclassified VUS results of deceased patients to family members: Current practices. S. Lascrain¹, D. Thull², A. Durst¹, T. Bear¹, PL. Mai². 1) School of Public Health, University of Pittsburgh Pittsburgh, PA; 2) Magee Womens Hospital of UPMC Pittsburgh, PA.

Purpose: A variant of uncertain significance result (VUS) is not uncommon in genetic testing. There is currently little regulation or guidance regarding the responsibility of genetics professionals when a VUS is reclassified after a patient has passed away. This study was designed to explore current practices and obtain opinions of genetics professionals on the disclosure of reclassified VUS results of a deceased patient to their relatives. **Methods:** A 35-question web-based survey was distributed to the >3,000 members of the National Society of Genetic Counselors on the listserv. **Results:** A total of 154 individuals completed the survey, 94% of respondents were genetic counselors and 45% reported between one and five years of clinical practice. Forty-five respondents (24%) reported receiving reclassified VUS results in deceased patients. Respondents were more likely to always attempt disclosure of variants reclassified as pathogenic (83%) vs. benign (73%). Most respondents reported not being aware of state or local policies on the disclosure of a deceased patient's reclassified VUS results. Approximately half of the respondents were not sure if an institutional policy existed. Respondents rated the impact of the results on family members and a duty to warn as extremely important factors in considering disclosure to relatives. Most respondents favored a legal mechanism to allow disclosure to relatives. A minority of respondents felt a consent obtained from the decedent prior to their death should be legally (10%) or ethically (27%) required for the updated results to be disclosed to relatives. **Conclusions:** Guidance on the disclosure of reclassified VUS results for a deceased patient is not readily available, though respondents who have experienced this situation commonly reported attempting disclosure in clinical practice. Additional research regarding the role for policy change or development is needed.

614F

A strategic approach to institutional genomics excellence: The Center for Pediatric Genomics. J.M. Furgason, H. Aungst, A.M. Rupert, P. Dexheimer, N. Felicelli, J. Chen, J. Hendershot, S. Huitger, C. Van Horn, P.T Komarovalu, J.T. Whaley, J.B. Harley, L.J. Muglia, P.S. White. Cincinnati Children's Hospital Medical Center, Cincinnati, OH.

The Center for Pediatric Genomics (CpG) is a facilitation-based effort within Cincinnati Children's Hospital Medical Center (CCHMC) to create a vibrant and cohesive culture around precision genomic medicine. The Center utilizes a novel combination of grassroots efforts and more traditional mechanisms to engage clinicians, researchers, and the local community by recognizing, connecting, and educating stakeholders across institutional, social, and educational boundaries. The relatively wide disparity in genomics knowledge among populations of clinicians, researchers, and patients means that finding common areas to interact with these groups can be difficult. CpG's approach examines sources of uncertainty around genomics and tailors individualized programs that generate confidence, competency, and proficiency. This approach focuses on three main pillars: constituent-centric literacy programs, collaborative genomic science, and a comprehensive precision genomics data strategy. Academic research hospital settings require literacy programs to engage a diverse group of caregivers, clinician scientists, and both basic and translational genomics researchers, each of which has varied historical exposure to genomics. CpG uses a unique engagement strategy that relies on stakeholders to inform an educational curriculum that focuses on topics such as genomic technology, testing indications, report interpretation, and clinically relevant knowledge gaps. In 2017, CpG established the Midwest's largest annual symposium for genomic medicine (Precision Genomics Midwest). For patients and the local community, CpG seeks to be a resource that makes precision medicine research transparent and accessible, revealing the power their participation has to shape and improve healthcare. For academic excellence, CpG has invested nearly \$5M to fund 48 peer-reviewed pilot projects that has incentivized over 400 CCHMC investigators to form teams of clinicians and genomic researchers to take on challenges in translational genomics. CpG also supports essential infrastructure for precision medicine, maintaining a "precision genomics ecosystem" for the storage, analysis and sharing of genomics data that is linked to over 1 million pediatric patient records and 800,000 biosamples across 120 projects. It is supported by a bioinformatics team that engages investigators across the institution to guide development of resources that satisfy the needs of both technical and non-technical users alike.

615T

Test2Learn: Leveraging personal genomic testing to advance primary care genomics education. *M. Massart¹, P. Empey².* 1) Family Medicine, University of Pittsburgh, Pittsburgh, PA; 2) School of Pharmacy, University of Pittsburgh, Pittsburgh, PA.

Precision medicine, is rapidly transforming healthcare but has outpaced medical education. Many barriers to the integration of genomics into primary care have been outlined with large emphasis on knowledge deficiency (3, 4). Practitioners have limited exposure to genetics testing, and there has been slow integration of genomics education into medical school, residency programs and continuing medical education curricula (1, 2, 5-16). We have tested the effectiveness of a participatory model (Test2Learn) as a highly effective method of providing essential genomic competencies for primary care practitioners. Specifically, we have created an innovative and broadly scalable online genomics education program designed to achieve the competencies outlined by the Genetics-Genomics Competency Center (G2C2) for physicians and the American Academy of Family Practice recommended competencies in Medical Genetics. The Test2Learn program is a participatory genomic testing (PGT) model previously developed at the University of Pittsburgh School of Pharmacy (17). This platform uses 23andMe kits for optional PGT and includes (A) ethics content; (B) asynchronous video content as an out of the classroom genetics review, (C) new in-person flipped-classroom lectures; and (D) new exercises leveraging the learners' experience with PGT and available genetic data. Since primary care providers practice a broad scope of medicine including prenatal genetics (carrier status), wellness, genetic cancer risk assessment, and pharmacogenomics, we have demonstrated that the Test2Learn education model can be extended to enhance the learning for Family Medicine residents by including these content areas. We present the Test2Learn educational platform as a method to provide precision medicine education to primary care providers as well as the results assessing its effectiveness in engaging learners, improving genomics knowledge, and the skills to integrate precision medicine into patient care. 1. Plunkett-Rondeau J et al., 2015. 2. Selkirk CG, et al., 2013. 3. Mikat-Stevens et al., 2015. 4. Suther S, Goodson P., 2003. 5. Baars MJ et al., Genet 2005. 6. Ginsburg GS, 2008. 7. Haga SB et al., 2013. 8. Houwink EJ et al., 2011 9. Rinke ML et al., 2014. 10. Weitzel KW et al., 2016. 11. Burke S, 2002. 12. Burke W 2002. 13. Cichon M, 2014. 14. McCarthy JJ, 2013. 15. Nelson EA, 2010. 16. Telner DE, 2008. 17. Adams SM, et al., 2016.

616F

Increased comprehension of a consent chatbot for research exome sequencing via rapid iteration. *T.J. Schmidlen¹, M. Snir², G. Snir², E. Simmons², H.L. Kirchner¹, D.H. Ledbetter¹, A.C. Sturm¹.* 1) Geisinger, Danville, PA; 2) Clear Genetics, Inc. San Francisco, CA.

Chatbots are computer programs that conduct conversations with humans. Geisinger's MyCode® Community Health Initiative partnered with Clear Genetics, Inc. to develop a consent chatbot for a large research biobank returning clinically actionable exome sequencing results. The MyCode® consent chatbot walks patients through components of the consent allowing them to opt to receive more or less detail on key topics (goals, benefits, risks, etc.). The platform uses conditional expressions, natural language processing, machine learning, and search indexing to reply to user questions. A link to the chatbot can be sent via MyGeisinger patient electronic health record (EHR) portal, via text message, or by email. It runs on any web browser and works on smartphones, tablets and computers. The chatbot is designed to record the decision (consent, considering, decline) in the patient's EHR. Comprehension of key concepts conveyed by the chatbot was collected via a 14-item multiple choice quiz. A link to the chatbot and comprehension quiz were posted to Amazon's Mechanical Turk with parameters restricting participant demographics to mirror age, income and education level of Geisinger patients. After one round of testing with 101 users (Group A), results were reviewed and the chatbot was updated to address questions with average correct scores of ≤75%. A second group of 102 users (Group B) tested the updated chatbot and comprehension results were compared to Group A. Total average comprehension score for all 14 items was high (80% Group A, 85% Group B). Group A scored <75% correct on 6 questions on the type and timing of results returned, purpose of the chatbot, and access to health records. A summary recap was added at the end of the chatbot to address these topics. Group B reviewed the chatbot with the summary recap and per-question average scores improved to >75% for all but one question. Updating the chatbot to include a summary recap resulted in significantly improved comprehension ($p=0.003$) of key elements of the MyCode® consent. The MyCode® consent chatbot presents an engaging, effective alternative to deliver content that can be challenging for patients to comprehend in traditional paper or in-person consent interactions. The quantitative nature of the chatbot platform allows for rapid learning and successful iteration to improve comprehension. Further patient engagement through focus groups, online surveys, and remote, unmoderated user testing is underway.

617T

Ethical review processes for bioinformatics research using human genome data: Analysis of specific requirements in the changing environment of personal data protection regulations. *M. Kokado, H. Okui, K. Kato.* Graduate School of Medicine, Osaka University, Suita, Osaka, Japan.

With dramatic advances in genomic research technology, especially the development of global database frameworks and information communication technology (ICT), the bioinformatics or statistical genomics research are making rapid progress. With respect to data sharing ELSI, concerns about privacy and security are emphasized and various devices are under consideration (e.g. S. Wang et al, 2017). In Japan, it is stipulated that the human genome research will be conducted in accordance with human genome research guidelines which have been established by the government. What the guidelines consider is the human genome research that starts with the collection of samples such as tissue or blood. However, there are increasing number of research projects using human genome data that are stored in the genomic databases such as National Bioscience Database Center (NBDC). At Osaka University, among the approved projects by the University ethics review committee for human genome research, 11 ongoing protocols are of such type and 6 of them have been approved in 2017. It is an urgent matter to clarify elements necessary for IRB review of such research and to establish a better research ethics review process. At the same time, regulations for personal data protection have been revised in many parts of the world including EU and Japan. The Japanese legislation, the APPI (Act on the Protection of Personal Information), has been revised and came into force in May 2017. In this paper, we will describe the current discussion about ethical considerations in human genomic data sharing and ethics review requirements for the human genome informatics research. We have reviewed procedures of various ethics review committees in Japan by consulting some human genome informatics researchers and IRB staffs. We have found that the increase in the bioinformatics research or statistical genomics research has been giving difficulties for ethics review committee members, because the changing regulations for personal data protection are very specialized. We have also found that there is a need in the committee to establish a standardized procedure for reviewing these bioinformatics projects.

618F

Frequency, characteristics, and impact of discordant lab to clinician genetic test result interpretation in a single pediatric medical center. *C. Berrios^{1,5}, C. Saunders^{1,2,5}, J. Thiffault^{1,2,5}, L. Willig^{1,3,5}, E. Farrow^{1,4,5}.* 1) Center for Pediatric Genomic Medicine, Children's Mercy, Kansas City, MO; 2) Department of Pathology and Laboratory Medicine, Children's Mercy, Kansas City, MO; 3) Department of Nephrology, Children's Mercy, Kansas City, MO; 4) Department of Pediatrics, Children's Mercy, Kansas City, MO; 5) School of Medicine, University of Missouri Kansas City, Kansas City, MO.

Interpretation of genetic variation remains a significant challenge to the field of clinical genetics. Efforts have been made to standardize lab interpretations, but ultimately clinicians returning results have the final responsibility for interpretation of genetic test results. A first step in learning how to best leverage both lab and clinician knowledge to improve variant interpretation and clinical care is determining the scope and factors behind discordant interpretations between labs and clinicians. This study seeks to determine the frequency, directionality, test characteristics, and clinical impact of discordant lab to clinician genetic test result interpretations in a single pediatric medical center. From the institutional molecular genetics lab database all sequencing tests ordered between April 2016 and March 2017 were pulled. Separate lists were compiled for those ordered by a medical geneticist and by other specialties. Each list was randomized and tests were sequentially pulled from the randomized list for chart review. Data collected included provider specialty, lab and clinician classification of each reported variant and overall result interpretation, and medical decisions based on the result. Those with zero variants reported or no documentation found of clinician interpretation were removed. To date, 52 tests ordered by non-genetic specialists have been reviewed, 7 were removed for zero variants and 8 for lack of documentation. Of the 37 remaining for analysis, a non-diagnostic or indeterminate lab result was interpreted by the clinician as diagnostic in 3 cases and a diagnostic lab result was interpreted by the clinician as non-diagnostic in 2 cases, for a total of 13.5% discordance in overall result interpretation. Discordant interpretation occurred in 15.4% of send out and 12.5% of internal tests reviewed, and in 14.2% of panels and 9.1% of exomes. Each affected clinical care; impacting treatment recommendations (2), screening tests (1), and recurrence risks (5). The data suggests that discordance between lab and clinician genetic test result interpretation is not uncommon, but larger studies from varied institutions are needed to better understand the frequency and impact. Studies on the reasons for discordance will allow the development of strategies to leverage each group's knowledge to improve interpretation, ensure appropriate medical care, and avoid patient confusion.

619T

Assessing the genetic literacy of the general population: A literature review. *N. Akiyama¹, M. Nishigaki².* 1) Center for Medical Genetics, Chiba Children's Hospital, Chiba, Japan; 2) Human Health Sciences, Graduate School of Medicine, Kyoto University, Kyoto, Japan.

[Background] Owing to the recent developments in genome analysis technology, genetic testing has become broadly available. Understanding its meanings and related ethical issues would enable the general population to make informed decisions, after obtaining sufficient information. However, the minimum genetic/genomic literacy level that is required in the general population has not been defined clearly. [Aim of the study] We aimed to conceptualize contents of genetic literacy (genetic/genomic literacy) and their evaluation methods in the general population. [Method] We gathered and examined reports that measured the genetic/genomic literacy of general citizens, high school students, and university students over the past 20 years; the literature search was carried out using the following four databases: PubMed, Web of science, Scopus, and ICHUSI (Japanese database for medical literatures). The abstracts of the paper, obtained as a result of the search, were screened and thoroughly reviewed. [Result] We obtained 36 articles that served our purpose. These were largely grouped into those targeting the general population and those targeting the high-risk group of individuals with specific diseases. Genetic/genomic literacy was defined as the ability to perform the following tasks and measured: (1) understand genetic terms/phenomena, (2) use genetic knowledge, and (3) recognize genetics as a familiar event. [Discussion] Genetic/genomic literacy should enable the general population to: (1) understand the terms and phenomena related to heredity and genetic testing, (2) think about the expected results of the test, and (3) explore and discuss related ethical, legal and social issues. Initiatives to improve genetic/genomic literacy should focus on not only increasing genetic knowledge, but also providing opportunities to understand how new technologies can affect our lives actually.

620F

Are we understanding each other? Early lessons from recruiting non-English speaking families for a study investigating the utility of WES in underserved populations. *B. Anguiano, S. Ackerman, J. Harris-Wai, H. Hoban, B. Lianoglou, M. Norstad, S. Outram, N. Rangwala, S. Rego, N. Sahin-Hodoglugil, B. Koenig.* UCSF, San Francisco, CA.

The Clinical Sequencing Evidence-Generating Research network is a national NIH-funded consortium that investigates the utility of genomic sequencing in diverse patient populations. The project encompasses six clinical sites, including the University of California, San Francisco, whose Program in Prenatal and Pediatric Genomic Sequencing (P3EGS) studies the clinical and personal utility of whole exome sequencing for children and pregnant women with suspected genetic conditions. Many prospective participants are monolingual Spanish speakers and the majority of clinical and research staff are not fluent in Spanish, necessitating a reliance on in-person or phone-based medical interpreters. The cultural and linguistic differences that characterize these encounters present unique challenges to recruitment in clinical settings. We elucidate the challenges encountered by the P3EGS in recruiting monolingual Spanish-speakers and outline initial strategies used to improve communication. Challenges include interpreters' lack of knowledge of genetic terminology, additional time required for an interpreted conversation with families, and increased likelihood of misapprehension when discussing complex topics such as insurance coverage, results interpretation, uncertainty, secondary findings, data sharing, and genetic discrimination. Strategies used to help overcome these barriers include a) a bilingual member of the team working directly with patients and families to assess the clarity and cultural appropriateness of professional interpretation and translated study materials, and b) observations of bilingual study enrollment sessions to assess the impact of professional medical interpretation on families' understandings of the research project and decisions about whether to participate. Insights gained were used to develop and refine English and Spanish-language study brochures tailored to prenatal and pediatric clinical settings, including reducing technical jargon and translating specific words to be comprehensible to a wide range of Spanish-speaking families. Preliminary findings suggest that more formal stakeholder engagement efforts with monolingual Spanish-speakers could be of benefit to the study. Further research is also needed to assess the experiences of non-English speaking participants to ensure they receive comprehensible information for complex clinical tests such as genomic sequencing.

621T

High throughput counseling: A model for genetic testing results disclosure and patient management. A. Arjunan, R. Ben-Shachar, J. Kostialik, G. Lizarin, K. Johansen Taber. Counsyl, South San Francisco, CA.

Objective: As the availability, uptake, and complexity of genetic testing expands in the general obstetrics setting, primary care providers increasingly must offer genetic services. Non-invasive prenatal screening (NIPS) and expanded carrier screening (ECS) have dramatically widened the scope of information available to patients and providers alike, underscoring the importance of timely results delivery. We describe an automated system that utilizes online education and tele-genetic counseling to aid women's health professionals in effectively delivering genetic testing results. **Methods:** Providers were notified upon availability of patient results. For ECS results indicating a high reproductive risk, patients received an automated email directing them to schedule a consultation with a laboratory-based genetic counselor. Screen-positive NIPS results were delivered to the ordering provider, who could request that a laboratory-based genetic counselor contact the patient. All other results could be viewed directly by the patient in an online portal and were accompanied by written and video resources. Regardless of result type, patients had the option to speak to a genetic counselor, either on-demand or by scheduling an appointment. **Results:** Through December 2017, 278,318 ECS results and 67,122 NIPS results were issued via the system. Of this cohort, 41,050 (15%) ECS patients and 4,673 (7%) NIPS patients elected genetic counseling, resulting in a total of 47,793 consultations. Interestingly, many of the genetic consultations (50% for ECS and 95% for NIPS) were requested by patients who received negative results. Regardless of result type, patients frequently utilized on-demand services (32% of ECS consultations and 70% of NIPS consultations). Median consultation duration was ten minutes and nine minutes for ECS and NIPS results, respectively. The mean patient satisfaction rating for consultations was 4.9/5.0. **Conclusion:** The automated system provides large-scale prenatal genetic testing results delivery, education, and counseling consistent with clinical guidelines. These services are imperative to quality clinical care as genetic testing uptake grows among the general obstetric population.

622F

Precision medicine & genomic education for clinicians: Effectiveness of a novel online CME program. E. Edelman¹, P. Tanner¹, L. Nicholson², S. McConnell³, T. Ingram¹, L. Steinmark¹, K. Johansen Taber⁴, K. Reed¹. 1) Genomic Education, The Jackson Laboratory, Bar Harbor, ME & Farmington, CT; 2) Scripps Translational Science Institute, La Jolla, CA; 3) American Medical Association, Chicago, IL; 4) Counsyl.

INTRODUCTION A number of challenges limit the integration of genomic technologies into patient care, one of which is the gap in genetics knowledge and skills among healthcare providers (HCPs). Indeed, HCPs have called for education to help navigate genomic testing decisions. While some education resources are available covering various genomic topics, evidence of the effectiveness or quality of such programs is limited. **OBJECTIVE** This study describes the development and evaluation of a novel genomics CME program. **METHODS** The Jackson Laboratory, American Medical Association, and Scripps Translational Science Institute developed "Precision Medicine for Your Practice," a 12-module, interactive online program that aims to help HCPs appropriately incorporate emerging genomic tests into patient care. The program framework was designed using adult education best practices such as patient-centered cases, skill-based practice, self-directed learning, expert feedback, and scaffolded instruction. Content was developed with input from target audience members and genetic experts. Modules were developed and disseminated on a biannual basis starting in 2016. The first five modules address: expanded carrier screening (M1), prenatal cell-free DNA (cfDNA) screening (M2), exploring somatic cancer panel testing (M3), interpreting somatic cancer panel results (M4), and genomic testing for the healthy patient (M5). M6-12 are in development and address diagnostic exome testing, pharmacogenomics, cardiogenomics, and neurogenomics. **RESULTS** As of 3/2018, 689 users enrolled in at least one module. M1 had 194 users, M2 142, M3 288, M4 160, and M5 136. 37% of users were physicians, 11% masters- and 7% bachelors-level nurses, 6% physician assistants, 6% genetic counselors, and 5% scientists. Among users who completed CME evaluations (32%), the average rating for the program was 4.5/5. After taking the program, 94% believed they were able to identify appropriate candidates for tests and interpret results for patient management. **CONCLUSION** Our approach to developing Precision Medicine for Your Practice was effective in helping HCPs of diverse backgrounds build skills around genomic testing. This program supports HCPs using genomic testing to identify preventative measures and treatments for appropriate patients, which can lead to improved patient outcomes. These findings can be used to further disseminate existing modules and also serve as a needs assessment for future education efforts.

623T

How is meiosis taught in high school? A critical gap for students. *D.L. Newman¹, G.E.C. Dy^{1,2}, L.K. Wright¹*. 1) School of Life Sciences, Rochester Institute of Technology, Rochester, NY; 2) Department of Biology, Washington University, Seattle, WA.

Understanding meiosis is essential for students, researchers, and practitioners in the field of genetics. While instruction on meiosis often begins in the middle school years, many students continue to struggle with these concepts throughout their undergraduate training. The DNA triangle framework (Wright et al, 2018), which links the chromosomal, molecular, and informational levels of DNA, can be used to conceptualize three important facets of meiosis: ploidy, homology, and homologous pairing. Previous work demonstrated that experts (university faculty) are able to appropriately integrate the molecular and informational levels of DNA to explain homology, integrate the chromosomal and informational levels to explain ploidy, and integrate the chromosomal and molecular levels to explain homologous pairing. Undergraduates, however, struggle to integrate multiple levels of DNA knowledge, especially at the molecular level. To better understand why this gap in molecular knowledge is so widespread, we interviewed 14 high school Biology teachers (experience ranging from 2 to 26 years) about their own mental models and their approaches to teaching meiosis. Perhaps not surprisingly, teachers held many of the same misconceptions documented in undergraduate biology majors. Analyzing the interview data through the lens of the DNA triangle, we found that the majority of teachers (11 of 14) did not incorporate *any* molecular-level knowledge into their own explanations of meiosis, and *none* emphasized DNA structure during meiosis instruction. Homology, for example, was explained at the chromosomal (same size and shape) and informational (same genes) levels, but not the molecular level (nearly the same sequence of DNA nucleotides). Analysis of 20 lesson plans about meiosis yielded similar results: instruction focused on the chromosomal and/or informational levels but *never* on the molecular level. During interviews, several teachers explicitly acknowledged the separation between instruction about DNA (i.e. structure and complementary base-pairing) from lessons on meiosis. While this work suggests a critical gap in the high school biology curriculum, it also highlights an opportunity to offer new professional development opportunities for high school teachers and curriculum developers. These results also partially explain why many undergraduate biology students have difficulty thinking about complex processes like meiosis at the molecular level.

624F

A teaching vignette: Our experience incorporating personalized pharmacogenetic testing into the PharmD curriculum. *S.S. Oh¹, R. Ta², J. Sims², J. Tinajero², A. Liu², T. Desai³, B.J. Guglielmo², S. Guo³, A.H.B. Wu⁴, E.G. Burchard^{1,2}*. 1) Medicine, UC San Francisco, San Francisco, CA; 2) School of Pharmacy, UC San Francisco, San Francisco, CA; 3) Bioengineering and Therapeutic Sciences, UC San Francisco, San Francisco, CA; 4) Laboratory Medicine, UC San Francisco, San Francisco, CA.

The popularity of direct-to-consumer genetic testing products and efforts to incorporate precision medicine in clinical practice place additional educational demands on providers. The utility of genomic data and their use in daily clinical practice depend on providers' knowledge and ability to interpret the data. However, the low prevalence of pharmacogenomics-related education in professional school curricula may hinder the acceptance of precision medicine in clinical practice. We therefore began to enhance pharmacogenetic (PGx) education by making it more personal. All first-year pharmacy students at the University of California, San Francisco are required to take Biopharmaceutical Sciences (BPS) 115 ("Genetics and Pharmacogenetics"). We piloted personalized PGx testing in 2013, when 22 BPS 115 students volunteered to be genotyped for pharmacogenomic variants in *CYP2C19*, a gene critical for drug metabolism. Course instructors devoted a "Reveal Day" lecture to review volunteers' allele distributions and the clinical implications of *CYP2C19* variants on drug metabolism. A qualitative evaluation of this pilot showed that students unanimously valued personalized PGx testing and enthusiastically encouraged its use as teaching material in future courses. We have since incorporated personalized PGx testing into BPS 115, and students' voluntary and anonymous participation in genotyping has increased. In 2017, we used 23andMe kits after receiving a grant to fund enough kits for the entire first-year class. Students are instructed on how to abstract their raw genetic data to interrogate PGx-related genes. The genome-wide SNP data allow us to test substantially more PGx genes and examine the influence of genetic ancestry on PGx. A major Reveal Day component now includes student-led activities such as role-playing patient/pharmacist and pharmacist/provider relationships using PGx test results. The role-plays have evolved in complexity and allow students to incorporate skills they are learning concurrently (e.g., motivational interviewing and navigating the subtleties of inter-professional relationships). We have shown that incorporating personalized PGx testing into the first-year PharmD curriculum leads to significant increases in knowledge (+1.0 Likert points) and more favorable attitudes (+0.3 Likert points) toward PGx and precision medicine. Our experience supports the use of personalized PGx testing as a teaching tool.

625T

Review of the current status of clinical practice guidelines in Japan for returning individual research results on genetic diseases. Y. Aizawa, F. Nagami. Tohoku Medical Megabank Organization, Tohoku University, Japan.

We explored clinical practice guidelines for genetic diseases in Japan and studied the genomic information selection criteria for the return of individual research results. In 2013, the American College of Medical Genetics and Genomics (ACMG) had proposed a list of specific genes underlying medically actionable diseases for sequencing in clinical settings as secondary findings; the list has since been updated. Since the list was tailored for the United States, it needs to be carefully examined for applicability in other countries. In case researchers wish to return the genomic results to participants, it would be necessary to choose returnable genomic information in the research project. In Japan, we are on the path to develop a practical guideline regarding return of genomic results in research settings. Toward that, we compared the description of genetic diseases in Japanese clinical practice guidelines with that in the 59-gene list proposed by ACMG. We categorized the 59 genes into 32 diseases, and searched for published guidelines on the web to examine the description regarding each genetic disease, including clinical characteristics, diagnosis/genetic testing, and management. We could find at least one guideline for all diseases. These guidelines were developed during 2010 to 2018. Guidelines regarding half of the diseases provided comprehensive description, but that for few others mentioned only the clinical characteristics. Approximately 90 % of the diseases had at least one guideline mentioned about genetic testing; 80 % of the diseases had at least one guideline mentioned about genetic counseling. Approximately half of the diseases had at least one guideline mentioned regarding relatives at risk or pre-symptomatic individuals. There was a variety of descriptions about diseases, since their prevalence varied greatly. Taken together, our findings suggested that it is important to gather updated information regarding each genetic disease, in addition to understanding the basic information from clinical practice guidelines. Therefore, we propose the need to consider genomic information selection criteria with respect to the health system of each country, and recommend additional discussion regarding the return of genomic research results for developing a practical guideline in Japan.

626F

Direct-to-consumer genetic testing for factor V Leiden and prothrombin 20210G>A: the consumer experience. S.L. Elson¹, N.A. Furlotte¹, B.S. Hromatka¹, C.H. Wilson¹, J.L. Mountain¹, H.M. Rowbotham¹, E.A. Varga², U. Francke^{1,3}. 1) 23andMe, Inc., CA; 2) Nationwide Children's Hospital, Columbus, Ohio, USA; 3) Stanford University Department of Genetics, Stanford, California, USA.

Background: Variants in clotting factor genes are responsible for inherited predisposition to venous thromboembolism (VTE), a condition characterized by formation of blood clots in veins. The two most common genetic risk factors for VTE are factor V Leiden in the *F5* gene, and prothrombin 20210G>A in the *F2* gene. There is limited consensus about which individuals should be tested for these two variants to inform treatment and prevention decisions. Given a gap between clinical guidelines for testing and consumer and public health organizations' interest in greater awareness of the condition, we examined the impact of direct-to-consumer (DTC) testing for VTE risk. **Methods:** 2114 customers (1136 mutation-positive cases and 978 mutation-negative controls) of the personal genetics company 23andMe, Inc., who had received results online for *F5* and *F2* variants prior to November 2013 and had consented to participate in research, participated in a survey-based study to assess sharing of test results with health care providers and family members, lifestyle changes based on testing, and emotional responses to testing. **Results:** Most participants (approximately 90% of cases and 99% of controls) had not previously had testing for *F5* or *F2* variants. Cases were more likely than controls to have a personal history of VTE, and most demonstrated understanding, based on their 23andMe results, that they were at higher than average risk for developing a VTE. More cases than controls reported (1) sharing their results with family and health care providers (35% versus 5%), (2) receiving lifestyle recommendations (31% versus 14%), and (3) making behavioral changes (e.g., taking steps to lose weight [25% versus 10%]; exercising more [30% versus 12%]) after receiving results. Both cases and controls were satisfied knowing their genetic probability for VTE (81% and 67%), and nearly all participants said they would choose to learn their genetic risk for VTE if given the option again. **Conclusion:** Individuals who received positive results for *F5* or *F2* variants were satisfied with having this knowledge and report sharing results with others and making positive behavioral and lifestyle changes.

627T

Engaging diverse communities with facilitated deliberation: Stakeholder perspectives from the Alabama Genomic Health Initiative. *S.J. Knight¹, G.S. Barsh², K.M. Bowling², A. Cannon¹, J.J. Cimino¹, G.M. Cooper², W.A. Curry¹, K. East², J. Edberg¹, M. Fouad¹, A.C. Hurst¹, M. Might¹, T. May², I.P. Moss¹, M. Nakano¹, J.H. Schach¹, B.M. Shaw¹, S.O. Sodeke³, S. Hardy¹, E. Miskell¹, J.L. Richards¹, B.R. Korff¹.* 1) University of Alabama at Birmingham, Birmingham, AL; 2) HudsonAlpha Institute for Biotechnology, Huntsville, AL; 3) Tuskegee University, Tuskegee, AL.

The Alabama Genomic Health Initiative (AGHI) is a population-based program designed to return results of medically-actionable genomic variants to 10,000 adult volunteers residing in all 67 counties of the state. Engaging urban and rural Alabama residents in genomic research has required a multi-level strategy to understand perspectives from community members, community and state leaders, clinicians, and healthcare system executives and managers. We are employing a novel approach, facilitated deliberative process meetings, to understand the wide range of expectations, needs, and priorities of stakeholders from Alabama communities. We have convened two meetings to date, one in northern and the other in central Alabama. Stakeholders attended along with AGHI researchers. The facilitated deliberative process included education and opportunities to ask questions of AGHI scientists and clinicians. We invited participants to offer recommendations for AGHI action on two topics: 1) Challenges of recruitment of underrepresented populations in biomedical research and 2) Value of return of genome sequencing results. Participants voted to establish the priority of action items. A total of 67 stakeholders and researchers attended the two 2.5 hour meetings. There was greater representation of whites (69%) at the northern location and African Americans (83%) at the central site. Despite initial concern about trust and privacy, the discussions indicated a strong interest in AGHI and receiving results from genome sequencing. Suggestions to expand enrollment in rural regions included offering "pop-up" clinics in remote areas, partnering with local businesses and schools, and using existing community health models and resources for outreach. Both groups indicated that waiting to receive results might be a time of anticipation and anxiety for participants. Stakeholder priorities were rapid return of results, expectations about timelines, support during the waiting period, and attention to literacy in informational materials. One group emphasized the importance of ensuring that participants understand the implications of negative results and both groups stressed sensitivity to participant needs and the setting when providing information. Facilitated deliberative process provides a way to understand perspectives on genomic research held by diverse communities. These meetings have informed AGHI recruitment and return of results, and ultimately, will inform dissemination of findings.

628F

ELSI research on the use of electronic methods for participant engagement in medical research: Analysis from the perspective of ethical principles. *A. Kogetsu, K. Kato.* Department of Biomedical Ethics and Public Policy, Graduate School of Medicine, Osaka University, Suita, Japan.

[Background] Recently, with the increase in the amount of data and the development and dissemination of ICT, electronic methods utilizing online systems are expected to be used for medical research. Electronic methods can be used for processes especially of the participant engagement in medical research. In addition, new projects which depend on electronic methods such as medical research using wearable devices or medical big data that accumulate in the medical facilities are also under development. However, it is not clear what kind of problems could arise when electronic methods are implemented for medical research. **[Objective]** To identify the benefits and problems associated with electronic methods for participant engagement in medical research and to examine its theoretical background. **[Method]** We divided participant engagement into four steps (i.e. recruitment, informed consent, provision of samples and data, and return of research results) and identified following points based on the eight principles of ethical clinical research¹⁾: 1) what kind of ELSI would be solved (i.e. benefit brought by electronic methods) and 2) what are the ELSI to be addressed when electronic methods are implemented? **[Result]** Electronic methods can be useful for solving following ELSI: temporal and spatial constraints of research participation, return of research results to participants, informed consent on secondary usage of data, guarantee of the right for participants to refuse to participate or to withdraw their consent, and establishing collaborative partnership. However, the major points of ELSI that should be addressed when using electronic methods are: 1) digital divide for those unfamiliar with ICT and bias associated with it, which are revealed by the principles of social value and fair participant selection, 2) privacy protection and transparency of data which are based on the principle of respect for participants, and 3) confirmation of understanding of participants and appropriate presentation of information which are associated with the principle of informed consent. **[Discussion]** Although electronic methods can give effective means to exiting ELSI in the current medical research, it will generate other ELSI. Addressing such issues appropriately will lead to better medical research. **[References]** [1] Emanuel, E. J., Wendler, D., & Grady, C. (2008). An ethical framework for biomedical research. The Oxford textbook of clinical research ethics, 123-135.

629T

TAG: A rapid assessment tool of genetic literacy. *H. Milo Rasouly, N. Cuneo, J.J. Thompson, M. Marasa, J. Wynn, M. Verbitsky, G. Hripacsak, C. Weng, P. Appelbaum, W.K. Chung, S.R. Bakken, A.G. Gharavi.* Columbia University, New York, NY.

BACKGROUND: As genetic testing is being offered more frequently, general public awareness of genetics' concepts is desirable. Knowledge of an individual's baseline genetic literacy (GL) may facilitate the informed consent and the return of results process by tailoring sessions to each individual's understanding and expectations. Currently, there are no rapid, self-administered, validated tests to assess GL. **METHODS:** As part of the eMERGE consortium (Electronic MEDical Records and GENomics), we recruited a diverse cohort of 1132 individuals for genetic testing for a targeted panel of 68 medically actionable genes. The informed consent process was administered by a team of trained clinical research coordinators, and all participants were asked to answer a 56-item questionnaire, including questions about education, health literacy, numeracy and the Terms Associated with Genetics test (TAG), a new GL test developed for this study. TAG is a term-recognition quiz informed by the METER (Medical Term Recognition Test) health literacy format, which includes genetic-specific words and pseudo-words, and can be completed in under two minutes. Its content validity was established through expert review. **RESULTS:** 69% of the participants completed the questionnaire. TAG construct validity was supported by the known groups differences (education, health literacy and self-reported genetic knowledge), and concurrent measures techniques (four genetic knowledge questions). TAG identified 262 (34%) individuals as having low GL. Independent of education level, health literacy, and numeracy, low GL was significantly associated with lower satisfaction from the informed consent process ($p < 0.005$), unrealistic expectations about genetic testing ($p < 0.005$), and limited understanding of the legal protections against genetic discrimination ($p < 10^{-6}$). Ten common genetic terms that providers may use in interactions with patients, including *actionability*, *pathogenicity* and *penetrance*, were not recognized as real words by the majority of the participants. **CONCLUSION:** The TAG test provides a simple method to identify individuals with low GL. We showed that low GL compromises the informed consent process and is associated with unrealistic expectations about genetic testing and lack of awareness of the limitations of legal protections. With the benefit of being a rapid and self-administered evaluation, it can inform future studies of GL and implementation of genetic testing in the clinic.

630F

Ancestry isn't the past, it's the future: Human genetics research and race. *V. Moreno.* Anthropology, San Francisco State University, San Francisco, CA.

This pilot study seeks to understand why scientists continue to use race as a proxy in their scientific research, and to explore the other forms and techniques geneticists use to study or classify race. Interviews were conducted with ten geneticists from the University of California, San Francisco to solicit their views on the scientific validity of race, the ethical implications of classifying race, and the emerging significance of ancestry in genetic science. As these interviews revealed, the requirements to receive funding from federal institutions influence scientists' use of racial classifications. To ensure diversity and follow OMB Directive 15, many scientists report their findings using the racial classifications described in the directive. However, a majority of the participants opt for using "ethnicity" or "populations" as a term for the genetic groups they study in their published research in order to present their findings in a more socially conscience approach. Options for geographic origin of ancestors along with selected race and ethnicity are included in the self-identification process of research participants, since scientists wish to utilize this information subsequently in their scientific research. Grouping large populations according to similar genetic ancestry is performed using the technique of Ancestry Informative Markers (AIMs), a more accurate form of categorization and replacement of referring to racial categories. Thus, AIMs and the transition to increasingly descriptive self-identification reporting demonstrates a progressive approach to studying race; though, socioeconomic, cultural, and environmental factors should receive more attention for their effects on health in human genetics research.

631T

Survey on the perception of germline genome editing among the general public in Japan. *K. Muto^{1,2}, A. Ngai¹, M. Uchiyama².* 1) The Institute of Medical Science, The University of Tokyo, Tokyo, Japan; 2) Graduate School of Frontier Sciences, The University of Tokyo, Chiba, Japan.

Genome editing of human embryos could become a fundamental treatment approach for genetic diseases; however, technical and ethical issues need to be resolved before its application in clinical settings. Presently, the Japanese government has issued a statement prohibiting human germline editing and emphasizing the need for discussions that include a wide range of perspectives. However, we need to consider how to involve the general public in discussions. Therefore, we conducted online surveys independently in 2017 (N=10,881) and 2018 (N=10,739) in Japan who indicated that their disease conditions are related to their genetic makeup, and clarified their attitude toward this technology. The results clearly indicated that the Japanese people generally accepted the use of genome editing for disease-related genes, and this trend was higher in 2018. However, many were concerned about the risks. In addition, many Japanese people did not understand the technology well. To improve awareness and understanding about genome editing, it is important that scientists and science communicators create opportunities for the public to participate in relevant discussions without harming vulnerable participants. It is also important to continuously track changes in the acceptance of genome editing by the public.

632F

Return of genomic results in the research context – review of the global situation. *F. Nagami, Y. Aizawa.* Tohoku Medical Megabank Organization, Tohoku University, Sendai, Miyagi, Japan.

In the past decade, the appearance of next-generation sequencing (NGS) has drastically reduced the time and cost of sequencing individual genomes. Many cohort or biobank studies perform genomic analyses, including whole-genome and exome sequencing, using NGS, and millions of personal genomes have been read. During the same period, the commercial 'direct-to-consumer' services that provide genomic data have flourished. Thus, receiving genomic results has become a normal occurrence for the public. However, for research projects, the return of genomic results has remained difficult and infrequent and the global situation remain uncharacterized. Here, we performed a review on the return of genomic results obtained from research projects and found such returns to still be rare. We systematically reviewed literature from two years (2016-2017) obtained from the public databases, among around extracted 500 articles on return of genomic results, we found that only 27 articles mentioned actual returning in the research context. Of these 27 articles, 22 projects aim or plan to return genomic results to the participants. Those articles were not confined to specific countries, but were distributed across United States, Britain, Canada, Sweden, Estonia, Singapore, Germany, Australia, Switzerland, and Japan. These projects belonged to various categories, such as basic research using biobank samples, research aimed to increase the clinical usage of genomic information, return of secondary findings from the research with cancer patients, etc. Some projects published the gene list of return, and many projects referred to the list issued by American College of Medical Genetics and Genomics (ACMG) in 2013, the only list published by the public organization. The ACMG list was designed for clinical purpose but it still seems to be useful for research projects. In conclusion, the number of research projects that sequence personal genomes is increasing; however, our review reveals that a very limited number of projects are willing to return genomic results to participants. The literature survey indicates that the process for returning genomic results from each research project is burdensome. However, public awareness regarding the use of genome information is growing. Thus, the return of genomic results is gaining significance. The early attempts by the projects reviewed here to return genomic results will provide insights into the logistics and management of return of genomic data.

633T

Airmen's attitudes toward genomic sequencing in the US Air Force:

Results from the MilSeq Project. D.K. Petersen¹, S. Pereira¹, J.O. Robinson¹, R.L. Hsu¹, M. Majumder¹, E. Parasidis², M. Mehlman³, K. Christensen⁴, M.D. Maxwell⁴, C. Blout⁴, M. Lebo⁵, R. Brenner⁶, C. Gardner⁷, M. De Castro⁸, R.C. Green¹, A.L. McGuire¹. 1) Baylor College of Medicine, Houston, TX; 2) Ohio State University, Columbus, OH; 3) Case Western Reserve University School of Law, Cleveland, OH; 4) Brigham and Women's Hospital, Boston, MA; 5) Partner's Personalized Medicine, Cambridge, MA; 6) AFMS Personalized Medicine, Falls Church, VA; 7) Joint Base San Antonio - Lackland AFB, JBSA-Lackland AFB, TX; 8) Keesler AFB, Keesler Air Force Base, MN.

Genomic sequencing (GS) in active-duty military healthcare settings comes with a unique set of considerations. GS has the potential to return information that may impact a service member's perceived mission-readiness, duty assignments, and career. Although scholars have commented on this issue, little is known about the perspectives of those serving in the military. The MilSeq Project is a pilot study investigating the use of GS in the clinical care of apparently healthy, active-duty Airmen (n=75) in the United States Air Force (USAF). Participants completed a baseline survey prior to enrolling in the sequencing study. Here we describe Airmen's baseline attitudes towards the use of GS in the USAF, including perspectives on hypothetical testing for non-disease traits. The majority of participants reported that they trusted the Air Force with their genetic information (79%) and expected that GS would not influence their career (89%). However, participants were largely neutral or negative toward hypothetical scenarios in which the USAF would require all Airmen to undergo GS (55%, 24%, respectively), as well as using genetic information to make decisions about deployment (60%, 34%) or duty assignments (66%, 29%). Most participants were supportive of hypothetical testing for traits that could affect duty performance. These included tolerance for sleep deprivation (76%), extreme environments (71%), and starvation (57%); personal traits such as injury recovery (81%), intelligence (76%), and impulsivity (58%); and mental health risks such as anxiety (73%), PTSD (68%) and substance abuse (66%). As technologies and our ability to understand complex traits improve, our results suggest that military members would generally be receptive to learning genetic information that could inform their risk for developing conditions that may affect their duty-related performance. Moreover, Airmen showed favorable attitudes about the use of GS in a US military setting when not used for deployment or assignment decisions. We discuss the implications of Airmen's perspectives on the breadth of uses of GS, as well as their views on when and how its information should be applied in the Air Force setting.

634F

The EXCEED study: Practical issues in design and conduct of a recall-by-genotype resource. N.F. Reeve¹, C. John¹, R. Hsu¹, R.C. Free², E.J. Hollox³, D.J. Shepherd¹, L.V. Wain¹, S.E. Wallace¹, M.D. Tobin¹.

1) Health Sciences, University of Leicester, Leicester, Leicestershire, United Kingdom; 2) NIHR Leicester Biomedical Research Centre – Respiratory, Glenfield Hospital, Groby Road, Leicester, LE3 9QP, United Kingdom; 3) Department of Genetics and Genome Biology, University of Leicester, Leicester, United Kingdom.

Resources like UK Biobank provide exciting opportunities for large-scale genetics and genomics, but the translation of findings to the clinic is challenging. Recall-by-genotype studies have a high potential to contribute to such translation, facilitating a deeper understanding of genetic variants that have been identified in genome-wide association studies (GWAS) as being related to complex traits and diseases. This approach involves bringing back participants based on their genotype(s) for detailed assessments chosen on the basis of the known or suspected mechanism of the relevant gene. Such studies can be observational in design or interventional, for example to evaluate the response or safety profile of a treatment conditional on genotype. Recall-by-genotype studies are expected to be highly valuable for medicines development and targeting of therapeutics in a precision medicine approach, but also raise questions as to how best to manage the recall process while respecting participant wishes. A strongly related issue is that of disclosing potentially sensitive results to participants. We have established the Extended Cohort for E-health, Environment, and DNA (EXCEED) including the capability to recall participants by phenotype or genotype. EXCEED has a comprehensive participant consent form, with 12 questions covering topics such as access to medical records, recall- by-phenotype and -genotype, and receiving genetic results. So far, 9,600 participants aged 40 to 69 years have been recruited to EXCEED, primarily from primary care services in Leicestershire, UK. All participants provide baseline lifestyle data, a saliva sample for DNA extraction, and consent to link to their primary healthcare record, and to recall-by-phenotype and -genotype. Some studies incorporating genetic analyses (such as Genomics England) actively seek clinically actionable variants. Most cohort studies do not but researchers accessing the resource may in the future through their investigations. In anticipation of this potential, at the time of consent, we asked whether participants would wish to be notified about clinically actionable variants; 99.5% of participants stated that they would wish to be informed in this situation. Understanding the reasons for participants' preferences, how these may change over time and how these can best be supported by future policies and procedures will be of key importance for longitudinal cohort studies.

635T

Privacy attitudes and information sensitivity perceptions are associated with health services utilization. C.K. Rubanovich¹, C. Cheung², C. Schairer³, C.S. Bloss^{3,4}. 1) San Diego State University/University of California San Diego Joint Doctoral Program in Clinical Psychology, San Diego, CA; 2) The Qualcomm Institute, California Institute for Telecommunications and Information Technology, University of California San Diego, La Jolla, CA; 3) Department of Psychiatry, School of Medicine, University of California San Diego, La Jolla, CA; 4) Department of Family Medicine and Public Health, School of Medicine, University of California San Diego, La Jolla, CA.

Introduction: The current "big data" era is one in which voluminous and granular personal data are collected on individuals and then stored. Given that the current era is also one in which concerns about privacy are growing, this study sought to identify which types of personal data are considered sensitive and whether perceived sensitivities are associated with health services utilization. **Methods:** Participants ($n = 102$) completed surveys including measures of sensitivity of personal information, privacy concern, and health services utilization as part of a larger study focused on privacy attitudes in relation to emerging health technologies. In terms of sensitivity of personal information, twelve types of personal data were rated on a scale of 1 to 10, from not at all to very sensitive. Health services utilization was measured using the Health Care Utilization scale from the Stanford Patient Education Research Center, which included the number of physician visits in the last six months. Descriptive analysis compared non-utilizers (0 physician visits) to utilizers (1 or more physician visits), and linear regression models sought to test the association between health services utilization and sensitivity ratings with age as a co-variate. **Results:** Overall, the most sensitive information types were: 1) bank account information ($M = 9.63$, $SD = 1.36$), 2) electronic health records ($M = 8.29$, $SD = 2.37$), and 3) genetic test results ($M = 8.18$, $SD = 2.62$). Compared to non-utilizers, health services utilizers rated ten out of the twelve types of personal data as more sensitive. Controlling for age, health services utilization significantly predicted sensitivity ratings of medication ($p = .008$), state of health at present ($p = .009$), fitness at present ($p = .046$), medical history ($p = .027$), addictions ($p = .049$), and electronic health records ($p = .027$). **Conclusions:** Prior health services utilization was significantly associated with perceived sensitivity of personal data for data types that are health-related in nature. These findings may help guide the design of systems that elicit patient preferences in sharing health data, for instance, systems that seek to leverage large patient electronic health record databases for research. More broadly, these findings can inform our understanding of individual differences in privacy attitudes and perceptions of personal data.

636F

Attitudes towards genetic testing among participants in the Jackson and Framingham Heart Studies. K.W. Saylor¹, L. Ekwunwe², D. Antoine-Lavigne², D.E. Sellers³, S. McGraw⁴, D. Levy^{5,6}, G.L. Splansky⁶, S. Joffe⁷. 1) Department of Public Policy, University of North Carolina, Chapel Hill, NC; 2) The Jackson Heart Study, Jackson, MS; 3) College of Human Ecology, Cornell University, New York, NY; 4) The Hastings Center, Garrison, NY; 5) Framingham Heart Study, Framingham, NY; 6) Population Sciences Branch, National Heart, Lung and Blood Institute of the National Institutes of Health, Bethesda, MD; 7) Department of Medical Ethics and Health Policy, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA.

PURPOSE: To identify research participant characteristics associated with positive and negative attitudes towards genetic testing. Findings can inform participant engagement and education. **METHODS:** Secondary analysis of survey data from participants in genetic testing components of the Jackson Heart Study (JHS, $n=960$, mean age=63, 66% female, 100% African American), Framingham Heart Study (FHS, $n=955$, mean age=58, 55.2% female, 99% white), and FHS-Omni Cohort ($n=160$, mean age=64, 62% female, 88% African American/11% mixed race). Three endpoints (a 3-item ordinal attitude scale, and positive and negative word scores based on a checklist of attitude descriptors) were dichotomized as close to the 25th percentile as the data allowed. Final logistic regression models include participant characteristics common to models with the lowest Akaike and Bayesian information criteria. **RESULTS:** Survey response rates were 67-73% across cohorts. Most participants had positive attitudes towards genetic testing (median score 4.3-5/5). Participants endorsed a median of 2 positive words and 0-1 negative words. "Hopeful" and "optimistic" were endorsed by 45-70% of participants, "enthusiastic", "excited", "mixed feelings", "concerned" and "cautious" by 25-55%, "confused", "indifferent" and "worried" by 6-16%, and "pessimistic" and "horrified" by 1-5%. Characteristics associated with favorable attitudes included: greater genetics knowledge and subjective numeracy (JHS); greater experience with genetic testing, subjective numeracy, and religious attendance (FHS); and not being employed (FHS-Omni) (see Table).

	JHS OR [95% CI]	p-value	FHS OR [95% CI]	p-value	FHS-Omni OR [95% CI]	p-value
<u>More negative attitude about testing</u>						
Experience with genetic testing			0.45 [0.23-0.90]	-0.023		
Genetics knowledge	0.87 [0.80-0.96]	-0.004				
<u>Endorse more positive words</u>						
Genetics knowledge	1.29 [1.16-1.43]	(<0.001)				
Subjective numeracy	1.28 [1.13-1.45]	(<0.001)	1.17 [1.04-1.31]	-0.008		
Not employed					2.15 [1.04-4.74]	-0.04
<u>Endorse more negative words</u>						
Subjective numeracy	0.88 [0.80-0.98]	-0.017				
Religious attendance			1.20 [1.04-1.38]	-0.012		

CONCLUSION: Attitudes towards genetic testing were generally positive among participants in two major cardiovascular cohort studies. Greater numeracy, and experience with and knowledge of genetics were associated with more positive attitudes. These findings can help researchers moderate excessive enthusiasm or skepticism about genetic testing.

637T

Ethnic identity and intentions to learn results in an exome sequencing study. E. Turbitt^{1,2}, MC. Roberts³, BM. Hollister¹, KL. Lewis⁴, LG. Biesecker⁴, WMP. Klein^{1,3}. 1) Social and Behavioral Research Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA; 2) Bioethics Core, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA; 3) Behavioral Research Program, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA; 4) Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA.

Background. Ethnic identity is part of one's social identity, derived from both knowledge about membership of one's ethnic group, and the value and emotional significance related to that membership. Endorsement of strong ethnic identity has been associated with better health outcomes. With evidence of racial disparities in genetic testing and research, there is a critical need to better understand the experiences of participants in studies involving genome sequencing among racially diverse cohorts. Considering its association with various health related factors, studying ethnic identity in the context of genome sequencing studies may add to our understanding of these disparities. We examined the potential association of ethnic identity with beliefs about and intentions to learn results of genome sequencing. **Methods.** A longitudinal cohort was recruited to implement genome sequencing among healthy participants who self-identify as African, African-American, or Afro-Caribbean, aged 45-65 years old ($n=408$). Before receiving sequencing results, participants completed a survey assessing a range of social and behavioral constructs related to health, genome sequencing and ethnic identity. Bivariate and moderation analyses were performed. **Results.** Stronger ethnic identity was positively correlated with anticipated value of sequencing results ($r = 0.14$, $p < 0.01$) and benefits from sequencing study participation ($r = 0.15$, $p < 0.01$), but negatively correlated with expectations for harms from sequencing study participation ($r = -0.12$, $p < 0.05$). For participants with stronger ethnic identity, there was a positive association of perceived value of results with intentions to receive results ($b = 0.56$, 95% CI: 0.24, 0.89, $p < 0.001$); whereas for those with weaker ethnic identity, there was no such association ($b = -0.093$, 95% CI: -0.36, 0.17, $p = 0.50$). **Discussion.** Individuals with stronger ethnic identity may attend more to the value of sequencing results when deliberating whether to receive results. The findings of our study highlight that investigators recruiting participants from diverse backgrounds should consider the influence of ethnic identity on participants' beliefs about sequencing participation and their intentions to receive genome sequencing results. In addition, understanding the variation in ethnic identity and its influence on genome sequencing perceptions and intentions can inform development of culturally-sensitive protocols for research and clinical applications.

638F

Repurposing of whole-exome sequencing data for personalized medicine. M. van der Lee¹, W.G. Allard², A.H.B. Bollen³, G.W.E. Santen³, C.A.L. Ruivenkamp³, F. Baas³, M. Kriek³, H.J. Guchelaar³, S.Y. Anvar^{1,2}, J.J. Swen¹. 1) Department of Clinical Pharmacy and Toxicology, Leiden University Medical Center, Leiden, Netherlands; 2) Department of Clinical Genetics, Leiden University Medical Center, Leiden, Netherlands; 3) Department of Human Genetics, Leiden University Medical Center, Leiden, Netherlands.

Background Whole exome sequencing (WES) is currently used on a wide scale within academic patient care. The resulting genetic data includes complementary information on pharmacogenetic (PGx) variants that could be used to optimize drug treatment where applicable. It allows personalized medicine rather than the established 'one size fits all' approach to prescribing drugs. Two consortia, the Dutch Pharmacogenetics Working Group (DPWG) and the Clinical Pharmacogenetics Implementation Consortium (CPIC), together provide therapeutic recommendations for more than 50 gene-drug pairs for patients with a known genotype. We set out to develop a new strategy to implement PGx screening into the routine genetic workflow and to combine this data with medication records and the DPWG guidelines. **Methods** From the DPWG guidelines, 45 actionable variants in 13 key pharmacogenes were selected for retrieval from WES data in a cohort of 500 child-parent trios. A trio based analysis was done to resolve phasing and identify *de novo* variants. Translation of variant-to-haplotype and haplotype to drug-metabolizer phenotype was done according to the DPWG guidelines. Medication records of included patients were matched to the drug-metabolizer-phenotypes based on specific codes for each pharmaceutical compound (ATC codes). **Results** 89% of the subjects carried at least one actionable variant, the majority carried at least 2 actionable variants (58.6%). Five variants were located in intronic or promoter regions which lacked coverage in WES data, leading to an inability to genotype HLA-A and HLA-B, and in many cases UGT1A1 and CYP2C19*17. We had access to medication records for 351 (23.4%) subjects, of whom 4.8% were exposed to a gene-drug interaction requiring a drug change or dose adjustment, of which the majority were with analgesics (76.5%). **Conclusion** This study shows the feasibility of repurposing WES data for PGx purposes. It identifies actionable PGx variants in almost 90% of the study cohort. Due to the limitations of WES data, variant coverage in intronic regions and the detection of Copy Number Variants, these numbers are an underestimation of the true number of actionable variants. The next step will be the implementation of this method as a routine pre-emptive approach into the genetic diagnostic setting, which will aid the use of PGx based prescribing in clinical practice.

639T

An innovative software to scale genetic counseling needs in a time of increasing demand. G. Snir¹, L. Diskin², N. Shrier-Shafir³, Y. Segal², B. Chen M.³, M. Mizrahi², S. Alboim², S. Katz², D. Cohen², A. Golan², M. Berger², A. Beit-Or², V. Nadler², T. Wolf², V. Katalan², N. Ash², Y. Wolf², M. Snir¹, H. Baris-Feldman³, M. Shohat^{1,2,4}. 1) ClearGenetics, San Anselmo, CA; 2) Maccabi Health Services, Tel Aviv, Israel; 3) Meuhedet HMO, Tel Aviv, Israel; 4) Sackler School of Medicine, Tel Aviv University, Israel.

Increased prenatal testing options combined with limited genetic counseling personnel and short OB visits, present a challenge in delivering genetic counseling. We have introduced an innovative software to improve identification of high risk pregnancies and expedite access to genetic counseling in two HMOs in Israel. The software automatically collects from EMRs the results of nuchal translucency, biochemical screening markers and ultrasound findings, and generates a detailed, personalized report at 17 weeks. The patient and treating OB receive the report that graphically visualizes: A. The fetus's individual risk for Down syndrome, other chromosomal anomalies and other non-chromosomal genetic diseases B. The impact of normal result in amniocentesis (with CMA) or NIPT, or other options on each risk C. The medical recommendation for genetic counseling, amniocentesis, and other tests D. Administrative information regarding cost and coverage. In the initial launch, the software thresholds were set at a low stringency to allow the maximal freedom for referral to genetic counseling. The thresholds were then adjusted by a group of geneticists who reached a consensus on each of the rules in the software algorithm. All cases were manually reviewed for the first 6 months (15,932 cases), to ensure the quality of the generated recommendations. A high agreement was observed between the genetic counseling team and software's recommendations. Since May 2017, the software analyzed 36,180 pregnancies. Ultrasound finding was found in 12.7% of the pregnancies and medical recommendations for amniocentesis was given to 6.9% of the women. Before implementation, about 28% of pregnancies were referred to genetic counseling, either for a medical reason or patient request. The other 72% weren't regularly informed about their risk or prenatal testing options. Three months following software implementation, the referral rate stabilized at 8%. All other 92% of women received the personal software report to discuss with their OB. We present a unique software that performs universal screening for genetic counseling for all pregnancies, using triage criteria defined by medical geneticists, genetic counselors and OBs. The personal report that patients receive, properly informs of the risk and provides a simple tool to understand the value of the available prenatal testing options. This process expedites genetic counseling in high risk pregnancies while ensuring high quality of care to all.

640F

A smartphone app to facilitate family communication of genetic results. C.J Jujjavarapu, L.A Amendola, M.H.P Horike-Pyne, S.D Mooney, G.P.J Jarvik. University of Washington, Seattle, WA., USA.

The convergence of cost-efficient, reliable genomic technology and increasing information about the impact of genomic variation on health presents an opportunity to advance genomic medicine. However, a key challenge is the barriers to sharing of genetic results with family members. Medical results from genetic tests are unique in healthcare because the results of one individual might be important to the healthcare of relatives. Despite this importance, numerous studies have documented lack of family communication about genetic risk. Controlled trials of genetic counseling-based interventions have shown improved satisfaction but not higher rates of result communication with family members. This supports the need for fresh approaches. We built ShareDNA, a free smart phone app, to lower the barrier to familial test result sharing by allowing those with genetic test results to easily share that result with anyone contactable by text or email. Increased sharing of family test results can improve both cost effectiveness and health outcomes. The ShareDNA app allows either a clinical lab or the user to upload test results, which the user can then share with anyone in their contact list or whose contact information they enter. The app was built using Xcode and will soon be available on the Apple App Store. The app provides an introduction on how to use this service, requires users to create a username and password and upload their genetic test results. The app also helps users select which contact(s) they would like to send their information to and then generate a pre-written message to send to them with the results. Links to relevant websites and a directory of genetics providers are included. Usability testing sessions have been performed to better understand how potential users experience and perceive ShareDNA. The app was well-received with minor wording and functionality modifications suggestions. Knowing the genetic results of family members is important for optimal care. For example, biologic relatives of a patient with a positive cancer genetic result can be tested for the family-specific pathogenic variant for ~ 1/10th the cost of a multi-gene panel test. Knowledge of a familial variant enables preventative and early detection strategies and is cost-effective (PMID:25940718). Implementing novel strategies to support increased familial sharing of results, such as the ShareDNA app, are essential to capitalize on the full potential of genomic medicine.

641T

The role of emerging disease-modifying therapies in reproductive decision-making for cystic fibrosis and Duchenne muscular dystrophy carrier couples. A.F. McCague^{1,2}, K.M. Stuttgen^{1,2}, J.M. Bollinger¹, R.L. Dvoskin¹, D.J.H. Mathews¹. 1) Berman Institute of Bioethics, Johns Hopkins University, Baltimore, MD; 2) Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD.

Increasingly, people are making reproductive decisions with genetic test results in hand. As disease-modifying therapies emerge for serious genetic disease, it is important to understand how people are evaluating and using their genetic test results, in order to assist the medical genetics community in designing screening and counseling protocols. For example, there are now FDA-approved small molecule therapies for cystic fibrosis (CF) that have the potential to dramatically improve patient outcomes. New therapies are also becoming available for Duchenne muscular dystrophy (DMD). This study sought to evaluate the impact of disease-modifying therapies on reproductive decision-making for CF and DMD carrier couples. We conducted semi-structured interviews with 19 DMD mutation carriers and 10 CF mutation carriers whose partners are also carriers, including couples with and without affected children. Interview questions addressed the following domains: familiarity with the disease prior to testing; how genetic test results were communicated in the family; reproductive decision-making; if and how disease modifying therapies affected that decision-making; and views regarding preconception genetic testing. Among CF carriers, there was an even split between those who said that disease-modifying therapies had no effect on their reproductive decision-making, those who indicated that a therapy would change their reproductive planning only if it represented a cure, and those who said that they would be more willing to risk having an affected child due to the existence of therapies. For the majority of DMD carriers, the therapies either had no effect or would change their reproductive decisions only if they represented a cure. A minority responded that given the existence of an approved therapy, they would be more willing to risk having an affected child. Two minority themes present in the DMD—but not the CF—group were caveats about the cost and availability of therapies, as well as a willingness to risk having a carrier daughter, given available therapies. Others in both disease groups said they probably would have made the same decision, but the existence of therapeutic intervention may have lessened their anxiety surrounding the decision. These views were affected by numerous factors, including whether the couple had affected children and the severity of their child's disease, their personal experience with and knowledge of the therapies, and personal religious views.

642F

Development of patient "profiles" to tailor counseling for incidental genome sequencing results. L. Carlsson¹, C. Mighton^{1,2}, M. Clausen², S. Shickh^{1,2}, S. Casalino², T.H.M. Kim³, L. McCuaig^{1,2}, E. Joshi^{1,2}, N. Baxter^{1,2}, A. Scheer^{1,2}, C. Elser⁴, A. Eisen⁷, M. Evans^{1,2}, R.H. Kim^{6,8,9}, S. Panchal⁸, T. Graham⁷, M. Aronson^{1,8}, C. Piccinin⁸, L. Winter-Paquette^{1,8}, K. Semotiuk^{1,8}, J. Lerner-Ellis^{1,8}, J.C. Carroll^{1,8}, J. Hamilton⁵, E. Glogowski³, K. Schrader⁴, K. Offit⁵, M. Robson⁵, K. Thorpe¹, A. Laupacis^{1,2}, Y. Bombard^{1,2}. 1) University of Toronto, Toronto, ON; 2) St. Michael's Hospital, Toronto, ON; 3) GeneDX, Gaithersburg, MD; 4) BC Cancer Agency, Vancouver, BC; 5) Memorial Sloan Kettering Cancer Center, New York, NY; 6) University Health Network, Toronto, ON; 7) Sunnybrook Health Sciences Centre, Toronto, ON; 8) Mount Sinai Hospital, Sinai Health System, Toronto, ON; 9) The Hospital for Sick Children, Toronto, ON.

Background: Despite guidelines recommending return of medically actionable incidental results (IR), research indicates patients are interested in learning results beyond these. Providers should engage patients in shared decision-making prior to genome sequencing (GS), but this can be extremely time consuming given the myriad of IR and variation in preferences. There is a need to streamline pre-test counseling for IR. We identified patient profiles based on shared characteristics of patients' decision-making processes for learning IR. **Methods:** We conducted semi-structured interviews with participants as a qualitative component of a randomized trial of a decision aid (the Genomics ADVISER) for selecting IR. Participants had a personal or family history of cancer, received uninformative genetic test results, making GS potentially useful. Interviews explored factors participants considered when deliberating over learning IR. Interviews were transcribed and analysed by thematic analysis. Transcripts were coded, and themes identified using constant comparison. **Results:** Participants (n=31) were mostly female (28/31) and about half over 50 years old (16/31). We identified key characteristics that informed preferences for IR: contextual factors, disposition, attitudes toward IR, concerns about IR, and perceived utility of IR. Based on how these characteristics were shared by groups of participants, we identified five patient profiles. *Information Enthusiasts* self-identified as "planners" and valued learning most/all IR to enable planning and disease prevention because "knowledge is power". *Concerned* individuals defined themselves as "anxious," were reluctant to learn IR, anticipating negative psychological impacts from IR. Individuals of *Advanced Life Stage* primarily considered implications of IR for younger family members. *Reassurance Seekers* were reassured from previous negative genetic test results, which shaped their expectations for receiving no IR: "hopefully [GS will] be negative, too. And then I can rest easy". *Contemplators* were discerning about the value and limitations of IR, heavily weighing the health benefits with the impacts of not being able to "un-know" information. **Conclusions:** These profiles are a starting point towards tailoring shared decision making for IR to address the variation in patients' preferences. These profiles could guide clinicians' counseling by providing a framework to address common values, concerns and misconceptions about IR.

643T

Primary care providers' perspectives on the use of genome sequencing in research among healthy children. E. Joshi¹, C. Mighton^{1,2}, M. Clausen¹, S. Casalino¹, T.H.M. Kim^{1,3}, C. Kowal³, C. Birken³, J.H. Maguire^{1,4,5}, Y. Bombard^{1,2}. 1) Li Ka Shing Knowledge Institute, St. Michael's Hospital, Toronto, Canada; 2) Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON; 3) The Hospital for Sick Children, Toronto, ON; 4) Department of Pediatrics, University of Toronto, Toronto, ON; 5) Department of Nutritional Sciences, University of Toronto, Toronto, ON.

Background: There is growing interest in using genome sequencing (GS) for healthy children in research studies to improve scientific knowledge of genetic variation and better characterize the relationship between genetics, environment and disease. Very little research has explored primary health-care providers' views on using GS in this context. This study aimed to explore physicians' perspectives on the use of GS in healthy children and overall perspectives towards the use of this technology in their care for patients.

Methods: We conducted in depth, semi-structured interviews with physicians who care for children in a longitudinal cohort study called TARGet Kids!. No GS was offered as part of this study. Interviews were audio-recorded, transcribed verbatim and analyzed by thematic analysis. Transcripts were coded for themes using constant comparison. **Results:** Participants were predominantly family physicians (10/16) and had practiced for over 10 years (11/16). Participants' perspectives on GS for healthy children diverged into two groups: (1) those who valued using results from GS in practice, and (2) those who were concerned about the current lack of evidence to support clinical use of GS. Those who focused on the value described a multitude of ways that they could use GS results to inform their patients' care through monitoring and preventive actions. They also expressed a desire to offer all GS results to parents, as they presumed parents would see value in the information. Participants who expressed concern about the current lack of evidence on the utility of GS questioned the clinical validity and utility of results, and challenged the appropriateness of using research data in their patients' care. They also favoured returning only medically actionable results to parents. Providers in both groups acknowledged that offering GS to healthy children conflicts with established screening principles. In addition, both groups shared concerns about the analytic validity and interpretation of GS results, the need for additional resources, potential for increased health system costs and inequities, and the possibility for genetic discrimination if results from GS research were to be adopted into clinical practice. **Conclusions:** Primary healthcare providers recognize that they will inevitably play a role in the use of GS towards precision medicine, but differ in their perceptions of the appropriateness and utility of GS in the care for healthy pediatric patients.

644F

Quality of life is a major driver of patients' preferences for incidental genome sequencing results. C. Mighton¹, L. Carlsson^{1,6}, M. Clausen², S. Shickh^{1,2}, S. Casalino², T.H.M. Kim³, L. McCuaig^{1,2}, E. Joshi^{1,2}, N. Baxter^{1,2}, A. Scheer^{1,2}, C. Elser⁴, A. Eisen⁷, M. Evans^{1,2}, R.H. Kim^{6,8,9}, S. Panchal⁸, T. Graham⁷, M. Aronson^{1,8}, C. Piccinin⁹, L. Winter-Paquette^{1,8}, K. Semotiuk^{1,8}, J. Lerner-Ellis^{1,8}, J.C. Carroll^{1,8}, J. Hamilton⁵, E. Glogowski⁵, K. Schrader⁴, K. Offit⁵, M. Robson⁵, K. Thorpe¹, A. Laupacis^{1,2}, Y. Bombard^{1,2}. 1) University of Toronto, Toronto, ON; 2) St. Michael's Hospital, Toronto, ON; 3) GeneDx, Gaithersburg, MD; 4) BC Cancer Agency, Vancouver, BC; 5) Memorial Sloan Kettering Cancer Center, New York, NY; 6) University Health Network, Toronto, ON; 7) Sunnybrook Health Sciences Centre, Toronto, ON; 8) Mount Sinai Hospital, Sinai Health System, Toronto, ON; 9) The Hospital for Sick Children, Toronto, ON.

Background: Secondary or incidental results (IR) from genome sequencing (GS) are typically categorized based on their medical actionability (MA). However, patients are interested in learning results beyond those that are MA, as these findings can have personal utility. It has been suggested that personal utility should be considered in the evaluation of genetic tests, but it is a challenging concept to measure. We explored how patients making hypothetical decisions about learning IR perceive actionability and how actionability shapes preferences for IR, with the goal of capturing both personal and clinical utilities of learning IR. **Methods:** We conducted a qualitative study within an RCT of a decision aid (DA) for the selection of IR from GS (The Genomics ADVISER). Participants were adult cancer patients who had previously received an uninformative result from genetic testing related to their cancer. Participants used the DA or spoke with a genetic counselor to choose between five categories of IR: 1) Medically actionable, 2) common disease SNPs, 3) rare genetic disorders, 4) early-age brain diseases and 5) carrier status. Interviews were audio-recorded and transcribed verbatim. Transcripts were analyzed using thematic analysis; data was labelled with thematic and descriptive codes using constant comparison. **Results:** We interviewed 31 participants, who were predominately female (28/31) and about half were over 50 years old (16/31). Participants perceived actionability in all types of IR, not only those considered MA. Participants expressed that learning IR could allow them to take actions to control disease outcome, or to take control over aspects of their personal life in preparation for possible illness. An overarching motivator for both wishing to learn and not wishing to learn IR was preserving quality of life. For some participants, learning about any type of IR was perceived as a way to maintain their quality of life through taking actions such as disease prevention or preparation. For others, choosing not to learn particular categories of IR was seen as a way to preserve quality of life by avoiding potential negative psychological impacts of learning IR. **Conclusions:** Preserving quality of life was a strong motivator in patients' preferences for and against learning IR. Quality of life measures that incorporate the value of information and non-health outcomes would provide a more comprehensive evaluation of genetic tests and the impacts of IR from GS.

645T

Whole genome sequencing: Parent-reported outcome measures. J. Nelson, M. Brown, L. Villarroel, D. Salyakina, J. McCafferty, P. Jayakar, K. Schain, B. Totapally, M. Diaz-Barbosa, J. Tsasaki, A. Quittner. Research Institute, Nicklaus Children's Research Institute, Miami, FL.

Objectives: The use of genomic sequencing has the potential to significantly alter clinical practice, patient care, and health outcomes. Although the application of these techniques is moving forward rapidly, little is known about the psychosocial, behavioral or ethical effects of this technology, particularly in pediatrics. The aims of this study are to develop the first parent-reported outcome measures for families participating in whole genome sequencing (WGS). **Methods:** Open-ended, qualitative interviews were conducted with 11 parents whose children were undergoing WGS; about half were infants who underwent rapid WGS (rWGS) and half were children in our Odyssey genetics program. Parents were interviewed at time of enrollment and then at return of results. Two brief screening measures of depression (PHQ-9) and anxiety (GAD-7). Interviews were audio-recorded and transcribed for coding using NVivo. **Results:** To date, 15 families have been enrolled (rWGS = 6; Odyssey = 9); complete data were available on 11 families. The sample was racially and ethnically diverse; half identified as Hispanic. Mean age of children rWGS = 9.0 months; Odyssey = 7.4 years. Psychological screening at time of enrollment for those undergoing rWGS indicated that 50% to 67% of parents reported elevated depression and 67-100% reported elevated anxiety. For those in the Odyssey program at time of enrollment 33% of parents were elevated in both depression and anxiety. To date, the majority of infants undergoing rWGS have had a pathological variant identified (e.g., BRAT1, MAGEL2), whereas only 1 child in the Odyssey program has been identified with a new genetic variant. Themes emerging from the qualitative data for the rWGS group suggest that: 1) parents had limited understanding of information provided in the consent; 2) parents had very high expectations that the genetic diagnosis would lead to a significant change in management; 3) parents wanted better communication with their providers. Families in the Odyssey program had lower expectations for a new genetic diagnosis and reported less psychological distress. **Conclusions & Future Directions:** These preliminary results illustrate the importance of developing outcome measures that reflect the parents' experience with these new genetic testing technologies. Several implementation challenges were identified in launching this new personalized medicine program from both a family and provider perspective.

646F

Temporal differences in concerns, motivations, and attitudes regarding predispositional genome sequencing among healthy adults: Findings from the PeopleSeq Consortium. E.S. Zoltick^{1,2}, M.D. Linderman^{3,4}, D.E. Nielsen⁵, W.N. Betting¹, M.A. McGinniss⁶, E. Ramos⁷, M.P. Ball^{8,9}, D.G.B. Leonard¹⁰, S. Pereira¹¹, S.S. Sanderson^{4,12}, S.D. Crawford¹³, R.C. Green^{1,14,15,16}, the PeopleSeq Consortium. 1) Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Boston, MA; 2) Section of Preventive Medicine and Epidemiology, Department of Medicine, Boston University School of Medicine, Boston, MA; 3) Department of Computer Science, Middlebury College, Middlebury, VT; 4) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 5) School of Human Nutrition, McGill University, Montreal, Quebec, Canada; 6) Illumina Inc., San Diego, CA; 7) Geisinger, Danville, PA; 8) Harvard Personal Genome Project, Harvard Medical School, Boston, MA; 9) Open Humans Foundation, Boston, MA; 10) Department of Pathology and Laboratory Medicine, Robert Larner, M.D. College of Medicine of the University of Vermont, Burlington, VT; 11) Center for Medical Ethics and Health Policy, Baylor College of Medicine, Houston, TX; 12) Department of Behavioural Science and Health, University College London, London, UK; 13) SoundRocket, Ann Arbor, MI; 14) Harvard Medical School, Boston, MA; 15) The Broad Institute of MIT and Harvard, Cambridge, MA; 16) Partners Healthcare Personalized Medicine, Boston, MA.

Purpose: To compare differences over time in participants' concerns, motivations, and attitudes following genome sequencing in the PeopleSeq Consortium, a collaboration of academic and commercial predispositional sequencing projects in healthy adults. **Methods:** Web-based surveys were administered to healthy adults who obtained sequencing at four programs. Participants were categorized as early (2014-2015) or recent (2016-2017) respondents based on the year of survey completion. Chi-square tests were used to compare outcomes between early (n=292) and recent (n=244) respondents. Statistical significance was set at p=0.01 given the large number of analyses. **Results:** Study surveys were completed a median of 8 months after disclosure of genomic results. Age, race, education, marital status, health status, and prior genetic testing did not significantly differ between early and recent respondents. A slightly higher percentage of early respondents were male (66% vs 56%; p=0.02). When deciding to pursue sequencing, being very concerned about receiving unwanted information was significantly higher among early respondents (14% vs 4%; p<0.001). The following were less likely to be very important motivators for pursuing sequencing for early compared to recent respondents: "Interest in finding out what I can do to improve my health" (38% vs 51%; p=0.002) and "There is a medical condition in my family which may be genetic" (20% vs 30%; p=0.009). There was no significant difference between the groups on a composite measure of perceived utility of results. However, a higher percentage of early respondents strongly agreed that they learned something to improve their health that they did not know before (21% vs 13%; p=0.02) and were very satisfied with their decision to obtain sequencing (73% vs 60%; p=0.001). Early and recent respondents did not differ on their attitudes on access to sequencing without a medical professional or inclusion of results in medical care. **Conclusion:** Among early adopters of predispositional genome sequencing in the PeopleSeq Consortium, attitudes on access to sequencing and use in medical care did not differ between more recent and earlier respondents. Concerns and motivations did differ between the two groups, and may reflect a decrease in fear surrounding personal genomic information and an increase in anticipated health benefits. However, growing expectations of personal sequencing results may exceed current available evidence of clinical utility.

647T

Clinician and parent perceptions of genomic sequencing in the Neonatal and Pediatric Intensive Care Units: A blinded randomized clinical trial at Rady Children's Hospital. J. Cakici^{1,2,3}, M. Gaughran¹, C. Clarke¹, M. Clark¹, C. Bloss², D. Dimmock¹, S. Kingsmore¹. 1) Rady Children's Institute for Genomic Medicine, San Diego, CA; 2) University of California- San Diego, San Diego, CA; 3) San Diego State University, San Diego, CA.

Introduction: Congenital anomalies are a leading cause of admission to a Neonatal Intensive Care Unit (NICU) or a Pediatric Intensive Care Unit (PICU). Studies have shown that genomic testing in newborns has resulted in a diagnostic yield between 20-58%, leading many researchers and clinicians to believe that early diagnoses may have clinical utility. However, there is no universally accepted definition of clinical utility. **Methods:** Clinicians and parents of newborns enrolled in a randomized controlled trial of rapid whole exome sequencing versus rapid whole genome sequencing of infants in the NICU and PICU completed baseline and follow-up surveys. To date, 154 newborns less than four months of age have been enrolled at Rady Children's Hospital in San Diego. Analysis and reporting are limited to symptom-driven pathogenic or likely pathogenic variants. Parents are also enrolled to allow each arm to reflex from a singleton to trio analysis to determine the need for trio genomic sequencing. Enrolled parents are asked to provide optional feedback at the time of enrollment, following results, and at one year post-enrollment. **Results:** Of the 154 patients tested, 34 (22%) have received a diagnosis, but clinician feedback from 145 clinician surveys shows 77% of clinicians who responded (141) found genomic testing to be Useful (48 of 141; 34%) or Very Useful (60 of 141; 43%). Clinicians also provide standardized information regarding what, if any, changes were made to clinical care. Parental perceptions of utility, harm, benefit, and understanding are also being explored. Unlike other studies of parental perceptions, only 14% (31 of 220) have post-graduate education, whereas 45.5% (100 of 220) do not have education beyond high school and 40.5% (89 of 220) report having some undergraduate college education making this cohort more representative of NICU and PICU demographics. Of the 220 surveys received, three parents reported of harm to the parent (3 of 220 parents; 1%) and two reported of harm to the child (2 of 100 probands; 2%). Alternatively, 74% (77 of 104) indicated that the results were useful. **Conclusion:** This study provides insights into both clinician and parental perceptions of genomic testing in newborns that can be used to create rigorous definitions of clinical changes in management and provide a deeper understanding of parental perceptions in order to shape best practices and adoption of genomic testing in the intensive care setting.

648F

Pragmatism in pharmacogenetics trials: A PRECIS-2 perspective on the Integrating Pharmacogenetics in Clinical Care (I-PICC) Study. C.A. Brunette¹, S.J. Miller¹, N. Majahmal², C. Hau¹, L. MacMullen¹, S. Advani¹, A.J. Zimolzak^{1,2}, J.L. Vassy^{1,3,4}. 1) VA Boston Healthcare System, Boston, MA; 2) Boston University School of Medicine, Boston, MA; 3) Harvard Medical School, Boston, MA; 4) Division of General Internal Medicine and Primary Care, Brigham and Women's Hospital, Boston, MA.

Background Conducted in real-world clinical settings, pragmatic trials yield generalizable scientific evidence about novel healthcare innovations. Pharmacogenetic (PGx) testing in patient care is an emerging medical innovation in need of outcomes assessment. The I-PICC Study is an ongoing randomized trial collecting patient outcomes after *SLCO1B1* PGx testing for simvastatin myopathy risk. Examining the pragmatism of the I-PICC Study can help determine the generalizability of its eventual results. **Methods** We mapped the I-PICC Study design using the PRagmatic EXplanatory Continuum Indicator Summary 2 (PRECIS-2) tool, which scores 9 clinical trial domains on a 5-point Likert continuum from explanatory to pragmatic. **Results** The I-PICC Study design is highly pragmatic in all PRECIS-2 domains (scores 3-5). 1) **Eligibility** criteria for providers and patients are inclusive and align with intended intervention recipients. 67% of available primary care providers (PCPs) were enrolled between July 31, 2016 and April 18, 2018. Of 6,142 eligible patients encountered, 81% were exposed to the intervention via an enrolled PCP. 2) Informatics methods are leveraged to automate PCP and patient **recruitment**. Patient contact is limited to a study letter and telephone consent. 3) The **setting**, 8 primary care clinics across the VA Boston Healthcare System (VABHS), matches where PGx testing might be applied in clinical medicine. 4) The study fits within the VABHS **organization** and requires no specialized training or resources beyond the *SLCO1B1* test, which is sent to a reference lab through usual send-out processes. 5) PCPs have **flexibility** in how they **deliver** PGx results to patients. 6) Patients have **flexibility** in how they **adhere** to any recommendations from the PGx results. 7) All **follow-up** occurs observationally through usual care, except a brief 12-month phone survey. 8) The **primary outcome**, LDL cholesterol, is a routine, clinically relevant biomarker but is a proxy for cardiovascular risk. 9) **Primary analysis** will include data from all enrollees as randomized. **Discussion** The I-PICC Study was designed to minimize interruption of clinical workflow as it enrolls and tracks patients in a trial of PGx testing in primary care. Its high level of pragmatism should enhance the generalizability of its findings to other healthcare systems interested in the potential impact of PGx testing on patient outcomes. The PRECIS-2 tool may be useful for the design of future trials of precision medicine.

649T

Projected prevalence of actionable pharmacogenetic variants and associated drug use in the VHA patient population: Impact and implications of adopting clinical pharmacogenetics implementation consortium (CPIC) guidelines for treatment of veterans. C. Chanfreau¹, J.A. Lynch^{1,2}, S.L. DuVal³, L.E. Hull⁴, S.M. Damrauer³, M.S. Icardi⁴, S. Tuteja³. 1) VA Salt Lake City Healthcare System, Salt Lake City, UT; 2) VA Bedford Healthcare System, Bedford, MA; 3) Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 4) VA Iowa City Healthcare System, Pathology and Laboratory Medicine Services, Iowa City, IA.

Background Pharmacogenetic (PGx) tests identify genetic variations associated with drug metabolism, and have the potential to increase treatment efficacy and reduce the risk of adverse event. The evaluation and use of PGx within the Department of Veterans Affairs (VA) is an important topic due to high prevalence of Veterans with several chronic conditions and on multiple medications. We developed a model to estimate the impact and implications of adopting Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines in the treatment of the VA population, taking in account the genetic make-up of the population and the prevalence of drug use. **Objective** To estimate the prevalence of actionable gene variants with CPIC level A evidence (strong to moderate) and use of associated drugs among VA enrollees. **Material and Methods** We used variant frequencies from the NIH 1000 Genomes project and demographics of the VA population (20% African/80% Caucasian) to estimate the prevalence of CPIC level A variants among 8.9 million Veterans. We examined the use of associated drugs using VA pharmacy data (2012-2017). **Results** Most VA enrollees (98%) would carry at least one actionable variant. Drugs associated with those variants represent 7% of VA drug orders (n=110 millions). Over the past five years 4.1 million VA pharmacy users received at least one CPIC level A drug, with 27% receiving two drugs, 14% receiving three drugs, and 11% receiving 4 or more. Drug-gene interactions with potentially greater impact included *CYP2C9*2*/Warfarin (20% predicted carriers among VA enrollees and 8% of VA pharmacy users, respectively), *CYP2C19*17*/Citalopram (5% homozygous carriers; 9% users), *CYP2D6*4*/Tramadol (3% homozygous carriers; 16% users), and *SLCO1B1*5*/Simvastatin (26% carriers; 32% users). **Conclusions** Our findings illustrate the potential clinical impact of PGx testing to inform clinical care in a large integrated healthcare system. As PGx testing becomes more prevalent it has the power to transform the prescription practices for many common drugs. Combining VA comprehensive electronic health records data with genetic information ordered in practice, or collected from studies (e.g., the Million Veteran Program), will be instrumental in evaluating real life outcomes associated with PGx testing in a diverse population.

650F

Utilization of drugs with FDA and CPIC guidelines and pharmacogenetic testing in pediatric clinical settings. D. Salyakina, I. Sotto, S. Roy, W. Wang, J. Gruber, S. Bahrami, R. Alkhouri, N. Mulas. Research Institute, Nicklaus Children's Hospital, Miami, FL.

Use of pharmacogenetic testing (PGx) in clinical settings is costly and still limited. As medical practice moves towards integrating PGx into pediatric clinic, it is important to assess the efficacy and utility associated with implementation. At the same time, it is not clear to which extent genetic testing is utilized in pediatric clinical settings. The primary goal of this study was to evaluate which medications labeled by the FDA as "Actionable PGx", "Testing required" or "Testing recommended" are utilized at the Nicklaus Children's Hospital and if all patients receiving these medications underwent genetic testing prior to treatment. In order to address this goal, we have retrospectively reviewed medical records for all patients that have received medications with PGx recommendations in 2013-2017. Out of 352 gene drug pairs documented by CPIC, 129 met inclusion criteria. Out of these, 39 pairs had clear CPIC guidelines. Further, 18 drugs from this list were utilized at Nicklaus Children's Hospital. The average number of patients receiving one of these drugs per year has ranged between 22 and 15,033. Thiopurines had clear guidelines and had PGx testing available for *TPMT* gene in the medical health records. *TPMT* gene testing utilization was retrospectively reviewed in 176 patients who received thiopurines as part of their clinical care. Only 36% of these patients had received PGx test prior to medication and 11 patients had a phenotype of "Intermediate Metabolizer" and needed additional monitoring. Our analysis has revealed that different service lines utilize inconsistent workflows. For instance, 75% of patients who received thiopurines in the gastroenterology (GI) service line did not have genetic test done. At the same time, other service lines predominantly tested *TPMT* enzyme activity levels in lieu of a genetic test. There is urgent need to standardize the application of PGx monitoring strategies in clinical practice for medications that pose increased risk of toxicity. In order to evaluate the importance of PGx in pediatric clinical settings, we still need to better understand the rate of toxicities associated with these medications, and the overall cost associated with the toxicities compared to the PGx testing recommended by FDA and CPIC.

651T

Development and validation of the multi-dimensional index of clinical utility: A novel measure to assess value and aid decision-making in genomic medicine. R.Z. Hayeems¹, S. Luca¹, A. Bhatt¹, E. Pullenayegum¹, M.S. Meyn², W.J. Ungar¹. 1) The Hospital for Sick Children, Toronto, Ontario, Canada; 2) UW Center for Human Genomics and Precision Medicine, University of Wisconsin, Madison, WI.

Purpose: Clinical utility is a term used to describe the value of genetic tests, but lacks a specific definition and measurement strategy. While the laboratory performance of genetic testing is improving significantly, policymakers require evidence of clinical value to inform decision-making. This study aims to develop and validate a novel measure of clinical utility for the field of genomic medicine. **Methods:** A literature-derived index of items reflecting on the concept of clinical utility was generated. Semi-structured interviews were conducted with clinicians to refine the concept and provide feedback on the structure, understandability, and importance of each item. Items were revised and grouped into 3 core domains. Clinicians then participated in an online Delphi survey and expert panel to rank item importance, rate clinical sensibility, and establish a scoring algorithm. **Results:** The literature review identified 26 plausible items for the index. Interviews were conducted with 35 clinicians from 9 specialty areas. Providers defined clinical utility as a multi-dimensional concept related to diagnostic thinking, patient management, family-centred care, and system efficiency. With the exception of oncology, there was general agreement across speciality groups on the importance of each item. Less important or redundant items were removed and ambiguously worded items were revised. A total of 113 clinicians completed a Delphi survey, rating the importance and sensibility of a reduced set of 20 index items. Of these clinicians, 45 (39.8%) were genetics professionals (MD or genetic counsellor) and 68 (60.2%) were specialists (e.g., pediatrician, neurologist). The majority agreed that the index items accurately reflected the concept of clinical utility (81.4%) and were clear and easy to understand (75.2%). Item mean rankings and qualitative feedback were synthesized to create a 17-item index with 4 domains, for which a scoring algorithm was established by 10 experts. Reliability and validity testing of the final index is underway. **Conclusion:** The development of a clinician reported outcome measure to capture the concept of clinical utility in genomics is an essential prerequisite for assessing the value of rapidly evolving genetic testing technologies on the cusp of clinical translation. Evidence of this sort will inform reimbursement and implementation decisions related to this complex technology.

652F

Assessment of physician recommendations of warfarin genotype testing in a virtual clinic. B.M. Hollister, K.E. Abdallah, W.D. Kistler, S. Persky, V.L. Bonham. Social and Behavioral Research Branch National Human Genome Research Institute, National Institutes of Health, Bethesda, MD.

Advances in pharmacogenomic research have led to a better understanding of genetic variants that influence an individual's drug response. Currently, there is limited adoption of pharmacogenomic testing in clinical settings; however, warfarin response genotype testing has been under consideration for clinical use since 2007. Due to warfarin's narrow therapeutic range, identifying the optimal therapeutic dose for patients can be challenging and requires extensive monitoring of the patient. Variants in *CYP2C9* and *VKORC1* can be useful in predicting the best starting dose for patients. However, the trials evaluating the implementation of genotype-guided warfarin dosing have yielded mixed results, resulting in limited uptake of genotype testing. Despite investigations exploring the effectiveness of genotype-guided dosing, little is known about the willingness to integrate this testing by physicians. To address this, we used a virtual clinical interaction, which enables the exploration of this inquiry in a controlled setting, to investigate the use of warfarin genotype testing by physicians and their views on this test. We recruited internal and family medicine residents (N=153) from 7 residency programs in the United States. Residents were asked to interact with a virtual patient who was diagnosed with a deep vein thrombosis and expressed interest in genetic testing after receiving a recommendation for warfarin therapy. Following the virtual interaction, we asked the residents about their familiarity with pharmacogenomic testing and, specifically, warfarin genotype testing. Of the 153 residents who participated, none recommended warfarin genotype testing during the interaction with the virtual patient. 38% indicated that they were familiar with warfarin genotype testing after the virtual interaction; however, only one resident had ordered the test previously. 13% indicated that they felt warfarin genotype testing is useful. This lack of familiarity is unlikely due to limited knowledge of basic genetics, as most participants scored at least 70% on the Genetic Variation Knowledge Assessment Index. The absence of recommendation for warfarin genotype testing by residents in our sample suggests that the adoption of testing may be impacted by knowledge of the availability of genotype testing and its perceived utility in clinical practice. These results highlight a need for education and research regarding the implementation of pharmacogenomic testing.

653T

Assessment of the human bioresource in K-NIH for precision medicine.

J.W. Kim, E.J. Kim, J.H. Park, H.H. Kim, J.Y. Jung, J.H. Kim. Korea Institute of S&T Evaluation and Planning, Seoul, South Korea.

Cohort-based and epidemiological information, genomic information, and human-derived resources are essential elements for the development of precision medicine and clinical applications. In Korea, efforts are being made to build a national research resource infrastructure, but the issue of openness and usability of research resources continues to arise, necessitating the reorganization of research resource and the associated management system. This study evaluates the status of the major human resources currently held by the K-NIH and the usefulness of the research in precision medical research. The goal of this research is to identify and analyze the status of the collection, management, and utilization of research resources such as cohort and human resources that have been established by the National Institute for Health Research and to evaluate the utilization value as a precision medical research resource. As a result of the evaluation, it is difficult to use the cohort research resources possessed by National Institutes of Health Genome Center directly in the precision medical researches, considering the purpose of collection, the number of research subjects, and the nature of collected information and data. There is currently a lack of connections with external clinical information such as medical records and health checkup data, therefore the maturation of all necessary conditions needs to be set and met. Although the Ministry of Health's genome resources are produced in a significant amount based on the SNP chip information, it is limited to the use of precision medical care in consideration of the cohort management status, genetic information, and differences in the level of information linked to each platform. Overall, the research resources collected to so far are long-term follow-up cohorts and while the value of epidemiological information is recognized, there are limitations on the types of research resource data provided and lack of linkage with clinical information. For the purposes of precision medical research, a new cohort can be established by including the existing cohort at the National Institute of Health Genome Center. A much higher level is required in terms of the quantity and quality of collected data and human resources than in the past. For precision medical research, the full-scale medical cohort study, linkage with electronic medical record (EHR) should be established.

654F

Attitudes of health professionals about the Cree encephalitis and Cree leukoencephalopathy education and carrier screening program, a population-based carrier screening program in a Canadian first nation.

J. Le Clerc-Blain¹, V. Gosselin², A. Bearskin³, J.E. Torrie⁴, G.A. Mitchell^{1,5,6}, B.J. Wilson⁷, A. Richter^{1,5,6}, AM. Laberge^{1,5,6}. 1) CHU Sainte-Justine Research Center, Montreal, QC, Canada; 2) Formerly Cree Board of Health and Social Services of James Bay, Chisasibi, QC, Canada; 3) Eeyou Awaash Foundation, Chisasibi, QC, Canada; 4) Cree Board of Health and Social Services of James Bay, Mistissini, QC, Canada; 5) Medical Genetics Division, CHU Sainte-Justine, Montreal, QC, Canada; 6) Department of Pediatrics, University of Montreal, Montreal, QC, Canada; 7) Division of Community Health and Humanities, Memorial University of Newfoundland, St. John's, NL, Canada.

Background: A population-based carrier screening program (CSP) was started in 2006 for Cree Encephalitis (CE) and Cree Leukoencephalopathy (CLE), two neurodegenerative autosomal recessive conditions with high carrier rates in Cree communities around eastern James Bay (JB) (Canada). Developed by local health authorities in collaboration with the Eeyou Awaash Foundation (a community family support group), the CE-CLE CSP informs the population about CE-CLE, carrier screening, and reproductive options for at-risk couples. **Objective:** Describe the attitudes of local health professionals (HPs) about the CE-CLE CSP. **Methodology:** HPs working in Cree communities around JB were asked to complete a survey on their knowledge of CE-CLE and attitudes about CE-CLE CSP. Hypothesis-driven statistical analysis was performed to test for associations between HPs' attitudes and their characteristics. **Results:** Response rate was 59.4% (164/276). Participants included nurses (102), general practitioners (19), community health representatives (28) and others (12). HPs' comfort level with providing information about the CSP was associated with their knowledge level ($p < 0.001$), program awareness ($p = 0.018$), and having cared for a family with an affected child ($p < 0.001$), but not with number of years working in JB, job type, family history, or spoken languages. HPs who felt they did their job when someone chose screening and HPs who worried that a couple didn't understand what it's like to have an affected baby when they choose to keep an affected pregnancy were both more likely to perceive a higher pressure to screen ($p = 0.037$ and $p = 0.011$, respectively). Importance given to informed consent for screening was associated with knowledge level ($p = 0.023$), but not with perceived pressure to screen, level of concern about perceived potential screening impacts, or program awareness. Perception of program goals was associated with job type ($p = 0.001$), but not with level of knowledge, having cared for a family with an affected child, program awareness, number of years working in JB, or level of concern about perceived potential screening impacts. **Conclusion:** Our findings suggest that training of HPs who are meant to discuss carrier screening with their patients should focus not only on increasing their knowledge and awareness of the program, but also provide opportunities to meet with families who were affected with the targeted diseases and address perceptions of pressure to screen and program goals.

3063W

CD163: A possible novel candidate gene for Moyamoya disease. L.

Pisani¹, K. Kolla², R. Arora³, A. Blevins⁴, K. Monaghan⁴, J. Juusola⁴. 1) Dept. of Pediatrics, Division of Clinical Genetics, Northwell Health, NY; 2) Comprehensive cancer center, Medstar Georgetown Hospital, D.C; 3) Dept. of Neurology, Northwell Health, NY; 4) GeneDx, MD.

Moyamoya disease (MMD) is a progressive disorder characterized by the stenosis or occlusion of the internal carotid artery bifurcation, resulting in the development of an abnormal collateral vasculature at the base of the brain. Cerebral hemorrhage and strokes are the main clinical features, as well as chronic migraines, visual disturbances, developmental delay and seizures. Its etiology is still largely unknown, though the p.R4859K polymorphism in *RNF213* has been recently identified via GWAS as important risk for MMD in the East Asian population. We report a 25 year-old woman of Brazilian descent with stenosis of the right middle cerebral artery and history of right centrum semiovale ischemia at age 17, followed by development of chronic migraines. She was 29 weeks pregnant with her first child. Family history was remarkable for MMD in a maternal uncle and his 10 year-old daughter, as well a severe stroke and aneurysm in the patient's mother at age 43, and fatal strokes in the 40s-50s in multiple maternal relatives, over three generations. Given the clinical presentation and finding compatible with MMD, the extensive family history, and the urgent need of an action plan for her delivery, Whole Exome Sequencing (WES) was elected. WES revealed a heterozygous p.Q676X variant in *CD163*, maternally inherited. This variant is predicted to cause loss of normal protein function through protein truncation or nonsense-mediated mRNA decay; it has not been observed in large population cohorts, nor has been reported as pathogenic or benign. CD163 is a scavenger receptor expressed on the surface of most monocytes and tissue macrophages and, in its soluble form, in the serum. Amongst its functions, it has been demonstrated that CD163 is involved in the adhesion of monocytes to the endothelial cells, and C163+ macrophages have been demonstrated in tissues actively regenerating after an ischemic injury. CD163 has also proven to be a mediator of the late down regulation of both acute and chronic inflammation. While not all the functions nor ligands of CD163 have been completely defined, evidence suggests that this protein plays a role in the pathogenesis of MMD, as in a recent study elevated levels of soluble CD163 have been reported in affected individuals, significantly higher than those of controls, irrespective of the patients *RNF213* genotype. Further testing of family members and research is necessary to better clarify the role of *CD163* in the pathogenesis of MMD.

3064T

Novel risk genes and mechanisms implicated by exome sequencing of 2756 individuals with pulmonary arterial hypertension. N. Zhu^{1,2}, M. Paucilo^{3,4}, K. Lutz^{3,4}, C.L. Welch¹, J. Wang^{1,2}, C. Gonzaga-Jauregui⁵, J.G. Reid⁶, J.D. Overton⁵, A. Baras⁵, A. Walsworth^{3,4}, Y. Shen^{2,6}, W.K. Chung^{1,7,8}, W.C. Nichols^{3,4}.

1) Department of Pediatrics, Columbia University Medical Center, New York, NY; 2) Department of System Biology, Columbia University, New York, NY; 3) Division of human genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 4) Department of Pediatrics, University of Cincinnati, College of Medicine, Cincinnati, OH; 5) Regeneron Genetics Center, Regeneron Pharmaceuticals, Tarrytown, NY; 6) Department of Biomedical Informatics, Columbia University, New York, NY; 7) Herbert Irving Comprehensive Cancer Center, Columbia University Medical center, New York, NY; 8) Department of Medicine, Columbia University Medical Center, New York, NY.

Pulmonary arterial hypertension (PAH) is a rare disease characterized by remodeling of pulmonary arterioles leading to increased arterial pressure and resistance, right ventricular hypertrophy and heart failure. Despite clinical advances, the long-term prognosis remains poor for many PAH patients. Genetics play an important role in the pathogenesis of PAH, and there are eleven known risk genes underlying ~70% of familial cases (FPAH) and ~10% of sporadic idiopathic cases (IPAH). However, the majority of idiopathic cases, as well as PAH associated with other diseases (APAH) remain genetically undefined. To identify new variants in known risk genes, we performed targeted and exome sequencing in 2756 PAH cases from the US PH Biobank. The cohort included 1124 IPAH, 1225 APAH (58.9% connective tissue disorders, 21.8% congenital heart disease and 19.3% other), and 97 FPAH cases. Most of the patients were adults (mean age of onset = 38.6 years) and, as expected for this disease, the female: male ratio was 4:1. We observed that 278 cases (10% overall: 12.7% IPAH, 5% APAH, 76.3% FPAH) carried rare predicted pathogenic variants in at least one of the 11 known risk genes (*ACVRL1*, *BMPR1A*, *BMPR1B*, *BMPR2*, *CAV1*, *EIF2AK4*, *ENG*, *KNCK3*, *SMAD4*, *SMAD9*, *TBX4*). Eight patients carried rare deleterious variants in two risk genes. Of note, 114/235 (48.5%) of the variants in known risk genes are novel. To identify new risk genes, we performed gene-based rare variant association tests using 1832 unrelated European cases and 12,927 European controls (7509 gnomAD, 5418 Columbia controls). The frequency of rare synonymous variants was identical in cases and controls, supporting negligible batch effects. Using a variable threshold of predicted deleteriousness for missense variant inclusion, we identified *KLK1* (kallikrein-1, $p=1.30E-07$) as a new candidate risk gene. For IPAH (813 cases), *KLK1* ($p=2.00E-08$, 10 carriers) and *GGCX* ($p=9.02E-07$, 12 carriers) reached genome-wide significance. *KLK1* is expressed in pulmonary arterial cells and has an established role in vascular remodeling and has not previously been associated with PAH. Other top-ranked associations included an enrichment in SMAD3 transcription targets (*RCVRN*, *MYO5B*, *NOTCH4*, *FLNA*, *PAX6*, *NR3C1*, *SMAD5*; $p=0.0009$) and calcium ion binding regulation (*PNLIPRP1*, *SPARC*, *NOTCH4*, *MAN1B1*; $p=0.009$). These data increase the number of known risk genes and suggest new mechanisms for IPAH that could provide targets for therapy.

3065F

A new form of inherited cardiomyopathy leads to changes in heterochromatin structure in man and mice. *N. Abdelfatah¹, R. Chen², D. Belke¹, C. Fielding³, K. Ito³, J. Gross², H. Duff¹, B. Gerull^{1,2}.* 1) Department of Cardiac Sciences and Libin Cardiovascular Institute, University of Calgary; 2) Comprehensive Heart Failure Center and Department of Medicine I, University Hospital Würzburg, Würzburg, Germany; 3) Transgenic Services, Centre for Genome Engineering, Cumming School of Medicine, University of Calgary.

Introduction: Nuclear envelope proteins have been shown to play an important role in the pathogenesis of inherited dilated cardiomyopathy (DCM). We recently discovered a novel homozygous mutation in the LEM domain of the inner nuclear membrane (INM) protein LEMD2 causing juvenile onset of arrhythmic cardiomyopathy and bilateral cataract in the Hutterite population. Previous studies on patient fibroblasts suggested that the mutation may cause a proliferation defect and cell senescence. In addition, myocyte disarray and extensive interstitial fibrosis along with extensive changes of myocyte nuclei were observed in cardiac tissue of a deceased patient. **Purpose:** To understand the role of LEMD2 in the heart using an in vivo knock-out (KO) and knock-in (KI) mouse model. **Methods and Result:** In vivo homozygous KO, as well as KI mouse models, were created using CRISPR-Cas9 system. Survival studies were accomplished for both lines at different embryonic stages. KO mice died between embryonic day 10.5 and 11.0 and were smaller in size. Histology showed ventricular wall thinning and abnormalities of the outflow tract development. On the other hand, most homozygous KI mice survived until 3 weeks of age before they died due to severe heart failure. Histology already at E13.5 and P21 demonstrated ventricular wall thinning and dilatation of the right and left ventricle with extensive cardiac remodeling and fibrosis. Furthermore, TEM of hearts from both lines revealed remarkable changes of myocyte nuclei similar seen in cardiac tissue from patients. Next, echocardiography was used to determine cardiac function at 3 weeks old KI mice. Homozygous KI mice showed significantly decreased fractional shortening and ejection fraction compared to wildtype littermates (9.39 ± 5.6 vs. 23.427 ± 2.2 , $p < 0.0001$ for FS (%); 20.322 ± 11.4 vs 47.345 ± 3.7 , $p < 0.0001$ for EF). Homozygous hearts also showed dilatation and wall thinning indicated by decreased value of LVIDd ($p < 0.01$) and IVSd ($p < 0.01$), respectively. The phenotype represents the human homolog of dilated cardiomyopathy in mice. **Conclusion** Knock-out of LEMD2, as well as a mutated LEM domain in mice, lead to remarkable changes in nuclei with abnormal heterochromatin formation as seen in humans. Homozygous mice of both models show structural and functional abnormalities during cardiac development suggesting the involvement of LEMD2 in this process. Future work will determine the underlying mechanisms causing cardiomyopathy in mice and man.

3066W

Characterization of Egyptian genetic variation in relation to cardiovascular phenotype. *Y. Aguib^{1,2}, M. Allouba^{1,2}, R. Walsh², M. Elkhateb¹, N. Shehata^{1,2}, A. Afify¹, S. Halawa^{1,3}, R. Govind², N. Whiffin², B. Sami¹, D. Yehia¹, S. ElShorbagy¹, E. Abdel Rahman¹, W. ElMozy¹, A. Essam¹, M. Maghawry¹, S. Lee², S. Cook^{2,4}, P. Barton², J. Ware², H. Aguib^{1,2}, A. ElGuindy^{1,2}, M. Yacoub^{1,2}.* 1) Aswan Heart Centre, Magdi Yacoub Foundation, Aswan, Egypt; 2) Imperial College London, UK; 3) American University in Cairo, Egypt; 4) National Heart Centre Singapore, Duke-National University of Singapore, Singapore.

Background & Objectives: Genetic and phenotypic databases for variant interpretation are key for the development of precision medicine, taking ethnicity into consideration. The Egyptian and North African population is not represented in the Exome and Genome Aggregation Consortia (ExAC and gnomAD), which are the most comprehensive population data sets, posing a major challenge in data interpretation. To address this, a study of 1000 Egyptian healthy volunteers (E-HVols) was initiated. **Methods:** Recruited E-HVols undergo cardiovascular phenotyping (CMR, ECG and ECHO). Genetic make-up is studied using a panel of 174 genes implicated in Inherited Cardiac Conditions (ICC). Variant calling is performed using a custom designed bio-informatics pipeline based on the Genome Analysis ToolKit and variants are annotated using CardioClassifier. Results are compared to existing reference datasets. CMR studies are conducted to investigate geometry, dynamism and function in the different heart chambers, aorta and pulmonary. Computerised algorithms were developed for detailed morphology analysis of healthy aortic and ventricular structure. Feature Tracking for LV motion analysis and computational modelling were used to capture and quantify dynamic changes of both shape and flow and to identify disease-based variations in patients. Also, circulating protein biomarkers (serum/plasma) are assessed. **Results:** A database comprising clinical data, genetic variation in ICC genes and imaging data derived morphological parameters was established. 2% among self-declared E-HVols were diabetic (excluded), 22.3% had a BMI >30. NGS Data (408/650) obtained provide an ethnically-matched control data set and are utilized for better interpretation of variants in our clinical setting. In core HCM/DCM genes, for instance, novel rare variants were identified in E-HVols, including a truncating variant in *TTN* and in 9 Individuals homozygosity was identified (28 core HCM/DCM analysed). Long-term clinical follow-up will reveal the impact of identified variants and can impact their classification. For 200 E-HVols demographic and anthropometric data, advanced imaging combined with 3D modelling and novel computerized analysis were analysed and used as a powerful tool to define geometry and function of the aortic root and aorta. **Conclusions:** Expanding the number of this initial sample will enhance the clinical utility of our data and their integration in precision medicine approaches.

3067T

Defining the genetic architecture of hypertrophic cardiomyopathy in the Egyptian population.

M. Allouba^{1,2}, Y. Aguib^{1,2}, R. Walsh², R. Govind², N. Whiffin², S. Halawa^{1,3}, A. Maher¹, A. Afify¹, A. Galal¹, R. Buchan², H. Kassem^{1,4}, B. Samy¹, R. Azer¹, M. Elkhateb¹, F. Mohamed¹, A. Elguindy^{1,2}, S. Cook^{1,5}, P. Barton², J. Ware², M. Yacoub^{1,2}. 1) Aswan Heart Centre-Magdi Yacoub Foundation, Aswan, Egypt; 2) Imperial College London, UK; 3) The American University in Cairo, Egypt; 4) Faculty of Medicine Alexandria University, Egypt; 5) National Heart Centre Research Institute Singapore, Duke-National University of Singapore.

Introduction: Hypertrophic cardiomyopathy (HCM) is a heterogeneous disease with regards to phenotype, genotype and mode of inheritance. It affects 1:500 individuals in the general population and appears to be influenced by ethnicity. In Egypt, HCM constitutes a substantial burden of disease, yet, its genetic determinants have not been adequately studied. **Aim:** To analyse the genetic architecture of HCM in the Egyptian population in comparison to a Caucasian HCM cohort as well as to global and ethnically-matched healthy controls. **Methods:** In this work, genetic data from 500 HCM patients and 408 Egyptian healthy Volunteers (EHVOLs) are studied using the Illumina TruSight Cardio Sequencing Kit comprising 174 genes with reported roles in Inherited Cardiac Conditions (ICC). **Results:** Currently, 1098 Egyptian patients, with a mean age of 34.9±18, have been recruited to the National HCM Program at our hospital. 285 patients have underwent myectomy with a 3-year survival rate of 98.9%. Analyzed data reveal unique features of the Egyptian HCM cohort. A novel truncating variant in *MYH7* was identified among 8 Egyptian patients. This significantly enriched ($p < 0.0005$) variant in the clinical cohort over local controls as well as its absence in ExAC suggests that it could be reclassified from a “variant of uncertain significance” to likely “pathogenic” as per the guidelines by the American College of Medical Genetics and Genomics (ACMG). Induced pluripotent stem cell (iPSC)-derived cardiomyocytes are being generated to investigate the level of expression of the truncated variant relative to the wild type allele. Its functional role remains to be studied. In addition, a set of 16 core HCM genes was selected for the analysis of the frequency of homozygous variants among Egyptian patients. 6 and 9 patients harbored rare and common homozygous variants in the selected genes, respectively. This reflects the high rate of intermarriages in this population. **Conclusions:** The genetic architecture of HCM in the Egyptian population shows several unique features with regards to the novel truncating variant in *MYH7* and homozygosity in core HCM genes. This findings could provide new mechanistic insights into the pathogenesis of HCM and improve the efficacy of current treatment strategies.

3068F

Identification of genetic defects in phenotypically defined subtypes of patients with Ebstein anomaly.

R. Cabrera¹, M.C. Miranda-Fernández¹, C.J. Hernández-Toro¹, V.M. Huertas-Quiñones^{1,2,3}, P. Laissue³, S. Ramírez-Oyaga¹, M. Carreño¹, I. Pineda¹, C.M. Restrepo², C.T. Silva³, A. García^{1,3,4,6}, N. Sandoval^{1,3}, M. Barrera-Castañeda¹, K.J. Moreno Medina¹, R.J. Dennis Verano^{1,5}. 1) Fundación Cardioinfantil - Instituto de Cardiología, Bogotá, DC, Colombia; 2) Universidad Nacional de Colombia, Bogotá, Colombia; 3) Universidad del Rosario, Bogotá, Colombia; 4) Universidad El Bosque, Bogotá, Colombia; 5) Pontificia Universidad Javeriana, Bogotá, Colombia; 6) Universidad Militar Nueva Granada, Bogotá, Colombia.

Objectives Ebstein Anomaly (EA) is a rare congenital heart defect (CHD) with a poorly characterized genetic etiology and heterogeneous clinical presentation. As a reference institution for complex CHD, we have assembled one of the largest single center cohorts of patients with EA. We hypothesize that phenotypic diversity in Ebstein Anomaly patients reflects an underlying heterogeneity in genetic causes, and that major genetic causes can be found in at least some patient subgroups. **Methods** A comprehensive cross-sectional phenotypic characterization of 147 EA patients from one of the main referral institutions for CHD in Colombia was carried out. The most prevalent comorbidities and distinct subgroups within the patient cohort were identified through multivariate cluster analysis. Patients from the most homogeneous phenotypically-defined subtypes were submitted to genome-wide molecular analyses (Whole Exome Sequencing and/or cytogenetic microarray analysis). Relevant variants were identified by gene function filtering and population probability analysis. **Results** The data show that EA is a heterogeneous disease, frequently associated with cardiovascular and noncardiovascular comorbidities. Patients with WPW and SVT represent a homogeneous subgroup that presents with a less severe spectrum of disease and better survival when adequately managed. Similarly, patients with a syndromic presentation also represent a distinct subgroup. Whole exome sequencing can identify variants in known and novel EA-associated genes that are highly likely to be causative in a significant number of familial and isolated cases of the WPW/SVT subtype of EA patients. Some of these genes have been implicated in the transcriptional regulation of the Endothelial-Mesenchymal Transition underlying Tricuspid valve delamination. Furthermore, syndromic patients are likely to show chromosomal abnormalities, and two novel chromosomal regions deleted in syndromic patients with EA were identified. Genetic causes in non-WPW/SVT/nonsyndromic EA cases remain to be identified. **Conclusions** Multivariate analysis can reveal clinically relevant patterns in comorbidity datasets of phenotypically heterogeneous disease cohorts. This analysis can identify relevant subtypes of EA patients and empowers the identification of genetic variants. This approach can be used in the identification of novel causal genes in CHD and other rare diseases.

3069W

The phenotype of a CRISPR mouse model of a human gap junction C2 (connexin 47) mutant causing lymphedema. R.P. Erickson¹, R. Kyiat¹, D.J. Mustachich¹, M. Bernas², M. Witte¹. 1) University of Arizona, Tucson, AZ; 2) TCU and UNTHSC School of Medicine, Ft. Worth, Texas.

Ferrell, et al (doi 10.1016/j.ajhg.2010.04.010) reported 2 mutations of *GJC2* (*CX47*) in 2 families of dominantly inherited lymphedema with variable penetrance and expressivity while Ostergaard, et al (doi 10.1036/jmg.2010.085563) soon added 2 more lymphedema-causing mutations in *CX47* families with full penetrance and great variation in expressivity. We sought to make a mouse model of lymphedema caused by *Cx47* mutations using the CRISPR approach to create mice carrying one of these mutations, the R260C mutation. Due to differences in sequence, the human R260 is at position R257 in mice, and, due to codon usage differences, the mutation requires a 2 bp change in mice. Three guides were selected and produced by PCR using as a template an Addgene pX330 plasmid carrying the scaffold portion of the guide. SgRNAs (single guide RNAs) were designed to target R257 flanking sequences in the *Cx47* gene. Microinjection was performed by continuous flow injection of the Cas9/gRNA/ssODN mixture into the pronucleus of 1-cell C57BL/6J (BL6) zygotes. In addition to studying heterozygotes and homozygotes on the BL6 background, an F₁ was made between male BL6 carriers of the mutation and females of the 129/J and A/J strains. Heterozygous offspring were backcrossed to their mothers to create an N₁ generation with 75% of the new genetic background. Evans blue dye (EBD) (~50ul of 5 g/dl), was injected intradermally into all four paws, ears and snout serially to sequentially highlight the regional lymphatic systems and the central collection system which were examined by sequential dissection of the peritoneal cavity (iliac and cysterna chyli), thoracic cavity (thoracic duct), axillary, jugular, popliteal and sacral regions. Alterations in the lymphatic system were only found (and confirmed by blinded observation) in homozygous mutant mice on the BL6 background, consisting of an increase in the number of lymph nodes in the jugular region, as well as an increase in the number and/or size of lymph nodes in the iliac region. We detected no phenotypic alterations in heterozygotes on the other 2 (75%) backgrounds and no lymphedema was identified in any of the mice. Thus, the phenotype was not obviously lymphedema-causing but altered the lymphatic architecture by creating extra lymph node niches. These results confirm a role for Connexin 47 in lymphatic development but also indicate differing phenotypic consequences between man and mouse.

3070T

SMAD4 rare variants in individuals and families with thoracic aortic aneurysms and dissection. D. Guo¹, X. Duan¹, E. Regalado¹, J. Cosell^{2,3}, A. Estrera⁴, H. Safi⁵, M. Bamshad^{6,6}, D. Nickerson⁵, S. LeMaire^{2,3}, J. Backer⁷, D. Milewicz¹, Montalcino Aortic Consortium; University of Washington Center for Mendelian Genomics, Seattle, USA. 1) Department of Internal Medicine, The University of Texas Health Science Center at Houston McGovern Medical School, Houston, TX. 1) Department of Internal Medicine, The University of Texas Health Science Center at Houston McGovern Medical School, Houston, USA; 2) Michael E. DeBakey Department of Surgery, Baylor College of Medicine, Houston, USA; 3) Department of Cardiovascular Surgery, Texas Heart Institute, Houston, USA; 4) Cardiothoracic Vascular Surgery, The University of Texas Health Science Center at Houston McGovern Medical School, Houston, USA; 5) Department of Genome Sciences, University of Washington, Seattle, USA; 6) Department of Pediatrics University of Washington, Seattle, USA; 7) Center for Medical Genetics, Department of Cardiology, Ghent University Hospital, Ghent, Belgium.

Pathogenic variants in a number of genes lead to heritable thoracic aortic disease (HTAD) have been identified and the majority of these genes encode proteins involved in smooth muscle cell contraction and adhesion to the extracellular matrix or the transforming growth factor beta signaling pathway. In particular, *SMAD4* pathogenic variants cause juvenile polyposis (JPS) and hereditary hemorrhagic telangiectasia (HHT), and 40% of these cases also have thoracic aortic disease. *SMAD4* variants have not been reported in thoracic aortic disease families without JPS-HHT and thus far the variants reported are loss of function or missense variants located in the MH2 domain of the protein. Analysis of exome sequencing data from 223 unsolved families with HTAD identified a rare variant in *SMAD4*, c.290G>T (p.Arg97Leu), that segregated with aortic disease in family TAA281. In this family, the proband and her brother had ascending aortic dissections at the age of 24 and 41 years, respectively, and their father had an ascending aortic aneurysm repair at age 75; none of them had JPS-HHT. This *SMAD4* variant, located in the MH1 domain, is absent in population databases, predicted to be damaging, and has a CADD score of 33. Cellular studies using the HEK293T and an immortalized smooth muscle cells (SMC) lines revealed that the *SMAD4* p.Arg97Leu alteration increased *SMAD4* ubiquitination and 26S proteasome-mediated protein degradation. SMCs infected with lentivirus expressing the *SMAD4* p.Arg97Leu variant demonstrated reduced expression of SMC contractile proteins when compared to SMCs infected with wildtype *SMAD4*. In addition, exome sequencing data of 355 individuals with dissections under the age of 56 years and no family history or syndromic features identified two other rare nonsynonymous *SMAD4* variants: p.Met24Val identified in a 37 year old patient with type A aortic dissection and p.Pro246Thr in a 54 year old patient with type A aortic dissection, and no features of JPS-HHT. These variants have very low minor allele frequencies in the gnomAD database and CADD scores of >15. These findings support the conclusion that *SMAD4* rare missense variants disrupting *SMAD4* cellular function can lead to thoracic aortic disease in individuals who do not have JPS or HHT.

3071F

Genetic variants in familial abdominal aortic aneurysms. A. Ijzma¹, D. Heijlsman², H.T. Bruggenwirth², D. Majoer-Krakauer². 1) Erasmus Medical Center, Clinical Bioinformatics Unit, Department of Pathology, Rotterdam, Netherlands; 2) Erasmus Medical Center, Department of Clinical Genetics, Rotterdam, Netherlands.

Introduction: Abdominal Aortic Aneurysm (AAA) has a prevalence of 5% in the elderly population. An AAA occurs when the aorta below the renal arteries expands to a diameter of 3cm or more. In the Netherlands, each year approximately 5000 AAA patients are hospitalized and around 750 people die due to AAA rupture. **Hypothesis:** Approximately 20% of AAA patients are familial and our hypothesis is that genetic predisposition is a significant cause for abdominal aneurysm pathology. Our goal is to identify the genes that play a role in the formation of AAA. **Methods:** Our study population consists of approximately 1250 AAA patients. So far we sequenced 548 of these AAA patients. Complete Genomics whole genome sequencing (WGS) was performed in 3 families (15 individuals) and whole exome sequencing (WES) on the illumina platform using Agilent Haloplex and CRE sureselect exome capturing technology was performed in 71 families (175 individuals) and 358 single familial AAA patients. Burden analysis was used to identify genes enriched in our AAA population. **Results:** We present the detailed workflow of the analysis of the genomics data. We will discuss several candidate genes identified such as the enrichment for variants we identified in the COL4A2 gene. **Conclusions:** In 118 out of 512 families a variant in a diagnostic AAA gene was found. Analysis of all genes in the exome dataset led to the identification of several candidate genes that show variants in more than one AAA family and that have not been linked to AAA before.

3072W

Investigation of the molecular basis of systemic to pulmonary arteriovenous malformations. R.P. Martin¹, D. Valle¹, H. Dietz¹, N. Sobreira¹, S.E. Mitchell¹. 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Department of Radiology, Johns Hopkins University School of Medicine, Baltimore, MD.

Arteriovenous malformations of the pulmonary vascular (PAVM) bed are anomalies characterized by vascular shunts in normally ventilated lung tissue. PAVM with a systemic source is a rare condition and considered to be a specific entity (SPAVM). Dyspnea and the classical triad of cyanosis, clubbing of the digits and polycythemia, as a result of undersaturation in conventional PAVM, are not present in SPAVM systemic because it is a left-to-right shunt. Because of lack of symptoms, the discovery of these malformations is often an accidental finding. Most patients are young asymptomatic adults, predominantly male. SPAVMs are subject to systemic pressure, and so may lead to pulmonary hypertension or may grow and rupture. The molecular basis of SPAVM remains unknown. As part of the Baylor-Hopkins Center for Mendelian Genomics project, we performed whole exome sequencing on 6 unrelated probands with SPAVM. One proband was part of a multiplex family with 9 affected individuals from 3 generations presenting with SPAVM and/or aortic root aneurysm in early age. Using PhenoDB analysis tool, we selected the heterozygous and homozygous, rare (MAF<1%) and functional variants in each proband. These genes and variants were evaluated for their ClinVar, HGMD, OMIM, GWAS, mouse model phenotypes and other annotations to identify the best candidate variants in each proband. Genes that were mutated in 3 or more probands were also selected. None of the probands had pathogenic variants in *ENG*, *ACVRL1*, *SMAD4*, *GDF2*, or *RASA1*. One proband had a rare, heterozygous variant in exon 8 of *EPHB4* (c.G1448A:p.R483Q). Amyere et al. (2017) described loss of function variants as the cause of multifocal capillary malformations (CM) associated with arteriovenous malformations (AVM), a phenotype that mimics *RASA1*-related CM-AVM1 and also hereditary hemorrhagic telangiectasia. Vivanti et al. (2018) described loss of function variants in *EPHB4* as responsible for vein of Galen aneurysmal malformation. The *EPHB4*-R483Q variant identified in one of our probands was reported in only 16 of 224462 alleles in gnomAD and is likely the cause of the SPAVM phenotype in this proband expanding the phenotype associated with pathogenic variants in this gene. Furthermore, we are also analyzing the genes mutated in at least 3 unrelated probands to possibly identify a novel candidate gene responsible for SPAVM. We suggest that SPAVM is a rare distinct phenotype with a still unknown molecular basis.

3073T

The novel RNF213 variant observed in patient with moyamoya and systemic artery stenosis diseases. O. Migita^{1,2}, M. Mizuno¹, M. Ozawa¹, K. Aso¹, H. Yamamoto¹, K. Hata². 1) St. Marianna University School of Medicine, Kawasaki-shi, Miyamae-ku, Kanagawa, Japan; 2) National Research Institute for Child Health and Development, Tokyo, Japan.

The *ring finger protein 213* (*RNF213*) gene was reported as a susceptible gene for moyamoya disease and intracranial major artery stenosis. According to ClinVar, which is curated database of published variants, 35 registered variants had been observed as risk, pathogenic variants or variants of uncertain significance. But the role of *RNF213* in moyamoya diseases is still unclear. A three year old boy presented with progressive mitral valve regurgitation. He was born to unconsanguineous parents. When he was first year old, he was referred to St. Marianna University Hospital by patient local physician because of his heart murmur which was found in his health check-up. He was received reconstructive surgery of mitral valve, because the regurgitation had been worsen. But four day after the operation, he had seizure and cerebral infarction. He had extensive cerebral infarction in the MCA, ACA area, and narrowing the basilar arteries. The possibility of moyamoya Disease and hereditary vascular diseases was suggested. To address to hereditary vascular disease, whole exam sequencing was employed to search for pathogenic mutations, and revealed that he had *RNF213* novel mutation. Subsequently, systemic MRI imaging showed that his artery disease were not showed only intracranial artery, but also renal artery, iliac artery and other systemic arteries. Further research should be needed to reveal the role of the *RNF213* variants in pathogenic machinery of moyamoya disease. This result of the present case may provide new insight for association between *RNF213* gene mutation and moyamoya disease.

3074F

Preliminary variant classification results from the Dilated Cardiomyopathy (DCM) Precision Medicine Study. A. Morales¹, D.D. Kinnamon¹, E. Jordan¹, J. Platt³, M. Vatta^{4,5}, M. Dorschner⁶, R.E. Hershberger^{1,2}. 1) Division of Human Genetics, Department of Internal Medicine, The Ohio State University, Columbus, OH; 2) Division of Cardiovascular Medicine, Department of Internal Medicine, The Ohio State University, Columbus, OH; 3) Stanford Center for Inherited Cardiovascular Disease, Stanford University, Palo Alto, CA; 4) Department of Medical and Molecular Genetics, Indiana University, Indianapolis, IN; 5) Invitae, San Francisco, CA; 6) Department of Pathology, University of Washington, Seattle, WA.

The DCM Precision Medicine Study, funded to recruit 1300 families of European and African Ancestry (EA and AA), as well as of Hispanic ethnicity (HE), aims to define the genetic basis of DCM. For the study, which involves returning results after analysis of 37 DCM genes, variant classification rules were developed based on the published *MYH7* ClinGen Cardiomyopathy Expert Panel (*MYH7* EP) approach. Modifications were made to 9/33 (27%) criteria evaluated by the *MYH7* EP, including use of hotspot data for *RBM20*, consideration of recessive DCM, and the development of evidence levels for variants encoding a premature stop of translation. Other modifications leveraged our DNA analysis pipeline, including automated confidence interval calculations for minor allele frequencies <0.05% (applied to all target genes), computational data predictions using REVEL scores, and a stand-alone benign classification for *TTN* missense variants. Having medical records to validate study participant phenotypes, *de novo* criteria required demonstrating parental absence of a phenotype. Similarly, the criterion for benign impact regarding concern for an alternate molecular basis for disease was deemed not applicable due to probands' phenotypes expected to be within the spectrum of the study genes. Variants that were of low-quality, synonymous, or outside canonical splice sites were not analyzed. Curation proceeded for 79 variants based on 27 criteria in 98 clinically confirmed probands (59 non-HE EA; 34 non-HE AA; 5 HE EA), revealing 45% with variants of uncertain significance (VUSs) and 11% with likely pathogenic (LP) variants, the latter found mostly (n=9) in non-HE EA. Multiple variants were identified in 19%. Variants reached LP classifications based on case and segregation data, rarity in reference populations, predicted loss of function, other known pathogenic missense change affecting the same codon, and hotspot location. While 9 of our 27 criteria (33%) depend on case or family data, none of the variants met *de novo* and strong segregation criteria. In conclusion, using a new variant classification approach tailored for DCM, approximately half of the probands had VUSs identified, most with LP variants are EA, and 19% have multiple variants. Our results not only highlight the value of multiple family units but also illustrate the challenges in variant interpretation for genetically heterogeneous conditions, warranting more research to move ahead with VUS resolution for DCM.

3075W

Autosomal recessive cutis laxa syndrome with cardiovascular involvement. H. Morisaki^{1,2}, M. Horiuchi³, M. Hibino⁴, Y. Saka⁵, R. Seki⁶, K. Kosaki⁷, T. Morisaki^{2,8}. 1) Dept Medical Genetics, Sakakibara Heart Institute, Fuchu, Tokyo, Japan; 2) Dept Bioscience and Genetics, National Cerebral and Cardiovascular Center, Suita, Osaka, Japan; 3) Div Perinatology and Gynecology, National Cerebral and Cardiovascular Center, Suita, Osaka, Japan; 4) Dept Cardiac Surgery, Nagoya University Graduate School of Medicine, Showa, Nagoya, Japan; 5) Dept Cardiovascular Surgery, Nagoya Daini Red Cross Hospital, Showa, Nagoya, Japan; 6) Dept Cardiovascular Internal Medicine, Sakakibara Heart Institute, Fuchu, Tokyo, Japan; 7) Ctr for Medical Genetics, Keio University School of Medicine, Shinjuku, Tokyo, Japan; 8) Div Molecular Pathology, Inst of Medical Science, The University of Tokyo, Minato, Tokyo, Japan.

Autosomal recessive cutis laxa syndrome type 1 (ARCL1) is a rare connective tissue disorder, characterized by generalized cutis laxa with pulmonary emphysema and/or vascular complications. ARCL1 is classified into three types (ARCL1A, ARCL1B and ARCL1C) according to responsible genes, fibulin-5 (*FBLN5*), fibulin-4 (*FBLN4/EFEMP2*) and *LTBP4*, respectively. The clinical severity of ARCL1 is highly variable, ranging from perinatal lethality to limited cardiovascular / respiratory involvement. Here, we report three ARCL1 cases with distinctive clinical features. The first case was a female fetus with *FBLN4* biallelic mutations presenting hydrops fetalis at 31st weeks gestation. Fetal echocardiography at 32wk revealed severe TR, MR, PR, with a highly tortuous, elongated and narrowed descending aorta. Autopsy revealed severe tortuosity of total aorta with thickened wall and narrow lumen, along with kinky aneurysm in subclavian and iliac arteries. The second case was a 4-month-old boy with compound heterozygous *FBLN4* mutations. Echocardiography revealed severe aortic regurgitation, dilatation of ascending aorta with isthmal narrowing and poststenotic dilatation. Brain MRI showed sporadic small infarction in addition to severe tortuosity in vertebral-basilar arteries. He presented general joint hypermobility and mild cutis laxa, though he didn't show any other characteristic skeletal features. Dilatation of ascending aorta was progressive and reached 44mm at 2y8m, followed by a successful surgical repair at 2y9m. The third case was a 26-year-old male with *LTBP4* mutations. He suffered asthma from early infancy and bone fractures in an elbow and fingers in his childhood. He felt shortness of breath at age 24 and visited our hospital. He presented marfanoid features with arachnodactyly, joint hypermobility, a characteristic face with retrognathia, bifid uvula and downslanting eyes, and redundant skin with sagging cheeks. Echocardiography revealed severe MR due to prominent MV prolapse with degenerative valves. Thoracic-abdominal-pelvic CT revealed severe emphysema and multiple bladder diverticula. He was successfully treated by mitral valve reconstruction and tricuspid valve annuloplasty to reduce MR/TR. Cutis laxa syndromes are highly heterogeneous with respect to organ involvement and severity. Since ARCL1 often accompanies cardiovascular complications, genetic study will help earlier diagnosis for presymptomatic cases to prevent further cardiovascular events.

3076T

Congenital mydriasis in the multisystemic smooth muscle dysfunction syndrome: A harbinger of progressive vasculopathy, expanding the phenotype of the ACTA2 R179C mutation. J.N. Peeden^{1,2}, R. Booher¹. 1) Medicine, East Tennessee Children's Hospital, Knoxville, TN., USA; 2) University of Tennessee Department of Medicine.

We report a patient who has the Multisystemic Smooth Muscle Dysfunction Syndrome (MSSMD), in order to draw attention to the progressive vasculopathy associated with this disorder and to expand the phenotype of the ACTA2 R179C mutation. The cardinal findings of which are present at birth. Born at term, He had hydronephrosis upon prenatal ultrasonography. He was noted to have a massively enlarged bladder and severe right hydronephrosis. He had no reflux. He would not void spontaneously. Upon cystoscopy, a rudimentary urethral valve was resected. He required intermittent catheterization to achieve bladder emptying. A small patent ductus arteriosus was confirmed on echocardiography. Ophthalmology consultation confirmed congenital mydriasis. Pupils were 7 to 7 ½ mm and non reactive. At six months, subclinical hypoxia was noted. Chest x-ray showed findings consistent with bronchopulmonary dysplasia. At cardiac catheterization systemic pulmonary artery pressure was found. There was a large PDA without evidence of coarctation. During surgery to repair the PDA, tissue was so friable and would not hold suture or bovine patch. Prolonged cross clamping of the aorta was required to achieve hemostasis, thereafter the patient developed spastic paraplegia. Other than paraplegia, the patient was normal neurologically. At age 5 he has an episode of left arm shaking that could not be stopped with pressure. EEG was normal. At age 8 he developed altered mental status. Multifocal strokes were present upon MRI, along with grossly abnormal neurovasculature. Multiple acute cortical infarctions due to hypoperfusion/infarction were found. MRA revealed a moyamoya pattern with fusiform aneurysmal dilation of bilateral petrous internal carotid arteries. He underwent bilateral encephalo-duro-arterio-myo-synangiosis successfully. He was found to have a de novo ACTA2 c. 535C>T in exon 6, leading to p.R179C amino acid substitution. Although the repair of PDA in children with ACTA2 mutations and the MSSMD syndrome have been successful, fatalities have occurred during surgery. Our patient survived, with significant neurological injury. Vascular friability, is we believe more likely with the R179C mutation. The knowledge then of the precise mutation (R179C) would identify a child at risk for surgical complications and progressive neurovascular disease; moreover a longitudinal review of our patient's clinical course expands the phenotype of this disorder.

3077F

Further defining the extent of pleiotropy for genes with *de novo* pathogenic variants predisposing to Moyamoya disease, intellectual disability, and congenital heart disease. A. Pinard¹, S. Guey², D. Guo³, S. Wallace⁴, N. Kharas⁵, E.S. Regalado¹, E.M. Hostetler¹, A.Z. Sharrief⁶, M.J. Bamshad⁵, D.A. Nickerson⁵, E. Tournier-Lasserre^{2,6}, D.M Milewicz¹, University of Washington Center for Mendelian Genomics. 1) Department of Internal Medicine, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA; 2) Inserm UMR-S1161, Génétique et physiopathologie des maladies cérébro-vasculaires, Université Paris Diderot, Sorbonne Paris Cité, Paris, France; 3) Department of Neurobiology and Anatomy, The University of Texas Health Science Center at Houston McGovern Medical School, TX 77030, USA; 4) Department of Neurology, The University of Texas Health Science Center at Houston McGovern Medical School, TX 77030, USA; 5) Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA; 6) AP-HP, Service de génétique moléculaire neurovasculaire, Centre de Référence des Maladies Vasculaires Rares du Cerveau et de l'œil, Groupe Hospitalier Saint-Louis Lariboisière, Paris, France.

Moyamoya disease (MMD) is a genetically heterogeneous occlusive cerebrovascular disease, often with pediatric onset, characterized by bilateral occlusion of the distal internal carotid artery, leading to compensatory collateral vessel formation at the base of the brain. Several genes have been identified increasing the risk for Moyamoya angiopathy but the majority of cases do not have an identified genetic trigger. Therefore, to identify novel genes for MMD, we performed whole exome sequencing in 41 trios and concentrated on *de novo* variants. Interestingly, we discovered 4 *de novo* rare pathogenic variants in three genes, *CHD4* (p.Arg1127Gln), *CNOT3* (p.Gln215* and p.Leu522Ile) and *SETD5* (p.Glu642Asnfs*15). *CHD4* *de novo* pathogenic variants are enriched in cases with congenital heart disease, intellectual disability, and distinctive dysmorphisms and including a case with MMD. *CNOT3* *de novo* pathogenic variants were reported in developmental disorder patients, and *SETD5* haploinsufficiency is established to predispose to global developmental delay with variable dysmorphic features. The phenotypes of our 4 *de novo* patients are overlapping with previously published phenotypes with 3/4 presenting with developmental delay, 2/4 intellectual disability, 2/4 facial abnormality, 1/4 bicuspid aortic valve, and 1/4 genitourinary defect. We then mined exome data from 158 additional probands with MMD, isolated or with affected family members, filtered for variants predicted to disrupt the protein sequence, and kept variants with a CADD score greater than 20. These three genes are invariant in the general population for both missense and loss of function (LoF) variants. Nevertheless, we found 9 additional rare pathogenic variants, including 5 missense and 1 splice donor site LoF variant in *CHD4*, 1 missense in two unrelated MMD probands in *CNOT3*, and 2 missenses in *SETD5*. *CHD4*, *CNOT3* and *SETD5* encode proteins involved in chromatin remodeling (component of NuRD, CCR4-NOT, and NCoR complex, respectively) and, together with a previously reported gene for MMD-like cerebrovascular disease, *YY1AP1* encoding a component of the INO80 chromatin remodeling complex, suggest altered chromatin remodeling as an underlying cause of MMD. Furthermore, this data support the hypothesis that MMD results from defective differentiation of smooth muscle cells in the cerebral arteries and expands our understanding of the phenotypic pleiotropy associated with genetic disruption of chromatin remodeling.

3078W

Genetic architecture of pulmonary hypertension as revealed by large gene panel sequencing in a clinical setting. F. Soubrier^{1,2}, M. Eyries^{1,2}, D. Montani³, S. Nadaud⁴, B. Gireud⁵, M. Levy⁶, A. Bourdin⁷, R. Tresorier⁸, A. Chaouat⁹, V. Cottin⁸, C. Sanfiorenzo⁸, G. Prevot¹⁰, M. Reynaud-Gaubert¹¹, C. Dromer¹², A. Houejeh¹³, K. Nguyen¹⁴, F. Coulet¹, D. Bonnet¹, M. Humbert¹. 1) Department of Genetics- GH Pitié-Salpêtrière, Assistance Publique-Hôpitaux de Paris, PARIS, Paris, France; 2) UMR_S 1166, Sorbonne-Université; INSERM and Institute for Cardiometabolism and Nutrition (ICAN), Paris; 3) Université Paris-Sud, Faculté de Médecine, Université Paris-Saclay, Service de Pneumologie, Centre de Référence de l'Hypertension Pulmonaire; INSERM UMR_S999, Hôpital de Bicêtre, AP-HP, Le Kremlin Bicêtre, France; 4) -M3C-cardiologie pédiatrique, hôpital Necker-enfants malades, AP-HP, Paris; 5) PhyMedExp, University of Montpellier, INSERM, CNRS, France. AND Département de pneumologie et addictologie; CHU Montpellier, Montpellier, France; 6) Service de cardiologie maladies vasculaires, CHU Gabriel Montpied, Clermont-Ferrand; 7) Département de pneumologie-CHRU Nancy-Université de Lorraine, Vandœuvre-lès-Nancy, Inserm, U1116; Université de Lorraine, Nancy, France; 8) Service de pneumologie, Centre national de référence des maladies pulmonaires rares, Hôpital Louis Pradel, université Claude Bernard Lyon 1, UMR754, Lyon; 9) Service de pneumologie, hôpital Pasteur-CHU Nice; 10) Service de pneumologie, hôpital Larrey, Toulouse; 11) Service de pneumologie, CHU nord de Marseille, Assistance Publique-Hôpitaux de Marseille (AP-HM), Marseille; 12) Service de pneumologie, CHU de Bordeaux hôpital Haut-Levêque, Pessac; 13) Service de cardiologie infantile et congénitale, CHRU Lille-hôpital cardiologique, Lille; 14) Département de Génétique Médicale, CHU la Timone enfants, AP-HM, Marseille.

Pulmonary arterial hypertension (PAH) and pulmonary veno-occlusive disease (PVOD) are closely related forms of pulmonary hypertension but in contrast to hereditary PVOD for which a single gene has been identified and whose mutations provoke a recessively transmitted disease with high penetrance, several genes have been implicated in PAH since the discovery of BMPR2 mutations in 2000. Here, we sought to determine the genetic architecture of PH according to gene mutations found in a cohort of French patients. Methods: Genetic analysis was prospectively performed on 268 PAH and PVOD patients (adult and pediatric cases). We developed a targeted capture strategy using SeqCap EZ choice library (Roche Nimblegen) coupled with multiplexing and high-throughput sequencing using the Miseq system (Illumina). We designed a gene panel including 9 genes analyzed for molecular diagnosis of PH (*BMPR2*, *TBX4*, *EIF2AK4*, *CAV1*, *KCNK3*, *SMAD9*, *ACVRL1*, *ENG*, *BMP9*). The *BMP10* gene, a *BMP9* paralog, was also included in the capture design. Standard bioinformatics pipelines were used for sequence analyses. Identified variants were classified according to ACMG. Following pedigree analysis and clinical examination, 181 patients were classified as sporadic idiopathic PAH (iPAH), 11 as familial PAH (fPAH), 13 as drug or toxin induced PAH and 63 as PVOD. A pathogenic mutation was identified in 49 of the 268 analyzed patients. Identification of bi-allelic *EIF2AK4* mutations in one initially suspected PAH case led us to reclassify this patient as PVOD. Thus 35 mutations were identified in iPAH patients (19.4% of iPAH), 6 in fPAH patients (54.5% of fPAH) and 8 in PVOD patients (12.5% of PVOD). 31 *BMPR2* mutations (25 iPAH, 6 fPAH), 5 *TBX4* mutations (5 iPAH), 2 *ACVRL1* (2 iPAH) 2 *BMP9* (2 iPAH), 1 *SMAD9* (1 iPAH) and 8 biallelic *EIF2AK4* mutations (8 PVOD) were found. A truncating mutation (C.370C>T, p.Arg124*) was found in the *BMP10* gene in a young female patient who was diagnosed with PAH at 11 yrs with severe evolution. No pathogenic mutation was identified in *KCNK3* and *CAV1* but 2 variants of unknown significance were identified in these genes. Conclusions: Gene panel sequencing is an efficient tool to improve the knowledge of PH genetic architecture. Our results confirm the involvement of genes previously described in hPAH, even rarely involved, reveal the emerging role of *TBX4* in adult PAH while this gene was initially described as a major gene in pediatric PAH and suggest *BMP10* as a new candidate.

3079T

High clinical impact *KCNQ1* mutation present in 1/40 Amish: Culturally-appropriate return of results from population whole exome sequencing (WES). E.A. Streeten¹, V. See², L.J. Jeng³, K. Maloney¹, T.I. Pollin¹, M. Daue¹, B. Mitchell¹, A.R. Shuldiner^{1,2}. 1) Endocrinology, University of Maryland School of Medicine, Baltimore, MD; 2) Regeneron Pharmaceuticals; 3) Cardiology, University of Maryland School of Medicine.

Background: Mutations in *KCNQ1* cause Long QT Syndrome, Type 1 (LQTS1), with an estimated frequency of 1/2000. From research WES of 4950 Old Order Amish, we identified an exonic missense variant in *KCNQ1* (c.671C>T, p.T224M), which we classified as likely pathogenic per ACMG/AMP guidelines. One in 40 Amish carry this variant, with carriers demonstrating longer EKG-derived QTc intervals compared to non-carriers (b=20 msec; p=10⁻²³). Since beta blocker (BB) therapy reduces the risk of syncope and sudden cardiac death (SCD) in LQTS1 by 70-90%, we performed follow up clinical phenotyping of the variant to inform treatment recommendations and enable genetic counseling. **Methods:** Following IRB approval, 124 Amish were notified that they carry a genetic change that put them at risk for heart problems. To date, 87 individuals have accepted an invitation to enroll in a study that would clarify their risk. At a first home visit, we obtained a comprehensive medical history including personal history of syncope, 3-generation family history (FH), repeat EKG, and blood for CLIA confirmation. After variant confirmation, we returned for a second home visit to provide treatment recommendations (Heart Rhythm Society guidelines) and genetic counseling. Twenty-six non-variant carrier Amish individuals were enrolled as controls. **Results:** All p.T224M variants were CLIA-confirmed. Mean QTc was 480±29 ms in cases, 423±20 ms in controls, p=0.000011 (normal ranges: M<450 ms, F<460 ms). Syncope was noted in 28 (33%) cases vs 4 (17%) controls (p=0.17), with non-vasovagal syncope in 7 (17%) cases vs zero controls (p=0.039). A FH of SCD in 1st or 2nd degree relatives was noted in 34 (40%) cases vs zero controls (p = 0.00006). A FH of SCD < 30 yo was noted in 9 (10%) cases, including two crib deaths, a 6 yo walking and a 12 yo swimming, vs zero controls (p=0.008). Of 36 visit #2 completed, 33 (94%) cases met guidelines for BB therapy. Of these, 19 (58%) agreed to start BB, which was coordinated with local primary care providers. Cascade testing was recommended for all 1st degree relatives. **Conclusions:** A *KCNQ1* likely pathogenic variant was present in 1/40 Amish and associated with a high risk of syncope and FH of early sudden death. Increased awareness, cascade and population screening, and treatment of affected individuals will likely prevent significant arrhythmogenic morbidity and mortality.

3080F

Targeting ER stress for collagen IV disorders reveals tissue specific outcomes and reduces intracerebral haemorrhaging. T. Van Agtmael¹, F. Jones¹, L. Murray¹, S. McNeilly¹, A. Dean¹, A. Aman², K. Kadler¹. 1) Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, G12 8QQ United Kingdom; 2) Wellcome Centre for Cell-Matrix Research, University of Manchester, Manchester, M13 9PT, UK.

Mutations in the collagen genes *COL4A1* and *COL4A2* cause Mendelian eye, kidney and cerebrovascular disease including intracerebral haemorrhage (ICH), and common collagen IV variants are a risk factor for sporadic ICH. Development of novel treatment for these diseases is urgently needed and requires understanding of pathomolecular disease mechanisms. While mutations cause endoplasmic reticulum (ER) stress and basement membrane (BM) defects, their relative contribution to disease and the potential of ER-stress as a therapeutic target remains unclear. To address this mice were treated with the chemical chaperones, PBA and TUDCA, as a preventative therapy from conception for 5 months or as a 1 month treatment from 4 months of age, when the disease is developed. This revealed that preventative oral treatment of *Col4a1* mutant mice with PBA and TUDCA reduced sporadic adult ICH and tubular kidney defects. In contrast, glomerular kidney disease and eye disease remained. Interestingly, ICH was also reduced following treatment of adult mice with established disease. PBA acted by reducing ER-stress which was accompanied by increased collagen IV incorporation into basement membranes (BM). However, chemical chaperone treatment did not improve BM structure with a reduced ability of basement membrane to withstand mechanical stress. Comparative analysis between ER stress, matrix defects and pathology suggested that Mendelian collagen disorders have tissue specific disease mechanisms. These data provide mechanistic insight into *COL4A1* disease, and identify that alleviating ER-stress is a valid therapeutic target for treating sporadic ICH which may be feasible using FDA-approved compounds. Future treatments for collagen IV patients will require stratification based on the clinical presentation in patients and mechanism of mutations.

3081W

A *PLN* nonsense variant causes dilated cardiomyopathy with heart failure in a novel autosomal recessive inheritance mode. D. W. Wang¹, Z. Li², P. Chen², J. Xu², H. Wang², D-W. Wang². 1) State Key Laboratory of Reproductive Medicine and The Center for Clinical Reproductive Medicine, Jiangsu Province Hospital, Nanjing, Jiangsu Province, China; 2) Division of Cardiology, Departments of Internal Medicine and Genetic Diagnosis Center, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China.

Background—Pathogenic variants in human phospholamban coding gene (*PLN*) are responsible for autosomal dominant hereditary dilated cardiomyopathy (DCM) with heart failure. **Methods and Results**—We performed high-depth targeted next-generation sequencing using a cardiomyopathy-panel containing 80 disease-related genes in 650 patients with cardiomyopathy to identify potential pathogenic *PLN* variants. A novel homozygous nonsense variant (p.Glu2Ter, c.4G>T) in *PLN* gene was identified in a 36-year-old male suffered from DCM with heart failure out of 650 unrelated cardiomyopathy patients. Whole-exome sequencing and Sanger sequencing were used to comprehensively evaluate the genetic cause of the proband and his pedigree. No more DCM-causing variant or likely pathogenic copy number variation was identified. The *PLN*-p.Glu2Ter variant is in an evolutionary highly conservative area across multiple species (phyloP score=2.7). The *PLN*-p.Glu2Ter variant was not detected in 800 unrelated healthy controls. Western blots showed that the variant significantly reduced the phospholamban expression. In pedigree analysis, we found that all heterozygous *PLN*-p.Glu2Ter carriers had normal heart size and cardiac function, which revealed an autosomal recessive inheritance mode. **Conclusions**—Our study identifies a novel pathogenic variant of *PLN*, and enriches a novel pathogenic inheritance mode of *PLN* causing DCM and heart failure.

3082T

A new catecholaminergic polymorphic ventricular tachycardia mechanism associated with numerous sudden deaths. M. Blancard¹, Z. Touat-Hamici¹, L. Yin², G. Vaksman³, N. Roux-Buisson⁴, V. Fressart⁵, D. Klug⁶, N. Neyroud⁷, I. Denjoy⁸, J. Ramos-Franco⁸, A.M. Gomez², P. Guicheney¹. 1) INSERM UMRS1166, Sorbonne University, Paris, France; 2) INSERM UMRS1180, Paris-Sud University, Châtenay-Malabry, France; 3) Hôpital Privé de la Louvière, Lille, France; 4) Université Joseph Fourier, Grenoble, France; 5) Unité de Cardiogénétique et Myogénétique, AP-HP, Groupe Pitié-Salpêtrière, Paris, France; 6) CHRU Lille - Hôpital Cardiologique, Lille Cedex, France; 7) AP-HP, Hôpital Bichat, Département de Cardiologie, Paris, France; 8) Rush University Medical Center, Chicago, IL 60612, USA.

Background: Catecholaminergic polymorphic ventricular tachycardia (CPVT) is a rare genetic arrhythmia triggered by exercise or acute emotion in patients with normal resting electrocardiogram. The major disease-causing gene for the dominant form, *RYR2*, encodes the cardiac ryanodine receptor (RyR2), while rare variants in RyR2-associated proteins (calsequestrin 2, triadin), encoded by *CASQ2* and *TRDN*, are identified in recessive forms of CPVT. We reported two French families with numerous sudden deaths and a novel *RYR2* variant. **Methods:** Family members were evaluated by exercise stress test, and CPVT genes were sequenced. Spontaneous Ca²⁺ release was evaluated by fluorescent microscopy in HEK293 cells expressing RyR2 wild type or the novel variant with Fura-2 AM as loading dye. Cells were challenged with either 8-Bromo-cAMP or forskolin to mimic adrenergic stress. **Results:** In a four-generation family, eight sudden cardiac deaths occurred before the age of 30 years. The proband had a first syncope at the age of 10, then experienced an aborted sudden cardiac death at the age of 14 in a context of emotion and moderate hypokalemia. The grand-parents were consanguineous. We identified a novel *RYR2* missense variant inducing the loss of a negative charge located out of the four classical domains where CPVT mutations are commonly found. This well conserved residue is located in a RyR2 domain of unknown function. Genotyping clearly showed that the affected subjects were all carriers of at least one copy of the *RYR2* variant. Three sisters were homozygous while eleven other patients were heterozygous carriers. We also identified the same variant at the heterozygous state in another family where two sudden deaths occurred. The analysis of spontaneous calcium release in presence of various calcium concentrations revealed no difference between the WT and the mutant RyR2 at basal conditions which excluded the most classical mechanism of a leaky mutant channel. WT-RyR2 transfected cells pretreated with stable c-AMP or forskolin showed the expected increase in SOICR activity while mutant cells did not present any enhanced activity. **Conclusions:** We identified a novel loss-of-function RyR2 variant, unresponsive to adrenergic stimulation, responsible for the high number of sudden deaths occurring in two families.

3083F

Six arterial tortuosity case with novel and known homozygous and compound heterozygote missense variations in the SLC2A10 gene: Genotype/phenotype correlations. A. Kablan¹, E. Mihci², A. Kocabas³, B. Nur², F. Kardelen³, M. Cousin^{4,5}, F. Uysal⁶, O. Bostan⁶, E. Cil⁶, E. Klee^{4,5}, S.G. Temel¹. 1) Uludag University, Faculty of Medicine, Department of Medical Genetics, Bursa, Turkey; 2) Akdeniz University School Of Medicine Department of Pediatrics Division of Pediatric Genetics Antalya, Turkey; 3) Health Science University, Antalya Training and Research Hospital, Department of Pediatric Cardiology, Turkey; 4) Center for Individualized Medicine, Mayo Clinic, Rochester, MN, USA; 5) Department of Clinical Genomics, Mayo Clinic, Rochester, MN, USA; 6) Uludag University, Faculty of Medicine, Department of Pediatric Cardiology, Bursa, Turkey.

Arterial tortuosity syndrome (ATS) is a rare autosomal recessive connective tissue disorder, mainly characterized by tortuosity and elongation of the large and medium-sized arteries with predisposition to stenosis and aneurysm. ATS is caused by mutations in the *SLC2A10* gene, encoding for the facilitative glucose transporter 10 (GLUT10) and is described typically in pediatric patients. We report 3 unrelated boy, 2 unrelated men and one girl with ATS. The main clinical findings of all our cases included elongated face, saggy cheeks, micrognathia, malar hypoplasia, joint hypermobility and hyperextensible skin. Sequencing of the *SLC2A10* gene in the probands revealed the presence of two novel pathogenic homozygous missense variants (c.727C>A/c.395C>A), one compound heterozygote missense variants (c.314G>A, c.727C>A) and one known homozygous missense variant (c.1465G>C) (2 case from sama family, one unrelated case). Heterozygous *SLC2A10* mutations were confirmed in both mothers and fathers. The missense variants lead to a p.Gln243Lys, p.Arg132Gln and p.Gly489Arg change in the seventh transmembrane domain, endofacial loop and in the twelfth transmembrane domain of GLUT10, respectively. Genotype/phenotype correlations with detailed clinical, radiological data and genetic findings of our cases will be discussed. Because ATS resembles some Connective Tissue Disorders (CTDs) like Loeys-Dietz and Marfan syndrome, timely differential diagnosis is extremely important for early diagnosis, management of the disease and for intervention of aneurysms to prevent serious vascular complications. Molecular diagnosis is mandatory to distinguish ATS from other CTDs with strictly overlapping clinical signs and with a high rate of more severe cardiovascular events and to define targeted clinical follow-up and timely cardiovascular surgery when needed. .

3084W

A novel de novo heterozygous variant in ACTC1 in a patient with a Noonan syndrome phenotype. V.E.H. Kim¹, M. Buscarilli¹, G.L. Yamamoto^{1,2}, R.S. Honjo³, J.R.M. Ceroni¹, A.C. Malaquias², C.A. Kim¹, A.A.L. Jorge², D.R. Bertola^{1,2}. 1) Departamento de Genética, Instituto da Criança dos Hospital das Clínicas da FMUSP, São Paulo, SP, Brazil; 2) Centro de Pesquisas sobre o Genoma Humano e Células Tronco – Instituto de Biociências – USP; 3) Unidade de Endocrinologia Genética LIM/25 - Disciplina de Endocrinologia do Hospital das Clínicas da FMUSP – HC – FMUSP.

The alpha-cardiac actin gene (*ACTC1*) has an essential role on the interaction domain with myosin heavy chain, mediating the heart function. *ACTC1* disturbance is a known cause of cardiomyopathy and arrhythmia, as well as structural cardiac defects, such as septal defects and valvar anomalies, including Ebstein anomaly. Furthermore, in a large Lebanese Maronite family, besides the cardiac anomaly, some of the affected individuals harboring a variant in *ACTC1* presented midline defects, suggesting a role not only in heart embryology. Noonan syndrome is a genetic heterogenous RASopathy with more than ten genes responsible for this disorder, characterized mainly by short stature, facial dysmorphisms, including hypertelorism, sternal deformity, cardiac defect and cryptorchidism. The diagnostic yield of the molecular analysis achieves approximately 80%. We report on an isolated case, a 13 year-old male patient presenting typical features of Noonan syndrome, with Ebstein anomaly, atrial septal defects and Wolf-Parkinson-White syndrome. Whole-exome sequencing of the patient and his parents was performed and a novel, de novo heterozygous missense mutation in *ACTC1* was identified. No pathogenic variants either in known genes associated with a RASopathy or in other genes associated with a monogenic disorder have been found. Since the employment of next-generation sequencing as a tool for the discovery of new genes responsible for monogenic disorders, six novel genes have been associated with Noonan syndrome in a small percentage of the cases. Still, the whole spectrum of molecular basis of this disorder is not completely known. We could not rule out for certainty that the *ACTC1* variant in the present patient is responsible only for the cardiac anomaly, but this is the second reported case harboring *ACTC1* variant and showing other signs besides heart defect. In the patients previously reported, although the authors did not consider the diagnosis of Noonan syndrome, several of them presented ocular hypertelorism and pectus deformity, cardinal features of Noonan syndrome. It is possible to speculate that this specific actin could play a role in a more widespread embryonal development leading to clinical features within a Noonan syndrome phenotype. FAPESP 2013/08028-1; CNPq 304130/2016-8.

3085T

Whole exome sequencing identifies a novel candidate gene in an Ashkenazi Jewish family with Tetralogy of Fallot. J. Liu, A. Sara, J.Y. Liu, P. Gupta, J. Fan, J. Wang. Medicine - Cardiology, University of California, Los Angeles, Los Angeles, CA.

Tetralogy of Fallot (ToF) is a congenital cardiac malformation consisting of four separate defects of the heart. Up to a quarter of ToF patients consist of individuals with chromosomal anomalies, such as trisomy 21 (Down Syndrome) or chromosome 22q11.2 deletion syndrome (DiGeorge Syndrome). However, due to intricate inheritance patterns and suspected multifactorial disease causation, around 70% of ToF cases are idiopathic. Yet, with a familial recurrence rate of 3%, an underlying genetic cause is strongly suspected. Here, we evaluate the genetic origin of ToF in a family of two affected sisters with unaffected parents. Using whole exome sequencing (WES) and computational analysis, we uncovered compound heterozygous variants in our affected patients within the DSPP gene. DSPP, or dentin sialophosphoprotein, is mainly established to serve a vital role in odontogenesis. While previous western blotting and RT-PCR analysis in mice demonstrate absence of DSPP in non-mineralized cardiac tissue, the gene's possible role in cardiogenesis has yet to be substantially characterized. Continued progress of this project aims to address the genetic contribution of DSPP as a developmental factor for ToF, which would facilitate the creation of targeted treatments in the future. Furthermore, the project contributes to the advancement of therapies that take individual variability in genes into account, and thus contributes data to a growing field of precision medicine.

3086F

Whole exome sequencing improves diagnostic precision in individuals affected with thoracic aortic aneurysm and dissection. B. Wolford¹, W. Hornsby², D. Guo³, W. Zhou¹, M. Lin², L. Farhat⁴, J. McNamara², A. Driscoll⁵, X. Wu⁶, E. Schmidt¹, C. Brummett¹, H. Patel¹, K. Eagle², D. Milewicz², J. Kitzman^{1,6}, C. Willer^{1,2,6}, B. Yang¹. 1) Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan; 2) Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan; 3) Department of Internal Medicine, University of Texas Health Science Center at Houston (UTHealth), Houston, Texas; 4) Department of Cardiac Surgery, University of Michigan, Ann Arbor, Michigan; 5) Department of Anesthesiology, University of Michigan, Ann Arbor, Michigan; 6) Department of Human Genetics, University of Michigan, Ann Arbor, Michigan.

Thoracic aortic aneurysm and dissection (TAAD) is a genetically heterogeneous and highly heritable disease involving dilation of the aorta (aneurysm) and possible tearing of aortic wall (dissection) or rupture. A genetic diagnosis is critical for precise clinical care since familial or syndromic TAAD is an indication for earlier surgical intervention due to increased risk of dissection or rupture of an aneurysm. This study demonstrates the use of research-level whole exome sequencing (WES) in the diagnosis and management of TAAD. WES was performed in 265 thoracic aortic dissection (Stanford type A and B) or rupture cases from the University of Michigan Cardiovascular Health Improvement Project (CHIP) Biorepository, and 265 age-, sex-, and ancestry-matched unaffected controls from the Michigan Genomics Initiative (MGI) biobank. After quality control, 249 non-trauma dissection cases and 258 controls were retained. Annotation of variants in 11 TAAD genes, identified *a priori*, was performed blinded to a sample's case or control status. Rare variants were annotated as pathogenic based on ACMG guidelines and established criteria for specific aortopathy genes or as variants of unknown significance (VUS) if lacking proof of pathogenicity. We replicated the WES-pathogenic variants using Molecular Inversion Probe Sequencing (MIPS). We identified 24 pathogenic variants across 6 genes (*COL3A1*, *FBN1*, *LOX*, *PRKG1*, *SMAD3*, *TGFBR2*) in 26 individuals representing 10.4% of cases and 0% of controls. For 17 individuals these results align with previous clinical diagnoses. However, for 9 individuals, research-level WES and annotation of pathogenic variants adds diagnostic precision by providing a diagnosis (n=8) or correcting a diagnosis (n=1). We also identified 4 individuals with diagnoses based on previous CLIA genetic testing which need revision after research-level WES and annotation as VUS. Although previous work demonstrates an increased burden of VUS in dissection cases < 56 years of age, VUS burden in aortopathy genes was not increased in our cohort ($\chi^2 = 2.94$, p-value=0.086). This demonstrates the need for greater understanding and careful interpretation of VUS in clinical practice. This work suggests that research-level WES increases diagnostic precision and can be used to support clinical decision making for TAAD patients and cascade screening of family members. We plan to return these results to patients in collaboration with aortic surgeons and genetic counselors.

3087W

Cardiomyopathy due to *PRDM16* mutation: First description of a fetal presentation, with possible modifier genes. G. Delplanq¹, A. Vitobello², A. Sorlin^{1,2}, C. Phillipe^{1,2}, Y. Duffourd³, F. Tran-Mau-Them^{1,2}, S. Moutton^{1,2}, C. Denis³, AL. Brue⁴, JC. Eichers⁵, S. Nambot^{1,2}, M. Chevarin², C. Poé², B. Khalilouk⁴, T. Rousseau⁴, S. Falcon-Eiche^{3,5}, FT. Harizay⁶, C. Chauvin-Robinet^{1,2}, L. Faivre^{1,2}. 1) Reference Center for Developmental Anomalies, Department of Medical Genetics, Dijon University Hospital, Dijon, France; 2) INSERM U1231, LNC UMR1231 GAD, Burgundy University, Dijon, France; 3) Pediatrics Department of Cardiology, Dijon University Hospital, Dijon, France; 4) Gynecology and Obstetrics Department, Dijon University Hospital, Dijon, France; 5) Department of Cardiology, Dijon University Hospital, Dijon, France; 6) Pathology Department, Dijon University Hospital, Dijon, France.

PRDM16 (positive regulatory domain 16, NM_022114.3, 1p36.32, #605557) located in the critical region for cardiomyopathy defined by Gajekka et al. (2010) encodes a zinc finger transcription factor involved in leukemogenesis, palatogenesis, neurogenesis, and brown adipose tissue differentiation. In 2013, Arndt et al resequenced *PRDM16* in 75 non syndromic individuals with left ventricular noncompaction (LVNC, #615373) and in a series of 131 cardiac biopsies from individuals with dilated cardiomyopathy (CMD, #615373). Eight of these samples were found to carry putatively deleterious sequence changes in *PRDM16*. We present the first case of *PRDM16* mutation in a fetus of a 37-year-old female, gravida 1, para 0. The third trimester obstetric ultrasound revealed a hydropic fetus with hydramnios and expanded hypokinetic heart. Familial history was negative. Exome sequencing identified a de novo c.1057C>T (p.Gln353*) heterozygous variant in *PRDM16* never reported in gnomAD. The CADD and GERP scores were respectively 38 and 4.56. In ExAC, the gene appears to be intolerant to the loss of function (pLI=1). It also identified three variants of unknown significance (VOUS), according to the ACMG (American College of Medical Genetics and Genomics) guidelines, in genes described in association with cardiomyopathy; two in the titin gene (TTN, NM_001256850.1): a de novo missense variant c.44319A>C (p.Lys14773Asn) and a splice variant c.33043+5A>G inherited from the mother; one in the Lim Domain-Binding 3 gene (LDB3, NM_001080116.1), c.349G>A (p.Asp117Asn) inherited from the mother. The familial segregation tests of all the variants were verified by Sanger sequencing. The parents elected to terminate the pregnancy at 31 weeks of gestation and 3 days, and a 2610g, 48cm, eutrophic fetus was delivered with facial dysmorphism due to hydrops. The phenotype showed a isolated cardiomegaly with a fibroelastotic thickening of the endocardium of the ventricles predominant on the left and a thickened fibrotic interventricular septum. Noted that echocardiographies of the parents were both normal. Besides to the *PRDM16* de novo probably pathogenic variant, the TTN and LDB3 VOUS could possibly contribute to the severity of the cardiac phenotype of the fetus. Because of the genetic heterogeneity of cardiomyopathies, large panels or even WES should be considered as the main approaches for the molecular diagnosis in particular in fetal presentations, where multiple hits seem to be common.

3088T

Genome-editing patient-specific induced pluripotent stem cells to investigate the functional role of *RARG* gene in doxorubicin-induced cardiotoxicity. H. Huang¹, E. Christidi¹, L. Brunham^{1,2,3}. 1) Centre for Heart Lung Innovation, Department of Medicine, University of British Columbia, Vancouver, BC, Canada; 2) Translational Laboratory in Genetic Medicine, Agency for Science, Technology and Research, Singapore; 3) Department of Medicine, National University of Singapore, Singapore.

Background: Doxorubicin is a commonly used chemotherapy drug that used to treat many adult and childhood cancers, but its clinical usefulness is limited by doxorubicin-induced cardiotoxicity (DIC). The risk of DIC increases sharply at doses above 550 mg/m². However, some patients suffer from DIC even at low doses, suggesting the presence of genetic susceptibility to this adverse drug reaction. **Approach:** The molecular mechanisms of DIC are still not fully understood. A genetic variant (rs2229774, p. Ser427Leu) located in Retinoic Acid Receptor Gamma (*RARG*) gene was identified by a genome-wide association study (GWAS) to be associated with DIC. The functional role of genetic variation in *RARG* on DIC is currently unknown. The goal of this study was to investigate the role of *RARG* in DIC. To do so, we performed genome editing in patient-specific induced pluripotent stem cells (iPSCs) to introduce the rs2229774 variant or inactivate the *RARG* gene, and characterized the cellular phenotypes of isogenic pairs of iPSC-derived cardiomyocytes (CMs). **Results:** Using CRISPR/Cas9 techniques, we successfully inactivated the *RARG* gene in iPSCs, which was confirmed by qPCR gene expression assay, and introduced the rs2229774 variant in both the heterozygous and homozygous state. With cell viability assays, we show that whereas presence of rs2229774 does not alter the response to doxorubicin, inactivation of *RARG* renders the cardiomyocytes less susceptibility to DIC. **Conclusion:** Our results provide evidence for a functional role of *RARG* in DIC, while at the same time suggesting that this effect may not be mediated by rs2229774. More broadly, our results demonstrate the utility of genome-editing in iPSCs to investigate the functional significance of GWAS variants and genes associated with drug response.

3089F

Genotype-phenotype: Phenotype differences when studying individual mutations in the cysteines of cbEGF-like domains in fibrillin-1 gene (*FBN1*).

J.A. Aragon-Martin, Y.B.A. Wan, M. Garcia, A.H. Child. Genetics Department, Molecular and Clinical Sciences Research Institute, London, United Kingdom.

Background: Marfan syndrome (MFS) is a monogenic autosomal dominant connective tissue disorder that is phenotypically heterogeneous [cardiovascular (Thoracic Aortic Aneurysm – TAA), ocular (Ectopia Lentis – EL), skeletal (Scoliosis)] and reported to be due to mutations in *FBN1*. There are 43 cbEGF-like domains that contain 6 highly conserved cysteine residues that form disulphide bonds between themselves: C¹-C³, C²-C⁴ and C⁵-C⁶. Previous studies *in vitro* (proteolysis of recombinant fibrillin-1 fragments) showed that individual mutations in cbEGF domains can affect these domains differently and may suggest an explanation for some genotype-phenotype relationship in MFS. **Methods:** Patients (n=75 – 33M:42F) with mutations in one of the 6 cysteine residues in any of the 43 cbEGF-like domains were entered in the study. Clinical data was obtained from the Marfan clinic at St George's Hospital and the genotypes from the *FBN1* mutation-screening database in the Sonalee Laboratory. A genotype-phenotype database was created with this data. A new column was added to record the name of the substituted amino acid. The results from the phenotypes thoracic aortic dissection and thoracic aortic aneurysm were combined and named TAAD since both of them can lead to death. Only TAAD and EL were studied since skeletal and other features can be seen in the general population. **Results:** When looking at cysteine residues in C¹, C², C³, C⁴, C⁵ or C⁶ individually: females were less likely to have TAAD phenotype but more likely to have EL phenotype when the amino acid substituted in C¹ was tyrosine (Y – p=0.0476) or in C³ was arginine (R – p=0.0286). When looking at all cysteine residues together: females were less likely than men to express TAAD phenotype (p=0.0145); females were less likely to have TAAD phenotype when the amino acid substituted was either Y (p=0.0052) or R (p=0.0112); both sexes were more protected from EL phenotype but had a high incidence of TAAD when the amino acid substituted was tryptophan (W – p=0.0476). **Conclusions:** Predisposition to TAAD or EL phenotype is partially determined by the amino acid replacing cysteine. It is imperative to look at a larger cohort (UMD) to clarify further these findings. Sex difference may be due to hormonal influence or protective gene on the X chromosome. We suggest looking at copy number variations and methylation to better understand variability in this cohort and also in family members with the same mutation.

3090W

Heterozygote Familial Hypercholesterolemia (HeFH) genotype-phenotype correlations among two AstraZeneca sponsored studies.

K. Carss¹, J. Hostyk², S. Petrovski¹, R. March¹, D.B. Goldstein², C. Haefliger¹. 1) AstraZeneca Centre for Genomics Research, Precision Medicine and Genomics, IMED Biotech Unit, AstraZeneca, Cambridge, UK; 2) Institute for Genomic Medicine, Columbia University Medical Center, New York, USA.

Heterozygote familial hypercholesterolemia (HeFH) is a disease of lipid metabolism that affects ~1/500 individuals and leads to early coronary heart disease. Here, we present whole-exome sequencing (WES) results from HeFH patients from two AstraZeneca clinical trials designed to assess the efficacy of rosuvastatin on this population (n=369 unrelated cases). The aims of this study were threefold. First, to identify pathogenic variants in known genes, second to identify putative novel candidate HeFH genes using a collapsing analysis approach, and third to leverage the rich phenotypic data available to identify genotype-phenotype correlations. Overall, pathogenic variants were identified in known HeFH-associated genes in 263 (71.3%) cases: 251 (68%) in *LDLR*, 10 (2.7%) in *APOB*, and 2 (0.5%) in *PCSK9*. We applied rare-variant gene-based collapsing analysis comparing 106 (28.7%) unsolved cases to a collection of ancestry-matched 5,930 sequenced population controls. While no novel gene achieved genome-wide significance ($p < 2.7 \times 10^{-6}$), this analysis has provided a list of candidate genes. We also identified several interesting genotype-phenotype correlations. Although the age of diagnosis is unknown, we found that patients with a pathogenic protein-truncating *LDLR* variant were recruited into the studies at a younger age than patients with a pathogenic missense *LDLR* variant (24.5 years vs 43 years median; $p = 0.032$). After correction for clinical covariates, patients with a *APOB* molecular diagnosis also have higher baseline high-density lipoprotein cholesterol (HDL-C) levels than patients without ($p = 5.2 \times 10^{-3}$). Specific recurring *LDLR* variants were found to affect baseline low-density lipoprotein cholesterol (LDL-C) levels, including ENSP00000454071.1:p.Pro309LysfsTer59 ($p = 1.7 \times 10^{-7}$). In conclusion, exome sequencing followed by a diagnostic variant screen, collapsing analysis, and genotype-phenotype correlation analysis yields deeper insight into the genetic architecture and clinical characteristics of the HeFH patient population.

3091T

Electronic health record phenotypes of patients with rare loss-of-function variants within genes associated with heart disease. C. Haggerty, L. Jing, D. Hartzel, D. Carey, B. Fornwalt on behalf of the DiscovEHR collaboration. Geisinger, Danville, PA.

The prevalence of rare, putative loss of function (LOF) variants in the human genome is much higher than expected. Thus, while specific variants are rare, their collective effect on patient health may be significant. This study tested this hypothesis for genes associated with heart disease using whole exomes from the DiscovEHR cohort of 92,445. We reviewed general cardiomyopathy and arrhythmia gene panels from 4 commercial labs and computed their union and intersection (i.e., genes appearing on all cardiomyopathy or all arrhythmia panels). We identified rare (MAF<0.001) putative LOF (frameshift, splice site, stop gain) and other protein altering variants (missense, stop loss), and cross-referenced with ClinVar. We selected: 1) cases: having a LOF or “two-star” (and above) pathogenic/likely pathogenic missense variant in any gene on the panel intersection list; and 2) controls: lacking any LOF or nonsynonymous variant (not annotated as benign/likely benign) on the panel union list. We performed genome-wide association (PheWAS), compared echocardiographic and electrocardiographic data linearly adjusted for age and sex, and modeled Cox proportional hazards for all-cause mortality adjusted for sex, body mass index and diabetes. 16,917 individuals (18.3%) carried none of over 40,000 variants in 110 genes associated with cardiomyopathy or arrhythmia (controls). 3,979 individuals (4.3%) carried at least one rare variant in the 52 consensus genes (cases). By age and sex-adjusted PheWAS, the cases had significantly increased odds for numerous cardiac diagnoses, including conduction disorders, cardiomyopathies, and heart failure (Table). Cases also had slightly lower left ventricular ejection fraction (1.1% absolute, $p=1 \times 10^{-6}$), and longer PR (3.6ms, $p=1 \times 10^{-6}$) and QTc (4.0ms, $p=4 \times 10^{-6}$) intervals. Finally, cases had 15% higher mortality risk ($p=0.02$). In conclusion, regardless of specific gene-disease associations, the presence of a rare LOF or known pathogenic missense variant in cardiomyopathy or arrhythmia genes is associated with increased risk for heart disease and early mortality, in contrast to previous data using such genome-first approaches. Future studies will incorporate such high level genomic data into polygenic risk models to potentially improve outcome predictions.

PheWAS findings		
	Odds Ratio	p
Other cardiac conduction disorders	8.4	6.00E-06
Other hypertrophic cardiomyopathy	3.7	1.00E-06
Primary/intrinsic cardiomyopathies	2.7	3.00E-26
Systolic heart failure	1.7	3.00E-10

3092F

Characterization of the genetic architecture and identification of novel genes and variants in an infantile cardiomyopathy cohort. C. Gonzaga-Jauregui¹, T. Lee², J. Staples¹, Y. Shen^{3,4}, J. Wang^{3,4}, F. Dewey¹, J.G. Reid¹, J.D. Overton¹, W.K. Chung^{2,5}. 1) 777 Old Saw Mill River Road, Tarrytown, NY, 10591 USA; 2) Department of Pediatrics, Columbia University Medical Center, New York, New York, USA; 3) Department of Systems Biology, Columbia University Medical Center, New York, New York, USA; 4) Department of Biomedical Informatics, Columbia University Medical Center, New York, New York, USA; 5) Department of Medicine, Columbia University Medical Center, New York, New York, USA.

Cardiomyopathies are a genetically and clinically heterogeneous group of disorders that affect the cardiac muscle. These disorders can cause the heart to become enlarged, increase its thickness, or reduce its contractility leading to reduced heart function and heart failure. Cardiomyopathies can be broadly divided into hypertrophic, dilated, restrictive, arrhythmogenic right ventricular, left ventricular noncompaction, and mixed. To date, several genes have been identified as contributing to different types of cardiomyopathy. Mutations in some cardiomyopathy associated genes can confer a spectrum of heart disease from hypertrophic to dilated cardiomyopathy, while others are associated with syndromic presentations in which cardiomyopathy is one component. We performed whole-exome sequencing in a cohort of 50 probands with cardiomyopathy of infantile onset, their parents, and other family members when available. We performed rare variant family based analyses, either under a trio-based approach if parents were unaffected or pedigree-guided in cases positive for a family history of cardiomyopathy. We identified known and novel rare variants in known cardiomyopathy associated genes in 17 probands, including variants in myosin and myosin binding structural proteins like *MYBPC3* and *MYH7*, and additional cardiomyopathy genes like *TTN*, *TNNI3*, *TNNC1*, *DES*, and *BAG3*. Additionally, we identified probands with variants in *PTPN11* and *RAF1*, both associated with Noonan syndrome that can include hypertrophic cardiomyopathy as a feature. We identified either *de novo* or compound heterozygous inherited variants in novel cardiomyopathy genes. These novel genes can be functionally classified as affecting cardiomyocyte structure or involved in signal transduction or energy homeostasis of the cardiac muscle. For example, we identified a *de novo* missense mutation in *MRAS*, a member of the RAS oncogene family expressed in the heart, in a child with prenatal hypertrophic cardiomyopathy. In another proband with a diagnosis of dilated cardiomyopathy, we identified compound heterozygous frameshift and missense variants in the myosin light chain kinase 3 gene, *MYLK3*, which has specific and high expression in cardiomyocytes and promotes sarcomere formation. These findings expand our understanding of the genetic architecture of genetic cardiomyopathies, the mechanisms involved in their pathogenesis, and potentially identify novel therapeutic targets.

3093W

Familial autonomic ganglionopathy and neurogenic orthostatic hypotension associated with rare *CHRNA3* variants. J.A. Phillips¹, C.A. Shibao², F. Vetrini³, P. Ward⁴, J.H. Sheehan⁵, J. Meiler⁶, J.A. Capra⁵, L. Fairbrother⁷, M.E. Koziura¹, J.D. Cogan¹, J.H. Newman², Y. Yang^{6,7}, R. Hamid¹. 1) Pediatrics, Vanderbilt University Medical Center, Nashville, TN; 2) Medicine, Vanderbilt University Medical Center, Nashville, TN; 3) Biochemistry, Vanderbilt University School of Medicine, Nashville, TN; 4) Chemistry, Vanderbilt University School of Medicine, Nashville, TN; 5) Biological Sciences, Vanderbilt University School of Medicine, Nashville, TN; 6) Baylor Genetics, Houston, TX; 7) Baylor College of Medicine, Houston, TX.

Aims: To determine the molecular basis of a new Mendelian cause of autonomic failure. **Methods:** The proband developed orthostatic hypotension (OH) and dizziness on standing at 39 yrs of age. At 40 yrs of age he was noted to have miosis, and hyperpigmentation of his back at the Vanderbilt Autonomic Dysfunction Center. Plasma catecholamines were low and remained so on standing but dopamine levels were wnl. He was diagnosed with neurogenic OH due to severe autonomic failure (AF) and miosis of unknown etiology, and telangiectasia macularis eruptiva perstans. His sister developed AF and miosis in her late teens. Their parents, brother and the proband's 3 children had no autonomic problems. At 64 yrs of age the proband was consented (IRB protocol NHGRI 15-HG0130) and enrolled by the Vanderbilt Undiagnosed Diseases Network (UDN) clinical site. **Results:** UDN WES results showed that the affected sibs are compound heterozygotes for c.907_908delCT (p.L303Dfs*115)/c.688G>A (p.D230N) variants in the acetylcholine receptor, neuronal nicotinic, alpha 3 subunit (*CHRNA3*) gene inherited from their mother and father, respectively, and their non-affected brother had a nl/p.D230N genotype. *CHRNA3* is a subunit of nicotinic acetylcholine receptors (nAChRs) that regulate blood pressure (BP) by release of catestatin, an inhibitor of catecholamine secretion and modulation of synaptic transmission in peripheral autonomic ganglia. The p.D230N variant was predicted to be damaging/probably damaging (SIFT/PolyPhen2) and by our structural modeling lies on the interface between *CHRNA3* and other nAChR subunits, and would destabilize the nAChR pentameric complex. Common variants (CV) in *CHRNA3* associate with inter-individual variations in autonomic function including BP, catestatin and epinephrine levels (PMID10318955). Knock-out (KO) mice without detectable *Chrna3* have severe AF and widely dilated pupils (PMID 10318955). Patients with AChR autoantibodies against the alpha 3 subunit develop severe AF secondary to autonomic ganglionopathy (PMID10995864). **Conclusions:** 1) Our sibs with severe AF and miosis are compound heterozygotes for *CHRNA3* rare variants (RV) that co segregate with their phenotype and have predicted effects on nAChR structure, and 2) since it has been shown that *CHRNA3* CV associate with variations in autonomic function and a murine *Chrna3* KO model has AF, we hypothesize that our affected sibs may have a new Mendelian AF disorder caused by their *CHRNA3* RV.

3094T

Rare variants in *CAPN2* underlie isolated hypoplastic left heart syndrome. J.J. White¹, E.I.T. Helle², M.M. Jenkins³, K.L. McBride⁴, S. Mital⁵, D.A. Nickerson⁶, J.X. Chong¹, E. Blue⁷, M.J. Bamshad^{1,6}, National Birth Defects Prevention Study, University of Washington Center for Mendelian Genomics. 1) Pediatrics, University of Washington, Seattle, WA; 2) Children's Hospital, University of Helsinki, Finland; 3) National Center on Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention, Atlanta, GA; 4) Center for Cardiovascular Research, Nationwide Children's Hospital, Department of Pediatrics, Ohio State University, Columbus OH; 5) Department of Pediatrics, The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada; 6) Department of Genome Sciences, University of Washington, Seattle, WA; 7) Division of Medical Genetics Department of Medicine, University of Washington Seattle WA.

Hypoplastic left heart syndrome (HLHS) is a severe congenital heart defect (CHD) characterized by hypoplasia of the left ventricle and aorta, stenosis or atresia of the aortic and mitral valves and an intact ventricular septum. HLHS represents ~4-8% of all CHD but accounts for 25% of deaths due to CHD. Seventy percent of HLHS cases are isolated (i.e., non-syndromic) but despite intense investigation, the genetic basis of HLHS remains largely unknown. We performed exome sequencing on ~330 probands with isolated HLHS (iHLHS) born to unaffected parents aggregated from four independent cohorts. Under *de novo*, recessive, and dominant models, only twelve probands (3.6%) had a rare, predicted pathogenic variant (five of which were *de novo*) in a gene previously known to underlie either iHLHS (*MYH6*, *NOTCH1*) or syndromic HLHS (*KDM6A*, *KMT2D*), and one proband had a known pathogenic variant per ClinVar in a gene (*ACTC1*) underlying other CHDs. No additional compelling candidate genes were identified under any Mendelian model. These results led us to hypothesize that iHLHS in this cohort was not due to single, large effect alleles in known or novel genes in >95% of families. Gene-based association testing of rare variants using SKAT-O in 211 iHLHS cases for which ancestry-matched controls without CHD (N=1,067) were available identified a single gene, *CAPN2* (p=3.8e-6), that reached genome-wide significance (3,048 genes with >=10 alt alleles; Bonferroni p<1.6e-5). This signal was driven by 11 (5%) cases heterozygous for rare (MAF < 0.001), predicted deleterious (9 with CADD>=15) missense variants in *CAPN2*. Homozygous loss of *Capn2* in mice results in a thin ventricular wall and is embryonic lethal, suggesting it is a promising candidate gene for iHLHS. However, all of the *CAPN2* variants in full trios (n=3) were inherited from an unaffected parent, suggesting that in those families, a variant in *CAPN2* is necessary but not sufficient to cause iHLHS. Using the presence of a *CAPN2* variant as a covariate, a conditional analysis revealed suggestive associations with *PITX2* (p=2.9e-5) and *ARHGAP31* (p=4.54e-5). *PITX2* is a key regulator of heart morphogenesis, and *ARHGAP31* underlies Adams-Oliver syndrome, a rare cause of syndromic HLHS. Two probands were heterozygous for variants in two of these three genes. Collectively, these results provide support that iHLHS is a digenic/oligogenic trait and identifies a novel pathway involved in its pathogenesis.

3095F

Aortic root dilation and mitral valve prolapse in a patient with Tatton-Brown-Rahman syndrome. H. Nguyen, AC. Tsai. Section of Genetics, Children's Hospital of Colorado, Aurora, CO.

Tatton-Brown-Rahman syndrome (TBRS; MIM #615879) is a recently described genetic overgrowth disorder characterized by tall stature, distinctive facial features and intellectual disability. Germline variants in *DNMT3A* (2p23.3) are causative for TBRS, and to date, fewer than 30 patients with over 20 variants have been reported. We report a 25 y. o. male with an initial clinical impression of Lujan-Fryns syndrome (LFS) based on overlapping clinical features with TBRS. Upon genetic testing by an expanded trio NGS panel, the patient was found to carry the *DNMT3A* variant c.1685G>A p.C562Y. While the C562Y variant has not been previously reported, a different missense change at this residue (C562R) has been described as likely pathogenic for TBRS. The physical features of our patient include: tall thin stature, autistic characteristics / repetitive behaviors, prominent glabella, deep set eyes, frontal upsweep, thick eyebrows, prominent nose, high palate, pectus carinatum, slumped posture with kyphosis, bilateral pes planus, and bilateral hindfoot deformities. He has moderate to severe intellectual disability and anxiety, agitation, and sleep disturbances that are non-responsive to psychotropic medications. Surprisingly, EKG abnormalities were detected upon medication screening, and a subsequent cardiac echocardiogram showed aortic root dilation (mild to moderate; aortic root 4.16cm Z-score 3.99) along with mitral valve prolapse. It has been suggested that phenotypic similarities exist between LFS and Marfan syndrome. Since these cardiac findings are known to be associated with Marfan-like syndrome, but not TBRS, we suggest an overlap in otherwise unestablished formal diagnostic criteria for TBRS and LFS. Interestingly, the causative gene for LFS, *MED12* has not been reported in TBRS. *MED12* is a coactivator involved in the regulated transcription of nearly all RNA polymerase II-dependent genes, and recent data indicate that DNMT DNA methyltransferases regulate RNA transcription by controlled methylation of CpG islands at gene promoters. As such, we postulate not only a clinical link, but a potential cellular link between the two disease mechanisms. Furthermore, given the potential association with cardiac defects, we recommend a baseline screening echocardiogram for patients diagnosed with TBRS due to pathogenic variants of *DNMT3A*.

3096W

Aortic dissection in a woman with a variant of uncertain significance in *TGFBR2* gene. D. Aguilar y Mendez¹, G. Torre Amione², F. Castilleja Leal³, A. Rojas Martinez⁴, J. Yañez Sanchez⁵, C. Britton Robles⁵. 1) Centro de Cancer de Mama Tec Salud, Monterrey, Nuevo Leon, Mexico; 2) Instituto de Cardiología y Medicina Vascular, Tec Salud, Tecnológico de Monterrey, San Pedro Garza Garcia, Nuevo Leon, Mexico; 3) Internal Medicine Institute, Escuela Nacional de Medicina Tecnológico de Monterrey, San Pedro Garza Garcia, Nuevo Leon, Mexico; 4) Tecnológico de Monterrey, Escuela Nacional de Medicina, Monterrey, Nuevo Leon, Mexico; 5) Instituto de Cardiología y Medicina Vascular, Centro Medico Zambrano Hellion, San Pedro Garza Garcia, Nuevo Leon, Mexico.

Case report: A 30 y.o. female debuting with a thoracic aortic aneurysm (TAA) was assessed by suspicion of Marfan Syndrome. The patient has not family history of Marfan Syndrome and does not meet the revised Ghent criteria. A molecular panel for variants associated to TAA, including Marfan Syndrome, Loeys Dietz syndrome, vascular Ehlers Danlos, and familial TAA genes associated (*ACTA2*, *CBS*, *COL3A1*, *COL5A1*, *COL5A2*, *EFEMP2*, *FBN1*, *FBN2*, *FLNA*, *MAT2A*, *MED12*, *MYH11*, *MYLK*, *NOTCH1*, *PLOD1*, *PRKG1*, *SKI*, *SLC2A10*, *SMAD3*, *SMAD4*, *SMAD6*, *TGFBR2*, *TGFB3*, *TGFBR1*, *TGFBR2*) was requested, since TAA has genetic etiology in about 5 percent of the cases. The sequence analysis demonstrated the presence of the germline variant of uncertain significance c.1493C>G (p.Pro498Arg) in *TGFBR2*. Variants in TGF-B receptor 2 (*TGFBR2*) are associated to the Loeys-Dietz syndrome (LDS), an autosomal dominant syndrome with widespread systemic involvement characterized by the triad of arterial tortuosity and aneurysms, hypertelorism, and bifid uvula or cleft palate. Mutations in the *TGFBR1* and *TGFBR2* genes are involved in LDS with frequencies of 20-25 and 55-60%, respectively. To assess the pathologic classification of the identified variant, first degree relatives of the proband (2 sibs and 2 children) were clinically examined and evaluated by echocardiography. Both children showed bifid uvula, umbilical hernia and subtle aortic anomalies by echocardiography; one of them had additionally joint hyperlaxity and food allergy, while the other had hypertelorism. Patient sisters showed myopia, food allergy and colitis. Her parents were not studied. Molecular studies also demonstrated the variant c.1493C>G (p.Pro498Arg) in *TGFBR* in both children, while the sisters were negative. We propose that variant c.1493C>G (p.Pro498Arg) in *TGFBR2* is pathologic. This case was solved by a close collaboration between cardiologists and geneticist. Multidisciplinary approach is essential to interpreting clinical relevance of monogenic diseases panel test results in order to provide an accurate, timely genetics diagnosis and ultimately improve the clinical care of patient and in risk relatives with cardiovascular diseases.

3097T

A national registry of patients with inheritable connective tissue disorders to identify genetic modifiers of clinical severity through whole exome sequencing. J.F. Calderon¹, Y. Jimenez¹, L. Cruz-Fernandes¹, G.M. Repetto^{1,2,3}, G. Lay-Son^{1,2,3}. 1) Center for Genetics and Genomics - ICIM, Universidad del Desarrollo, Santiago, RM, Chile; 2) Clínica Alemana de Santiago, Santiago, Chile; 3) Hospital Padre Hurtado, Santiago, Chile.

Aneurysms occur when an artery wall weakens and blood flow comes in between the different layers of the vessel and generates a bulge that can dissect and rupture. Young and middle-aged patients that suffer from aneurysms are usually carriers of genetic variants that predispose for this condition. This group of diseases is called Inheritable Connective Tissue Disorders (ICTDs), and are defined as those caused by mutations in genes that encode for proteins of the Extracellular Matrix (ECM). ICTDs affect patients since birth and their symptoms are present in the musculoskeletal system, skin, eye and respiratory system. However, the most life-threatening feature of these diseases is the enlargement and rupture of big vessels, most prominently, the aorta. These diseases include Marfan Syndrome (MFS), Loeys-Dietz Syndrome (LDS), Ehlers-Danlos Syndrome type IV (vEDS) among others. They are also referred to as TGF β vasculopathies because they are all caused by mutations in genes that code for different effectors or regulators of the Transforming Growth Factor beta (TGF β) signaling pathway, including ECM components that play active roles in the signaling activity of this pathway. We hypothesized that the wide range of phenotypic severity in patients with ICTDs is likely caused by unknown genetic variants in the genome that modify the ECM-TGF β signaling equilibrium beyond the effect of the pathogenic mutation. We aimed to collect a significant number of Chilean patients with ICTDs (n=200) in a database that encompasses every relevant clinical aspect in order to identify these genetic modifiers of clinical severity. Results: We have created a Chilean Registry of ICTDs as a platform for understanding the effect of genetic modifiers in the clinical course of patients with ICTDs. This online database integrates clinical information from many specialists in a patient-centered structure that will also hold Whole Exome Sequencing (WES) data with sufficient complexity to enable us to generate relevant genotype-phenotype correlations for the diseases studied. We have now identified more than 10 families with more than two affected members, with contrasting clinical courses. WES for these subjects is undergoing. With this information, we will interrogate molecular mechanisms behind this variation through in vitro, genome-engineering techniques and translate these findings into improved treatment of Chilean patients. This project is funded by Fondecyt Initiation Grant N°11170353.

3098F

Disruption of the gut barrier modulates lesion burden in mouse models of CCM. C. Gallione¹, A. Peiper¹, H. Pardo¹, E. Griffin¹, R. Lightle², T. Moore², R. Girard², D. Dalldorf¹, N. Hobson², Y. Cao², J. Koskimäki², R. Shenkar², I. Awad², D. Marchuk¹. 1) Duke University Medical Center, Durham, NC; 2) University of Chicago, Chicago, IL.

Cerebral Cavernous Malformations (CCM, MIM 116860) are clusters of dilated capillaries with defective endothelial barrier function, leading to repetitive brain hemorrhage with serious neurologic sequelae. The lesions can occur sporadically or can be inherited as an autosomal dominant trait caused by loss-of-function mutations in one of three genes. Recent work using Cre-inducible deletion of *Ccm* genes in mice has shown that early postnatal lesion development is stimulated by LPS released from the gut microbiome, signaling through the TLR4 receptor, and activating downstream pathways driving lesion genesis. However, the inducible KO model of CCM does not faithfully recapitulate relevant features of the human disease, including later age-of-onset, anatomic lesion location, or chronic hemorrhage. Thus, the intriguing "gut microbiome-to-brain" link of CCM pathogenesis requires further study. We have generated a stochastic, chronic CCM mouse model where the lesions develop chronically over time in the setting of *Ccm* heterozygosity, recapitulating the human disease. We here use this chronic CCM model to test whether the integrity of the gut epithelial barrier plays a role in CCM lesion formation. We administered Dextran Sodium Sulfate (DSS) to adult mice and weanlings in our mouse model and investigated lesion burden and lesion phenotypes. Since DSS is a rather severe treatment (commonly used to create mouse models of ulcerative colitis), we also employed a less severe and more physiologically-relevant method to disrupt the gut barrier. Tween-80 is a common food additive used as a food emulsifier, but chronic administration to mice will disrupt the gut barrier. Mice were thus treated with Tween-80 at a concentration used in processed foods. We found that the mean lesion volumes were significantly greater in brains of mice treated with either DSS or emulsifier compared to placebos, including greater lesional bleeding. These combined data further strengthen the hypothesis that CCM lesion development and hemorrhage in the brain is modulated by the microbial composition of the gut. Furthermore, these data motivate future studies in human CCM cohorts that attempt to reduce lesion burden and pathogenesis in the human disease by modulating the gut microbiome.

3099W

The molecular and physiological roles of ABCC6: More than meets the eyes. O. Le Saux¹, V. Pomozi¹, C. Brampton¹, J. Zoll¹, B. Calio¹, K. Pham¹, G. Kauffenstein^{2,3}, G. Leftheriotis⁴, L. Martin^{2,3}. 1) Cell and Molecular Biology, University of Hawaii John A. Burns School of Medicine, Honolulu, HI; 2) PXE Health and Research Center, University Hospital of Angers, 49045 Angers, France; 3) UMR CNRS 6015-Inserm 1083, School of Medicine, Bretagne Loire University, Angers, France; 4) Faculty of Medicine, University of Nice-Sophia Antipolis, 06107 Nice, France.

Ectopic mineralization of soft tissues occurs in aging, diabetes, hypercholesterolemia, and renal failure as well as genetic disorders. Several mechanisms have evolved to prevent calcification, notably the *ABCC6* gene. *ABCC6* is primarily expressed in liver, kidneys and facilitates the cellular efflux of ATP, which is converted into pyrophosphate (PPi) and adenosine, both major inhibitors of calcification. *ABCC6* deficiencies underlie the disorders pseudoxanthoma elasticum (PXE), and some cases of Generalized Arterial Calcification of Infancy (GACI, also associated with *ENPP1* mutations). The explication of these conditions places *ABCC6* as an upstream modulator of the well-known *ENPP1*→*CD73* pathway generating PPi and adenosine. The clinical manifestations associated with this pathway are a spectrum of related calcification diseases, PXE, GACI, as well as ACDC (Arterial Calcification Due to *CD73* deficiency). Studying the association of typical cardiovascular risk factors (hyperlipidemia, myocardial infarction) and *ABCC6* deficiency has revealed unrecognized facets of *ABCC6* functions probably related to its ability to modulate extracellular purines. Indeed, we found that *ABCC6* haploinsufficiency enhances dyslipidemia and atherosclerosis. Furthermore, we found that mice lacking *ABCC6* suffer less decrement in cardiac function after myocardial infarction than WT controls. Because *ABCC6* functions upstream of *ENPP1* and *CD73* enzymes that influence numerous physiological processes, we investigated the expression of genes implicated in nucleotide and phosphate metabolism in PXE patients and *Abcc6*^{-/-} mice and we found many tissue-specific alterations. Our results demonstrate that *ABCC6*'s primary function, facilitating the cellular efflux of nucleotides (including ATP) influences multiple pathways, giving it a pluripotent physiological role affecting the regulation of calcification, atherosclerosis, and cardioprotection.

3100T

Common variants in TGFβ/BMP9 pathway genes and disease severity phenotypes in hereditary hemorrhagic telangiectasia. L. Pawlikowska^{1,2}, J. Nelson¹, D. Guo¹, K. DeKay¹, K. Yan¹, C.E. McCulloch³, D.A. Marchuk⁴, M.T. Lawton⁵, H. Kim^{1,2,3}, M.E. Faughnan^{6,7}, The Brain Vascular Malformation Consortium HHT Investigator Group. 1) Center for Cerebrovascular Research, Dept of Anesthesia and Perioperative Care, University of California - San Francisco, San Francisco, CA; 2) Institute for Human Genetics, University of California San Francisco, San Francisco, CA; 3) Dept of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA; 4) Dept of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, NC; 5) Dept of Neurosurgery, Barrow Neurological Institute, Phoenix, AZ; 6) Division of Respiratory, Dept of Medicine and Li Ka Shing Knowledge Institute, St. Michael's Hospital, Toronto, Canada; 7) Division of Respiratory, Dept of Medicine, University of Toronto, Toronto, Canada.

Hereditary hemorrhagic telangiectasia (HHT) is an autosomal dominant disorder caused by mutations in TGFβ/BMP9 signaling pathway genes (most commonly in *ENG*, *ACVRL1*, *SMAD4*). HHT patients have skin and mucosal telangiectases that can cause chronic or severe bleeding, and brain, lung and liver vascular malformations (VM) and arteriovenous malformations (AVM) that can lead to ischemic stroke or intracerebral hemorrhage (ICH). The considerable clinical heterogeneity of HHT, even among family members with the same mutation, suggests a possible role for genetic modifier effects. Common variants in *ACVRL1*, *ENG*, *PTPN14* and *ADAM17* were previously reported to be associated with VM in HHT. We hypothesized that other common variants in TGFβ/BMP9 pathway genes modulate HHT disease severity phenotypes.

Methods. We genotyped 760 Caucasian HHT patients recruited by the Brain Vascular Malformation Consortium (BVMC) with the Affymetrix UK Biobank Axiom array. After QC (subject call rate ≥98%, variant call rate ≥99%, minor allele frequency ≥1%, HWE $p \geq 0.0001$), we evaluated 463 variants in or within 5 kb of 23 TGFβ/BMP9 pathway genes for association with 3 HHT severity phenotypes: any organ VM (brain VM, liver VM and/or lung AVM), anemia from severe bleeding (epistaxis and/or gastrointestinal bleeding), and stroke (ischemic stroke from lung AVM and/or intracerebral hemorrhage from brain VM). We used additive models adjusted for age, sex and 20 principal components and evaluated statistical significance for each phenotype using false discovery rates (FDR). **Results.** Among 712 Caucasian subjects, 490 had at least one organ VM (71%), 345 (51%) had anemia and 48 (7%) reported stroke. At FDR $q < 0.25$, no variants were associated with any organ VM, 2 variants in *PTPN14* were associated with anemia and 4 variants (in *ADAM17*, *PTPN14* and 2 in *SMAD7*) were associated with stroke. The strongest finding was for stroke and intronic rs116026081 in *ADAM17* (OR=8.5, $p=0.00002$, FDR $q=0.008$). **Conclusions.** Several common variants in the TGFβ/BMP9 signaling pathway were associated with anemia or stroke in HHT patients. We are replicating these associations in a second, independent cohort of 800 HHT patients. This work extends previously reported association of genetic variation in *ADAM17* and *PTPN14* with lung VM to different HHT severity phenotypes, suggesting that genetic modifiers in these loci may influence multiple aspects of HHT disease.

3101F

Recombinant human ENPP1-Fc protein prevents neointima formation in generalized arterial calcification of infancy by generating AMP. F. Rutsch¹, Y. Yan², I. Buers¹, K. Kintzinger¹, K. Askew², Y. Nitschke¹. 1) Muenster University Children's Hospital, Muenster, Germany; 2) Alexion Pharmaceuticals, 100 College St, New Haven, USA.

Introduction: Generalized arterial calcification of infancy (GACI) is associated with widespread arterial calcification and stenoses often leading to death in early life. Most frequently, the disease is caused by loss of function variants in *ENPP1*. Based on a literature review, arterial stenoses are present in at least 72.4% of GACI cases. *ENPP1* encodes for ectonucleotide pyrophosphatase/phosphodiesterase 1 (ENPP1), which cleaves ATP to generate inorganic pyrophosphate (PP_i) and adenosine monophosphate (AMP) extracellularly. The disease phenotype is recapitulated in *ttw/ttw* mice carrying a spontaneous *Enpp1* nonsense variant and in *Enpp1* knockout mice, the latter of which are known to develop increased neointima formation compared to wild type mice when treated with carotid ligation. This study was designed to identify the mechanism by which ENPP1 deficiency causes neointima formation leading to arterial stenoses and to effectively prevent neointima formation in an animal model of GACI by systemic administration of recombinant human (rh)ENPP1-Fc protein. **Methods and Results:** We evaluated the effect of rhENPP1-Fc on *ENPP1* silenced human vascular smooth muscle cells (VSMC) and on *ttw/ttw* mice treated with carotid ligation. We first demonstrated that silencing *ENPP1* in VSMC results in up to tenfold increase in proliferation relative to that of cells transfected with negative control siRNA. Addition of either rhENPP1-Fc, adenosine monophosphate or adenosine restored silenced *ENPP1*-associated VSMC proliferation. In contrast, neither PP_i nor etidronate, a current off-label treatment for GACI, had an effect on VSMC proliferation. *Ttw/ttw* mice treated with carotid ligation showed increased neointima formation compared to wild type animals reflecting intima proliferation. Subcutaneous administration of rhENPP1-Fc protein 1 mg/kg every other day reduced neointima formation induced by carotid ligation in the *ttw/ttw* mice to the same level as in wild type animals. **Conclusion:** We conclude that ENPP1 inhibits neointima formation by generating adenosine monophosphate. Subcutaneous administration of recombinant human ENPP1-Fc protein may serve as a therapeutic approach to effectively prevent neointima formation and arterial stenoses in GACI.

3102W

Metabolomic signature of angiopoietin-like protein 3 deficiency: Fasting and postprandial effects of loss-of-function genetic variants. E. Tikkanen¹, I. Minicocci², J. Hällström¹, S. Soidinsalo¹, P. Würtz¹, M. Jauhiainen³, V. Olkkonen^{3,4}, M. Arca². 1) Nightingale Health Ltd., Helsinki, Finland; 2)

Department of Internal Medicine and Medical Specialties, Sapienza University of Rome, Italy; 3) Minerva Foundation Institute for Medical Research, Helsinki, Finland; 4) Department of Anatomy, Faculty of Medicine, University of Helsinki, Finland.

Background - Angiopoietin-like 3 (ANGPTL3) is a promising therapeutic target for lowering atherogenic lipid levels and decreasing cardiovascular disease risk. **Aims** - We aimed to determine detailed metabolic effects of ANGPTL3 deficiency. **Methods** - The study population consisted of cases with familial combined hypolipidemia (FHBL2) caused by a loss-of-function mutation in ANGPTL3 gene (N=6 homozygous and N=32 heterozygous carriers) and their controls (N=38). We used nuclear magnetic resonance (NMR) profiling to quantify a wide range of metabolic measures, including lipid concentrations in lipoprotein subclasses, fatty acids, amino acids, and ketone bodies. We compared metabolic signatures of cases and controls in fasting state and after an energy-rich meal. In addition, we used genetic association data from up to 24,925 individuals to compare the metabolic signature of ANGPTL3 loss-of-function (LoF) mutation with an ANGPTL3 common variant (rs2131925) and a major triglyceride-rich lipoproteins-associated genetic variant (rs964184) in the APOC3-APOA5-APOA4-APOA1 gene cluster. **Results** - Individuals with ANGPTL3 deficiency (FHBL2) had markedly reduced levels of triglycerides, total cholesterol, free cholesterol and cholesterol esters in all major lipoproteins and their subclasses. They also showed reduced levels of most fatty acids, citrate and acetate, and elevated 3-beta-hydroxybutyrate, a surrogate biomarker of fatty acid beta-oxidation. Homozygous ANGPTL3 LoF variant carriers showed a significant postprandial reduction of cholesterol in large, medium and small VLDL particles as well as in circulating fatty acids, whereas the postprandial response in heterozygous carriers resembled that of controls. ANGPTL3 LoF, ANGPTL3 rs2131925_G and rs964184_C carriers displayed similar alterations in VLDL and LDL associated lipid parameters, but the common variant-associated HDL levels differed from those of ANGPTL3 LoF carriers. **Conclusions** - In addition to reductions in all major lipoprotein subclasses, ANGPTL3 deficiency results in elevated 3-beta-hydroxybutyrate suggesting enhanced fatty acid beta-oxidation as well as reduced circulating fatty acids putatively alleviating ectopic fat deposition.

3103T

Novel ACTN2 mutations are associated with cardiomyopathy and hypertrophy in human cardiac tissue and iPSC-derived cardiomyocytes. *M.E. Lindholm*^{1,2}, *H. Zhu*^{1,2}, *Y. Huang*¹, *E.A. Ashley*^{1,2}, *M. Wheeler*^{1,2}. 1) School of Medicine, Stanford University, Palo Alto, CA; 2) Stanford Center for Inherited Cardiovascular Disease, Stanford Cardiovascular Institute, Stanford University, Palo Alto, CA.

Introduction In cardiac and skeletal muscle, alpha-actinins are critical cytoskeletal proteins that anchor actin filaments within the sarcomere. Mutations in ACTN2, the isoform present in cardiac and skeletal muscle, have been associated with cardiac abnormalities, including arrhythmias, hypertrophic and dilated cardiomyopathies. However, the mechanisms behind how ACTN2 mutations lead to cardiac dysfunction remain poorly understood. The aim of the present study was to investigate the effects of two novel ACTN2 mutations on cardiac and skeletal muscle phenotypes in human tissue and patient-specific iPSC-derived cardiomyocytes. **Methods and results** We identified patients in the Stanford Center for Inherited Cardiovascular Disease database with ACTN2 mutations. We identified patients with negative clinical panel sequencing, sequenced ACTN2 and found six families with rare or novel ACTN2 variants using a custom mutation pipeline optimized for rare variant discovery. We identified one patient homozygous for a stop-gain mutation (p.Q860X) in ACTN2 and a family with an exon 8-10 deletion. In heart transplant tissue of the homozygous stop-gain patient, we observed mild hypertrophy and interstitial fibrosis. There was no evidence of variation in ACTN2 protein expression, indicating absence of nonsense mediated decay. We used siRNA to knock down ACTN2 in neonatal rat ventricular cardiomyocytes and a human myoblast cell line and observed dramatic changes in cell size and morphology. Patient-derived iPSC-cardiomyocytes were hypertrophic and displayed sarcomeric structural disarray compared to control iPSC-cardiomyocytes. **Conclusions** The molecular effects of ACTN2 on a cellular level and how it causes cardiomyopathy have not been fully elucidated. Here, we provide evidence that two loss of function genetic variants in ACTN2 lead to cardiomyocyte hypertrophy and cardiac abnormalities.

3104F

Patient-specific iPSC disease modeling reveals tissue-specific molecular mechanisms of LMNA-associated cardiac disease. *H.P. Widyastuti*^{1,2,3}, *M. Mehrabi*⁴, *C.H.H. Nguyen*^{2,3}, *L.A. McCarthy*⁴, *A. Grosberg*⁴, *M.V. Zaragoza*^{1,2,3}. 1) Department of Biological Chemistry, University of California, Irvine, Irvine, California, USA; 2) UCI Cardiogenomics Program, Department of Pediatrics, University of California, Irvine, Irvine, California, USA; 3) Sue and Bill Gross Stem Cell Research Center, University of California, Irvine, Irvine, California, USA; 4) Department of Biomedical Engineering and The Edwards Lifesciences Center for Advanced Cardiovascular Technology, University of California, Irvine, Irvine, California, USA.

Lamin A/C are intermediate filament proteins that construct the nuclear lamina of a cell encoded by the LMNA gene. Lamin A/C has been implicated in processes such as chromatin localization and cell differentiation. Consequently, mutations in LMNA gene have detrimental effect in cell functions which results in diseases collectively termed as laminopathies which include LMNA-related dilated cardiomyopathy. In a large affected family, we recently identified a heterozygous novel LMNA splice-site mutation (LMNA c.357-2A>G) predicted to result in exon skipping and premature truncation. Our group aims to elucidate the molecular mechanisms contributing to the pathology of LMNA-related cardiac disease by developing an in-vitro, patient-specific disease model. Dermal fibroblasts, obtained from affected and unaffected family members, were reprogrammed to induced pluripotent stem cells (iPSCs). One control and one patient-derived iPSCs were differentiated into cardiomyocytes (iPSC-CMs) for molecular and protein studies. Sanger sequencing of LMNA transcripts revealed evidence of the predicted exon skipping and frame shift exclusively in patient iPSC-CMs, indicating the expression of the mutant allele albeit at a low level compared to the normal allele. Western Blot studies showed that patient iPSC-CMs had a reduced level of Lamin A/C protein compared to control iPSC-CMs but no evidence of the mutant truncated protein. Interestingly, when the same studies were performed on patient and control fibroblast, mutant allele expression or reduction in Lamin A/C protein level was not observed. Taken together, our results provide insight to the molecular mechanism of LMNA-associated cardiomyopathy and reveal the tissue specificity of this mutation.

3105W

Associations of inflammatory cytokines (IL-6 and IL-16) with coronary artery disease in Sudan. *H.H. Musa¹, M.M. Yahia¹, I.H. Musa², E.A. Elbahir³.*

1) Medical Microbiology, Faculty of Medical Laboratory Sciences, University of Khartoum, Khartoum, Sudan; 2) Sudan Medical Specialization Boards, Khartoum, Sudan; 3) Sudan Heart Center, Khartoum, Sudan.

Background: Cardiovascular disorders are the second most common causes of adult deaths in sub-Saharan Africa. Lifestyle and environmental factors play an important role in coronary artery disease (CAD) development, and family clustering suggests a genetic predisposition. IL-6 and IL-16 play important role in the processing of inflammatory disease such as coronary artery disease. IL-6 was associated with acute ischemic condition, while IL-16 is multifunctional cytokines that was initially recognized as chemo-attract factor. The aim of this study was to detect the association of IL-6 and IL-16 genes with CAD in Sudanese population. **Methods:** A case control study was conducted for coronary artery disease patients who attended the Sudan Heart Institute and healthy controls from January 2014 to April 2016. Demographic data and medical histories were collected using a structured questionnaire. Blood samples from each participant were collected to estimate lipids profiles and DNA extraction. Total cholesterol, triglyceride, low density lipoprotein and high density lipoprotein were measured using an enzymatic colorimetric method. Genomic DNA was isolated using the G-spin Kits (INTRON Biotechnology). Mutation in IL-6 and IL-16 genes was detected using PCR-RFLP.

Results: Among the patients 64% were males and 41% were at 51-61 years, 63% were from Northern Sudan, physical activity was common in 60% of the patients. The risk factors for coronary artery disease were hypertension 62%, angina 62%, smoking 38%, diabetes 33%, tobacco user 31%, alcohol 24% and stroke 2%. Most patients have family history of stroke 93%, hypertension 36%, diabetes 34%, angina 9%, and heart attack 4%. Total cholesterol and low density lipoprotein were significantly different ($P > 0.05$) between cases and controls, and they were significantly associated with IL-16 genotype. While, IL-6 genotype has no significant effect on lipids markers. Multivariate logistic analysis indicated that all risk factors for coronary artery diseases were strongly associated with IL-6 genotype (P value < 0.05 , OR > 1). Whereas, IL-16 genotype was significantly associated with gender and physical activity. **Conclusion:** The study conclude that IL-6 and IL-16 genes can be use as a risk factor for coronary artery disease in Sudan.

3106T

Atrial fibrillation genetic risk differentiates cardioembolic stroke from other stroke subtypes. *C.D. Anderson^{1,9,10}, S.L. Pulit^{1,2,3}, L.C. Weng^{1,4}, P.F. McArdle⁵, L. Trinquart^{6,7}, S.H. Choi⁷, B.D. Mitchell^{8,9}, J. Rosand^{1,9,10}, P.I.W. de Bakker^{2,11}, E.J. Benjamin^{8,12,13}, P.T. Ellinor^{1,4,14}, S.J. Kittner^{5,16}, S.A. Lubitz^{1,4,14}.*

The Atrial Fibrillation Genetics Consortium, The International Stroke Genetics Consortium. 1) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA; 2) Department of Genetics, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands; 3) Li Ka Shing Centre for Health Information and Discovery, The Big Data Institute, University of Oxford, Oxford, UK; 4) Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA; 5) Department of Medicine, Program for Personalized and Genomic Medicine, University of Maryland School of Medicine, Baltimore, Maryland, United States of America; 6) National Heart, Lung, and Blood Institute's and Boston University's Framingham Heart Study, Framingham, MA, USA; 7) Department of Biostatistics, Boston University School of Public Health, Boston, MA; 8) Geriatrics Research and Education Clinical Center, Baltimore Veterans Administration Medical Center, Baltimore, MD, USA; 9) Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA; 10) J.P. Kistler Stroke Research Center, Massachusetts General Hospital, Boston, MA, USA; 11) Computational Genomics, Vertex Pharmaceuticals, Boston, MA, USA (current address); 12) Cardiology Preventive Medicine Sections, Evans Department of Medicine, Boston University School of Medicine, Boston, MA, USA; 13) Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA; 14) Cardiac Arrhythmia Service, Massachusetts General Hospital, Boston, MA, USA; 15) Department of Neurology, University of Maryland School of Medicine, Baltimore, MD, USA; 16) Department of Neurology, Veterans Affairs Medical Center, Baltimore, MD, USA.

Atrial fibrillation (AF) affects ~34 million individuals worldwide and is associated with a five-fold increased risk of ischemic stroke. Strokes caused by AF are classified as cardioembolic, and secondary prevention of cardioembolic (CE) stroke focuses on identifying AF as a potential cause and preventatively treating with anticoagulants. However, presence of AF can remain occult even after extensive workup, and whether AF in the context of stroke carries the same stroke risk as pre-existing AF is unknown. We thus sought to understand whether AF genetic risk is an important and potentially useful determinant of CE stroke risk. We first assessed whether stroke patients with AF have a genetic predisposition to AF and then evaluated the extent to which genetic risk factors for AF capture the heritability of CE stroke. We performed a genome-wide association of AF and CE stroke in 13,390 ischemic stroke cases and 28,026 referents. Associated SNPs ($p < 5 \times 10^{-8}$), in particular at the *PITX2* and *ZFHX3* loci, were consistent in effect across the two phenotypes, as expected. Using AF-associated SNPs ($p < 4.4 \times 10^{-4}$) from an independent meta-analysis, we observed strong correlation between previously-reported AF genetic risk, AF in the presence of stroke, and CE stroke (Pearson's $r = 0.77$ and 0.76 , respectively). Heritability estimates revealed AF and CE stroke to be similarly heritable (20.0% and 19.5%, respectively). We next derived an AF polygenic risk score (PRS) using 934 previously-implicated AF SNPs. The PRS, adjusted for clinical AF risk factors, was associated with CE stroke (odds ratio (OR) per standard deviation (SD) = 1.40, $p = 1.45 \times 10^{-6}$) and explained ~20% of the heritable component of CE stroke risk. The AF PRS also associated with stroke of undetermined cause ($p = 0.004$), but none of the other primary stroke subtypes (all $p > 0.1$). Despite the fact that CE stroke often affects older adults with multiple clinical comorbidities, increasing AF risk due to non-genetic factors, our results suggest that CE stroke patients have enrichment for AF genetic risk. Future work in AF and CE stroke may reveal how PRSs based on AF-associated SNPs could be used either to predict individuals at highest risk of CE stroke to be more closely monitored, or help distinguish between clinical etiopathologic subtypes of stroke.

3107F

An oligogenic model for congenital heart defects: whole-genome sequencing of congenitally corrected transposition of great arteries shows a mutation load effect at 228 loci involved in cardiac patterning.

S. Benko¹, F. Bajolle², Y. Benadjaoud¹, A. Desgrange³, F. Jabot-Hanin⁴, J-F. Delouze⁵, P. Nitschke⁴, S. Meilhac³, J. Amiel¹, D. Bonnet⁵, S. Lyonnet¹. 1) UMR-1163 INSERM, Institut Imagine and Université Paris Descartes, Paris, France; 2) Pediatric Cardiology, Hôpital Necker - Enfants malades, Paris, France; 3) Imagine-Institut Pasteur group of Heart Morphogenesis; 4) BIP-D Plateforme de Bioinformatique Paris-Descartes, Institut Imagine, Paris, France; 5) Centre National de Génotypage, Paris, France.

Congenital heart disease (CHD) is a major public health issue. About 20% of CHD correspond to misconnections between cardiac segments. Aiming to understand the genetic mechanisms that controls the alignment of cardiac segments, we focused on the congenitally corrected transposition of the great arteries (CCTGA), a typical example of cardiac misalignment, expected to result from anomalies of cardiac tube looping and outflow tract (OFT) rotation. CCTGA is characterized by a left/right inversion of the ventricles and the transposition of the great arteries, corresponding thus to a double atrio-ventricular and ventriculo-arterial discordant connections. We established a cohort of 47 non-syndromic and mostly sporadic CCTGA patients. The DNAs of CCTGA trios (patient and both parents) and siblings were submitted to next generation sequencing. The first analyses of our dataset under the hypothesis of *de novo* genomic alterations, did not reveal shared major gene(s). The data were further analysed under the hypothesis of a complex model of inheritance with incomplete penetrance. Candidate causative variations were prioritized using four criteria: (1) low frequency in public databases (<1%), (2) prediction of a deleterious effect, (3) evolutionary conservation, and (4) location in a gene involved in heart tube morphogenesis, OFT development and/or left-right asymmetry patterning (1,465 genes). A total gene set of 228 genes matched each of these criteria, with a mean number of 6.5 deleterious variant per affected case. We hypothesized that the highly heterogeneous combinations of susceptibility rare variants, mostly inherited from the healthy mother and father respectively, functionally converge to give rise to the CCTGA phenotype. To prove that a given allele combination is associated with CCTGA, we also showed a highly significant enrichment of rare variant in cases vs. parents and controls, any given deleterious variant combination within the CCTGA gene set being specific to the affected individual. Taken together, 44/47 CCTGA cases could be explained by a complex oligogenic mutation load of segregated alleles at loci mostly involved in heart tube looping, OFT morphogenesis and establishment of left-right asymmetry. Finally, the set of 228 CCTGA loci, comprised several genes whose LOF mutations are known to result in syndromic CHD, suggesting the CCTGA alleles are hypomorphic, and that CCTGA is a laterality defect limited to a cardiac segment.

3108W

Recurrent copy number variation in patients with thoracic aortic aneurysm affects dosage sensitive genes expressed in aorta. E. Brosens¹, J.M.A. Verhagen¹, H.T. Brüggewirth¹, T.J. van Ham¹, R.W.W. Brouwer², H. Eussen¹, R.M. van der Helm¹, T. Brands¹, W.F.J. van Ijcken³, W.G. de Valk¹, J.E.M.M. de Klein¹, H.B. Beverloo¹, M.A. van Slegtenhorst¹, R.M.W. Hofstra¹, J.A. Bekkers³, J.W. Roos-Hesselink⁴, M.W. Wessels¹, I.M.B.H. van de Laar¹.

1) Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, Zuid-Holland, Netherlands; 2) Center for Biomix, Erasmus University Medical Center, Rotterdam, Zuid-Holland, Netherlands; 3) Department of Cardiothoracic Surgery, Erasmus University Medical Center, Rotterdam, Zuid-Holland, Netherlands; 4) Department of Cardiology, Erasmus University Medical Center, Rotterdam, Zuid-Holland, Netherlands.

Thoracic aortic aneurysms (TAA) are life-threatening. As approximately 20% of patients have at least one affected first-degree family member, it often is considered an inherited disorder. Genes of the Transforming growth factor- β signaling and vascular smooth muscle cell contractility pathway have emerged as key players in TAA pathogenesis. Environmental factors and genetic modifiers, including copy number variants (CNVs), are thought to explain the clinical variability. Genes with changes in expression in the aorta as a result of CNVs could contribute to TAA formation. Therefore, we determined the CNV profiles of 167 patients and prioritized CNVs based on gene content and population frequency. Next, we quantified gene expression in 5 control aortic biopsies, 1 patient with Turner syndrome and 15 patients without a genetic diagnosis. We could confirm the presence of previously described modifying CNV in our cohort: 3p11 loss (n=2), 16p13 gain (n=2) and 22q11 gain (n=3). Moreover, we identified 67 rare CNVs, 54 of which were non-recurrent. CNVs were enriched for genes involved in fibroblast growth factor signaling and extracellular matrix organization. Four genes affected by a recurrent loss (including *EPHA3*) and 15 genes by a recurrent gain (including *MYH11*) had expression in aorta. Of the genes affected by a CNV 5 (*NAT8L*, *SLC19A3*, *LNFG*, *FGF7* and *ABCC6*) were differentially expressed between patient and control cohorts. Screening of 121 TAA patient exomes for variation in genes affected by the identified CNV (n=289) resulted in the identification of 63 putative deleterious variants in 24 genes. We did not detect loss of function variants in dosage sensitive genes (n=31), nor were recessive variants unmasked by a CN loss: all variants (n=18) in genes intolerant to recessive variation (n=44) were heterozygous. The expression of 5 genes on chromosome X in the Turner patient was reduced compared to controls: *PLCXD1*, *BMX*, *PLP1*, *GPM6B* and *F8A1*. Interestingly, the main TAA candidate gene on chromosome X (*BGN*) was highly expressed, but within the control aorta expression range. The impact on vascular development of changes in expression of main candidate genes is currently evaluated using zebrafish genetic models. To conclude, we could confirm the presence of previously described modifier CNVs and found new rare recurrent CNVs. CNV could contribute to TAA formation substantially as many patients have a rare CNV involving a gene expressed in the aorta.

3109T

A genome-wide polygenic risk score for low-density lipoprotein cholesterol is associated with coronary artery disease risk in >400,000 individuals in the UK biobank. K. Chaudhary, G. Nadkarni, R. Do. The Charles Bronfman Institute for Personalized Medicine, Department of Genetics and Genomics, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.

Plasma levels of low density lipoprotein cholesterol (LDL-C) is one of the strongest risk factors for coronary artery disease (CAD). Genome wide association (GWA) studies have identified associations between single nucleotide variants (SNVs) and LDL-C levels. Furthermore, polygenic risk scores (PRS) for LDL-C have been shown to be strongly associated with LDL-C and coronary artery disease (CAD) risk. In the present study, we evaluated the predictive power of PRS generated from different sets of SNVs ranging from LDL-C significant SNVs to the full set of SNVs genome-wide. We calculated PRSs using GWA summary statistics for LDL-C from the Global Lipids Genetics Consortium and tested for association with CAD risk in 21,000 cases and 387,961 controls of self-identified White British ancestry in the UK Biobank. We calculated PRS using three different sets of SNVs: (a) 138 loci significantly associated with LDL-C; (b) 46,766 SNVs using pruning thresholds of $r^2 < 0.8$ and association $P < 0.05$; and (c) the full set of 2,373,476 SNVs genome-wide with effect sizes adjusted for linkage disequilibrium (LD) using LDpred. We calculated the area under the receiver operating characteristic curve (AUROC), and performed logistic regression after adjusting for age, sex, four principal components and type of genotyping array. Across all three sets of PRS, we observed associations of the PRS generated from the 138 LDL-C loci with CAD risk when comparing the top 0.25 percentile of the PRS to the remaining 99.75 percentile (OR=1.4, $P < 0.005$). Stronger associations were observed for both the PRS generated from pruning (OR=1.9, $P < 1 \times 10^{-6}$) and LDpred (OR=1.5, $P = 0.0006$). In general, we observed an increase in the magnitude of the association between PRS and CAD risk in the extremes of the PRS distribution. In conclusion, PRS for LDL-C is strongly associated with CAD risk in a large study sample of White British individuals. Further work is in progress in examining additional sets of LDL-C SNVs for the effect of PRS on CAD risk. You may contact the first author (during and after the meeting) at kumardeep.chaudhary@mssm.edu.

3110F

The role ROR-alpha target genes post-mortem advanced atherosclerotic plaques and patients with CAD. N. Coban¹, A.F. Erkan², C. Gulec¹, B. Ekiç², F. Guclu-Geyik¹, M. Cavlak³, T. Erginel⁴, N. Erginel-Unaltuna¹. 1) Department of Genetic, Aziz Sancar Institute for Experimental Medicine, Istanbul University, , Istanbul, Turkey; 2) Department of Cardiology, Medical Faculty, Ufuk University, Ankara, Turkey; 3) Hacettepe University Faculty of Medicine, Department of Forensic Medicine, Ankara; 4) General Surgery Department, Istanbul Training and Research Hospital, Istanbul.

Aim: ROR-alpha is a member of nuclear receptor family of transcription factors, and therefore has ligand-dependent activity. Melatonin and cholesterol are two known ligands of ROR-alpha both were shown to be important for atherosclerosis. Aim of the present study was determined the expression profiles of *ROR-alpha* and its target genes (*ABCA1*, *CYP19A1* and *MIF*) in post-mortem advanced atherosclerotic plaques. In addition, the polymorphisms of two target genes have been investigated and associated with coronary artery disease (CAD). **Methods:** The advanced atherosclerotic plaques (CHD-P) and without plaque (CHD-nonP) artery samples from post-mortem cases with coronary heart disease (CHD). The *ROR-alpha*, *ABCA1*, *CYP19A1* and *MIF* expression levels were analyzed using qRT-PCR technique. In addition, unselected 332 Turkish CAD (coronary lesion with 50-100% stenosis) and control (coronary lesion with <30% stenosis) were genotyped for *MIF* and *CYP19A1* polymorphism using hybridization probes in RealTime PCR LC480 device. Blood samples were drawn before coronary angiography. Gensini and SYNTAX scores and myocardial blush grade (MBG) were assessed. **Results:** Expression results showed two target genes (*ABCA1* and *MIF*) were up-regulated and a target gene (*CYP19A1*) were down-regulated (fold change >1.5) in advanced atherosclerotic plaques. The *ROR-alpha* was not expressed significantly lower in advanced atherosclerotic plaques (CHD-P) as compared to CHD-nonP (Fold change, 0.9). In addition, *CYP19A1* rs10046 C/T, not *MIF* gene -173 G/C, polymorphism was significantly associated with the angiographic severity in patients with CAD ($p < 0.05$). **Conclusion:** It shows that reduced expression of *CYP19A1* may play an effective role in the development of atheroma. The expression levels of *ROR-alpha* some target genes were found to associations with advanced atherosclerosis in CHD cases causing sudden death.

3111W

Gain of function variation in *SOS1* as a cause of dilated cardiomyopathy.

J.R. Cowan, L. Salyer, D.D. Kinammon, A. Morales, E. Jordan, R.E. Hershberger. Division of Human Genetics, Department of Internal Medicine, The Ohio State University, Columbus, OH.

Background and significance: Dilated cardiomyopathy (DCM) is a genetically heterogeneous cardiac disease characterized by progressive ventricular enlargement and reduced systolic function. Genetic overlap is observed between DCM and hypertrophic cardiomyopathy (HCM), a primary cardiac manifestation in several distinct disorders arising from defective Ras signaling, including Noonan syndrome (NS). Here, we report genetic and functional analyses implicating the Ras signaling protein, *SOS1*, in the pathogenesis of DCM. **Methods and Results:** Exome sequencing identified an A-band Titin truncating variant [*TTN*tv; p.(Val30827Serfs*22)] in a large, multiplex pedigree with aggressive, early-onset DCM. Reduced severity in individuals carrying only the *TTN*tv indicated that it was unlikely to be the sole genetic factor operating in this family. The presence of a rare *SOS1* variant [p.(Glu191Lys)] cosegregating alongside the *TTN*tv in the most severely affected family members suggested its potential involvement as an additional risk factor. Exome sequencing results from four other families identified five additional *SOS1* variants with potential disease involvement [p.(Ile607Thr), p.(Gly719Ala), p.(Arg744Gly), p.(Asp910His), p.(Arg1201Trp)]. Impacted amino acids occupied a number of functional domains relevant to *SOS1* activity, including the N-terminal histone fold, as well as the C-terminal catalytic unit. As several lines of evidence have implicated dysregulated Ras signaling in the pathogenesis of DCM, we assessed functional impact of each variant on activation of ERK and AKT. Relative expression levels were determined by western blot in HEK293T cells transfected with variant or wildtype human *SOS1* expression constructs. Increased pERK expression relative to wildtype levels was seen for all six *SOS1* variants, with fold differences for four reaching statistical significance (1.58 - 2.47; p-values 0.0002 - 0.0311). No significant differences in AKT signaling were observed for any variants. These signaling profiles parallel those observed in NS and support adding DCM to the clinical spectrum associated with gain of function variation in *SOS1*. **Conclusions:** Our analyses provide the first reported evidence for gain of function variation in *SOS1* as a cause of DCM.

3112T

The effect of genetic determinants of arsenic metabolism efficiency on hypertension risk.

D.A. Delgado¹, L.S. Chen¹, F. Jasmine¹, M.G. Kibriya¹, T. Islam², A. Ahmed², Y. Chen³, H. Ahsan^{1,4,5,6}, B.L. Pierce^{1,4,5}. 1) University of Chicago, Chicago, IL; 2) UChicago Research Bangladesh, Dhaka, Bangladesh; 3) Department of Environmental Medicine, New York University School of Medicine, New York, NY; 4) Department of Human Genetics, University of Chicago, Chicago, IL 60615; 5) Comprehensive Cancer Center, University of Chicago, Chicago, IL 60615; 6) Department of Medicine, University of Chicago, Chicago, IL 60615.

Epidemiologic studies report inconsistent associations between inorganic arsenic (iAs) exposure and hypertension (HTN) at high and low-to-moderate levels of exposure. One limitation of these prior studies is the lack of individual-level measurements of arsenic and arsenic metabolites in urine. In the current study, we leverage SNPs known to influence arsenic metabolism efficiency in a Mendelian randomization (MR) approach to understand the relationship between arsenic exposure and HTN risk. We used data on three genetic variants known to influence arsenic metabolism efficiency (*AS3MT* SNPs rs9527 and rs11191527; and *FTCD* SNP rs61735836) and HTN and blood pressure phenotypes on 4,886 Bangladeshi individuals with a wide range of exposure to arsenic through naturally-contaminated drinking water. Arsenic metabolism efficiency is represented by the percentage of arsenic in urine that is dimethylarsinic acid (DMA), the end metabolite in a series of reduction and methylation reactions that act as a detoxification mechanism. We used two MR methods to estimate the association between arsenic-related SNPs and HTN risk: 1) a two-stage sequential regression, which uses individual level data on a subset of 1,663 individuals with arsenic metabolite data to generate genetically-predicted values of DMA% and 2) the inverse-variance weighted (IVW) and simple median meta-analysis methods, which uses summary statistics on up to 4,886 individuals. The high efficiency alleles of rs9527, rs11191527, and rs61735836 were strongly associated with DMA% ($\beta = 3.7$, $P = 1 \times 10^{-11}$; $\beta = 2.1$, $P = 6 \times 10^{-8}$; and $\beta = 5.4$, $P < 2 \times 10^{-16}$; respectively) and associated with increased odds of HTN, though not significantly (OR = 1.15, OR = 1.07, and OR = 1.04, respectively). Both MR methods suggest increasing DMA% is associated with increased odds of HTN, though the effect is not significant with the two-stage sequential regression (OR = 1.016 (0.994, 1.038)) and IVW method (OR = 1.020 (0.997, 1.043)). Under the simple median MR method, the high efficiency alleles of the arsenic-related SNPs are significantly associated with increased odds of HTN (OR = 1.036 (1.001, 1.075)). Analyses using continuous measures of systolic and diastolic blood pressure showed consistent results. In summary, we leveraged individual genetic variation in arsenic methylation capacity to evaluate the effect of DMA% on HTN risk. Under MR assumptions, our findings suggest efficient metabolism of iAs increases the risk of HTN.

3113F

Rare variants in previously identified linkage regions associated with carotid plaque in Dominican families. *N. Dueker¹, A. Beecham¹, L. Wang^{1,2}, S.H. Blanton^{1,2}, C. Dong³, B. Doliner¹, R.L. Sacco^{2,3}, T. Rundek³.* 1) John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL; 2) Dr. John T. Macdonald Department of Human Genetics, University of Miami, Miami, FL; 3) Departments of Neurology and Public Health Sciences, Miller School of Medicine, University of Miami, Miami, FL.

Carotid plaque is a subclinical marker of atherosclerosis and a well-established predictor of stroke and cardiovascular disease (CVD). We have previously shown measures of carotid plaque to be heritable in a sample of 100 Dominican families and found evidence for linkage and association of common variants (CVs) on 7q36, 11p15, 14q32 and 15q23 with plaque presence. Our current study aimed to further characterize these QTLs by identifying rare variants (RVs) (minor allele frequency < 5%) associated with plaque presence. To achieve this aim we sequenced the 1 LOD unit down region on 7q36, 11p15, 14q32 and 15q23 in twelve Dominican families with evidence for linkage to plaque presence. Using these data, we performed RV gene-based analyses on 454 genes within the four QTLs using the Sequence Association Test for familial data (F-SKAT). Gene-based analyses were performed under two filtering algorithms; exonic RVs and non-synonymous RVs. A Bonferroni correction was applied to correct for 908 tests ($p < 5.5 \times 10^{-5}$). Within the twelve families, sequencing identified 11,769 RVs, of which, 623 were non-synonymous. In exonic RV analyses, three genes showed suggestive association with plaque presence and nominal evidence for replication in an additional sample of 22 Dominican families with Exome Array data; *PARVA* ($p=0.03$ in sequenced and replication families), *DPP6* ($p=0.006$ sequenced families; $p=0.03$ replication families) and *NOM1* ($p=0.02$ sequenced families; $p=0.03$ replication families). We previously reported CVs in *PARVA* and *DPP6* to be associated with plaque; therefore, the current study extends these findings to suggest that RVs within these genes may also play a role in carotid plaque pathogenesis. Variants in these genes have been associated with coronary artery disease (*PARVA* and *DPP6*), heart failure mortality (*PARVA*) and idiopathic ventricular fibrillation (*DPP6*), making them excellent candidate genes for carotid plaque. In non-synonymous RV analyses, *NRDE2* ($p=0.04$ in sequenced families; $p=0.002$ in replication families) and *ZNF143* ($p=0.002$ sequenced families; $p=0.04$ replication families) showed suggestive association in the sequenced families and nominal evidence for replication. Taken together, these results provide evidence for a role of RVs within our previously-identified linkage regions on 7q36, 11p15, 14q32 and 15q23 in carotid plaque presence.

3114W

TRPV4 variants found to be suggestively associated with lung, renal and menstruation phenotypes in BioVU fail to replicate in UK Biobank and GERA cohorts. *M.G. Ehm¹, D.K. Ranatunga², K.K. Thai², M.N. Chiano¹, O. Kirichek¹, A. Lawton¹, T.J. Hoffman³, D.J. Behm¹, N. Risch², A.S. Go², C.A. Schaefer², D.M. Waterworth¹.* 1) GlaxoSmithKline, Research & Development, Collegeville, PA & Stevenage, UK; 2) Kaiser Permanente, Division of Research, 2000 Broadway, Oakland, CA 94612; 3) UCSF, 513 Parnassus Avenue, San Francisco CA 94117.

Transient receptor potential vanilloid-4 (*TRPV4*) encodes a Ca^{2+} -permeable, nonselective cation channel thought to be involved in the regulation of systemic osmotic pressure, being investigated pre-clinically for various indications including brain, cardiovascular and joint diseases. A *TRPV4* blocker (GSK2798745) was in development for pulmonary edema associated with heart failure. Current mechanistic understanding derives mainly from preclinical data. A phenotype-wide association study (PheWAS) for variants (*TRPV4* V562I and E840K) was conducted using 29,000 participants typed using exome genotyping chip and evaluated for 1,502 phenotypes within BioVU, a hospital electronic health record (EHR) linked biobank. While none of the results were statistically significant after correction for multiple testing, associations were observed between p.E840K variant and pulmonary edema (odds ratio [OR] 1.97, $p=4 \times 10^{-4}$) as well as five lung infection phenotypes (OR range 1.7-3.1, $p < 0.01$), renal failure and menstruation-related disorders ($p < 0.05$). These results aligned well with preclinical data and with human *TRPV4* expression, which is highest in the kidney and bronchial epithelia. We conducted two studies to assess if these results replicated: the UK Biobank (UKB) and Kaiser RPGEH GERA cohort, a population-based health care system sample of 100,000 with linked genotypic and EHR data. Both studies evaluated associations of genotyped or well-imputed ($R^2 > 0.80$) *TRPV4* variants p.V562I, p.E840K and a promoter variant (rs4766635) with MAFs of 0.012, 0.011 and 0.484, respectively with 21 phenotypes. No significant associations were observed between the *TRPV4* variants and phenotypes tested, either in GERA or UKB cohorts, nor the meta-analysis despite 80% power to detect genetic effects that overlap effect ranges identified at BioVU. The lack of replication could be due to the differing case settings across the cohorts, differing composition of cohorts with respect to selection of controls, differing uses of ICD codes to describe underlying clinical conditions or the initial association results were possibly false positives. Availability of multiple large, diverse, well-powered cohorts provide the opportunity to efficiently and quickly investigate genetic evidence for drug targets using real-world data which provides multiple insights into the epidemiology and biology of drug targets and associated conditions.

3115T

Transcriptome signature of hypertension in African Americans: Findings from machine learning. A. Gaye, S. Davis, G. Gibbons. National Institutes of Health, Bethesda, MD.

Background In multifactorial disease or related quantitative traits, the mechanism underlying the phenotype necessarily involves multiple determinants. Therefore, the full picture is more likely to emerge from investigations that integrate a range of interacting determinants. The aim of this work is to use a machine learning approach and leverage analysis of the transcriptome for the identification of functional/putative causal DNA variants that contribute to the pathobiological pathways that mediate a complex trait such as Hypertension. **Methods** The data consist of mRNA sequencing from whole-blood and genotype data (multi-ethnic array) from 178 participants of the Minority Health Genomics and Translational Research Bio-Repository Database (MH-GRID) study. A total of 16223 mRNAs (whole transcriptome) were used as predictors in a random forest classification analysis of Hypertension (HTN) to identify the set of mRNAs that predict HTN. Then the cis-variants that predict the expression of that set of mRNAs were identified. Finally, that group of cis-variants was linked back to the phenotype and cis-variants that predict Hypertension with high performance were identified. **Results** We identified 34 mRNAs that classify hypertensive vs. control with an area under the curve (AUC) = 0.93. Two of the mRNAs are encoded by genes that knock out animal models showed to be associated with vascular smooth muscle cell proliferation and abnormal development of intersegmental vessels. Three of the 92 cis-variants that, together, classify HTN with AUC=92 have been previously reported in GWAS of cardiovascular related phenotypes such as obesity, body mass index and heart rate. **Conclusion** Conventional statistical methods have provided extremely robust findings. However, due to various challenges many weaker but genuine signals are lost in the process whilst complex interactions are difficult to model. In this work, we used random forest classification to provide new insights and extend our knowledge. This unbiased approach revealed a set of mRNAs enriched for pathways that are plausible modulators of HTN by modulating vascular tone and/or vascular remodeling. The cis-variants that influence the expression of those mRNAs proved to be very good predictors of HTN.

3116F

Unraveling the relationship between blood pressure and kidney function: Genetic risk score phenome-wide association studies in the Million Veteran Program. A. Giri^{1,2}, J.N. Hellwege^{2,3}, J.M. Keaton^{2,3}, S.M. Damrauer⁴, S.L. DuVal⁵, C. Robinson-Cohen^{2,6}, E.S. Torstenson^{2,3}, O.D. Wilson^{2,6}, C.P. Chung⁷, K.A. Birdwell^{8,9}, Y.V. Sun², D. Klarin^{10,11}, C.P. Kovesdy², E.E. Siew^{6,13}, A. Kottgen¹⁴, C. Pattaro¹⁵, M. Wuttke¹⁴, M. Gorski¹⁶, A. Teumer¹⁷, P.S. Tsao¹⁸, C.J. O'Donnell¹⁹, P.W.F. Wilson²⁰, D.R. Velez Edwards^{1,2}, T.L. Edwards^{2,3}, A.M. Hung^{2,6}, CKDGen Consortium, on behalf of the VA Million Veteran Program.

1) Department of Obstetrics and Gynecology, Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA; 2) Biomedical Laboratory Research and Development, Tennessee Valley Healthcare System (626)/Vanderbilt University, Nashville, TN, USA; 3) Division of Epidemiology, Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA; 4) Department of Surgery, Corporal Michael Crescenz VA Medical Center & Department of Surgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia PA, USA; 5) VA Salt Lake City Health Care System, University of Utah School of Medicine, Salt Lake City, Utah, USA; 6) Division of Nephrology and Hypertension, Vanderbilt University Medical Center, Nashville, TN, USA; 7) Divisions of Rheumatology and Clinical Pharmacology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA; 8) Division of Nephrology, Department of Medicine, Nashville Veteran Affairs Hospital, Nashville, TN, USA; 9) Department of Epidemiology, Emory University Rollins School of Public Health, Department of Biomedical Informatics, Emory University School of Medicine, Atlanta, GA, USA; 10) VA Boston Health Care System, Boston, MA; Center for Genomic Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA; 11) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA; Department of Surgery, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA; 12) Nephrology Section, Memphis VA Medical Center, Memphis TN, USA; 13) Geriatrics Research Education and Clinical Center, Tennessee Valley Health System, Veteran's Health Administration, Nashville, TN, USA; 14) Institute of Genetic Epidemiology, Albert Ludwigs University of Freiburg, Germany; 15) Institute for Biomedicine, Eurac Research, Affiliated Institute of the University of Lübeck, Bolzano, Italy; 16) Department of Genetic Epidemiology, University of Regensburg, Regensburg, Germany; 17) Institute for Community Medicine, University Medicine Greifswald, Germany; 18) VA Palo Alto Health Care System; Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA, USA; 19) VA Boston Healthcare, Section of Cardiology and Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; 20) Atlanta VAMC and Emory Clinical Cardiovascular Research Institute, Atlanta, GA, USA.

Blood pressure (BP) homeostasis and kidney function are intimately linked, and chronic kidney disease (CKD) and hypertension (HT) often occur concurrently. However, whether high BP precedes CKD or gradually deteriorating kidney function is also a primary cause of HT is controversial, and can be confounded by subclinical disease in observational studies. We used genetic risk scores (GRS) for BP traits (systolic [SBP], diastolic [DBP] and pulse pressure [PP]) and kidney function (estimated glomerular filtration rate [eGFR]) to assess associations with clinically diagnosed phenotypes in a phenome-wide association study (PheWAS) in 188,008 European Americans from the Million Veteran Program (MVP). GRSs used European-derived effect estimates from the United Kingdom Biobank (UKB) study for BP-traits (N-SNPs = 545, N-Samples = 141k) and the CKDGen consortium for eGFR (N-SNPs = 128, N-Sample range = 239k-567k) for trait-increasing alleles from GWAS index single nucleotide polymorphisms (SNPs). We regressed 1,813 phenotypes onto each GRS using logistic regression adjusting for 10 principal components. SBP, BP and PP-weighted results were highly correlated. HT ($P < 1 \times 10^{-305}$) and essential HT ($P < 1 \times 10^{-305}$) were the strongest associations with SBP-weighted GRS followed by hypertensive heart and/or renal disease ($P = 3 \times 10^{-173}$) and hypertensive CKD ($P = 5 \times 10^{-141}$). In addition to disease groupings related to heart and vascular conditions, groupings related to the kidney including renal failure ($P = 3 \times 10^{-36}$), gout ($P = 2 \times 10^{-35}$), CKD ($P = 1 \times 10^{-26}$), and acute renal failure ($P = 1 \times 10^{-19}$) were among the top associations with BP-weighted GRSs. The eGFR-weighted GRS was inversely associated with kidney conditions including CKD ($P = 2 \times 10^{-100}$), renal failure ($P = 2 \times 10^{-94}$), CKD Stage III ($P = 1 \times 10^{-54}$), gout ($P = 3 \times 10^{-7}$), hypertensive CKD ($P = 4 \times 10^{-28}$), and hypertensive heart and/or renal disease ($P = 2 \times 10^{-19}$). The eGFR-weighted GRS was also positively associated with calculi formation including kidney calculus ($P = 3 \times 10^{-15}$), urinary calculus ($P = 4 \times 10^{-15}$) and ureter calculus ($P = 1 \times 10^{-5}$). However, eGFR-weighted GRS was not associated with essential HT ($P = 0.08$) or HT ($P = 0.09$). Additional analyses, adjusting for observed BP, eGFR, adjustment for GRSs, as well as Mendelian Randomization studies are underway. In this large GRS PheWAS study, genetically predicted BP-traits are associated with CKD, but genetically predicted kidney function is not associated with primary HT.

3117W

Rare variants in the 22q11.2 deletion region may contribute to the development of conotruncal heart defects in non-deleted patients. D.A. Hammond^{1,2}, D.M. Taylor^{1,2}, B. Morrow³, W.K. Chung⁴, Y. Shen⁴, J.W. Newburger⁵, C.E. Seidman⁶, M. Brueckner⁶, G.A. Porter, Jr.⁷, R.W. Kim⁸, D. Srivastava⁹, M. Tristani-Firouzi¹⁰, B.D. Gelb¹¹, A.J. Agopian¹², L.E. Mitchell¹², E. Goldmuntz^{1,2}. 1) Children's Hospital of Philadelphia, Philadelphia, PA; 2) University of Pennsylvania, Philadelphia, PA; 3) Albert Einstein College of Medicine, Bronx, NY; 4) Columbia University, New York, NY; 5) Children's Hospital Boston, Boston MA; 6) Yale University School of Medicine, New Haven, CT; 7) University of Rochester Medical Center, Rochester, NY; 8) Children's Hospital Los Angeles, Los Angeles, CA; 9) Gladstone Institutes/UCSF, San Francisco, CA; 10) University of Utah School of Medicine, Salt Lake City, UT; 11) Icahn School of Medicine at Mount Sinai, New York, NY; 12) UTHealth School of Public Health, Houston, TX.

Conotruncal defects (CTDs) are an important subtype of congenital heart defects that are associated with high morbidity and early mortality. It is estimated that approximately 80% of individuals with 22q11.2 deletion syndrome have CTDs, suggesting that genes within the chromosome 22q11.2 region may be critical to the development of CTDs. We therefore hypothesized that variation in genes in the region contribute to risk of CTDs in non-deleted individuals. In this study, we assessed the association of rare variants in genes within the chromosome 22q11.2 deletion region in individuals with CTDs who do not have a 22q11.2 deletion. **Methods:** Whole exome sequence (WES) data from 472 parent-child trios from the Pediatric Cardiac Genomic Consortium (PCGC) with a clinical diagnosis of CTDs were tested for association of rare variants (minor allele frequency <0.05) in genes within the 22q11.2 region using the family-based association test (FBAT) to obtain variant level p-values. The variant level p-values were then used to assess gene-level associations and to conduct a competitive gene-set analysis (all genes in the 22q11.2 region) via the Multi-marker Analysis of GenoMic Annotation (MAGMA). **Results:** We found a statistically significant association between CTDs and the set of all genes in the 22q11.2 region (p-value = 8.19×10^{-3}) and with two of the genes within the region: RAN-binding protein 1 (*RANBP1*, p = 0.03) and DiGeorge Syndrome Critical Region Gene 6 (*DGCR6*, p = 3.8×10^{-3}). **Conclusions:** Our findings suggest that rare variants within genes located in the 22q11.2 region contribute to the development of CTDs.

3118T

Spontaneous coronary artery dissection is infrequent in individuals with thoracic aortopathies despite shared genetic susceptibility. H.L. Hill^{1,2}, M. Ghanam¹, M. Yang^{1,2}, N.L. Pugh³, F.M. Asch⁴, W. Hornsby⁵, A. Driscoll⁶, J. McNamara⁷, C.J. Willer^{1,2}, E.S. Regalado⁸, GenTAC Investigators¹, MAC Investigators⁹, D.M. Milewicz⁶, K.A. Eagle¹, S.K. Ganesh^{1,2}. 1) Division of Cardiovascular Medicine, Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan; 2) Department of Human Genetics, University of Michigan, Ann Arbor, Michigan; 3) Center for Clinical Research Network Coordination, Biostatistics and Epidemiology Division, RTI International, Research Triangle Park, NC; 4) MedStar Health Research Institute, Hyattsville, MD; 5) Cardiovascular Health Improvement Project (CHIP) biorepository; 6) Division of Medical Genetics, Department of Internal Medicine, McGovern Medical School, The University of Texas Health Science Center at Houston, Houston, Texas; 7) Genetically Triggered Thoracic Aortic Aneurysms and Cardiovascular Conditions (GenTAC) Consortium; 8) Montalcino Aortic Consortium (MAC).

Spontaneous coronary artery dissection (SCAD) is an underdiagnosed nonatherosclerotic precipitant of acute coronary syndrome. Genetic testing in case series has identified pathogenic mutations in genes underlying familial thoracic aortic aneurysm and dissection (TAAD) and related disorders. We hypothesized that individuals with TAAD are at elevated risk for SCAD and that phenotypic profiling of affected individuals can inform clinical risk stratification. Three clinical databases enrolling individuals with TAAD (n=6,838) were queried for a history of SCAD events. Retrograde aortic and iatrogenic dissections were excluded. Queried datasets included the (1) NIH-supported multicenter Genetically Triggered Thoracic Aortic Aneurysms and Cardiovascular Conditions (GenTAC) consortium national registry; (2) Montalcino Aortic Consortium (MAC) international registry; and (3) Cardiovascular Health Improvement Project (CHIP) biorepository at the University of Michigan Frankel Cardiovascular Center. Phenotypic profiles of individuals experiencing SCAD were compared against unaffected GenTAC controls using tests of independence. Querying 6,838 individuals with genetically-associated TAAD identified 11 cases of SCAD (0.2%). The GenTAC (n=3,542) and MAC (n=987) registries included 8 (0.2%) and 3 (0.3%) cases of SCAD, respectively. No SCAD events were identified in the CHIP biorepository (n=2,309). Genetic testing of aortopathy genes performed as part of routine clinical care confirmed pathogenic mutations in 7 (63.6%) SCAD subjects affecting the *COL3A1* (n=2), *TGFBR2* (n=2), *TGFBR1* (n=1), *FBN1* (n=1), and *PRKG1* (n=1) genes. Affected individuals were predominantly Caucasian (n=9, 81.8%) females (n=9, 81.8%) experiencing SCAD at an average age of 35.6 ± 11.0 years. Dissections affected main or proximal coronary arteries in 5 cases (45.5%). Additional dissections of the internal carotid (n=2), iliac (n=2), vertebral (n=1), and subclavian (n=1) arteries were identified in SCAD cases. Prevalence of SCAD in the setting of heritable thoracic aortopathy is low, consistent with reports of low frequency of pathogenic findings in known arteriopathy genes among individuals primarily presenting with SCAD. While our phenotype analysis was limited by small case counts, the presence of aortic branch vessel dissections may suggest an increased risk for SCAD. Our findings implicate SCAD as part of a systemic vasculopathy and support evidence for genetic susceptibility.

3119F

Genome-wide linkage and association studies of circulating apolipoproteins CIII and E levels in Amerindians. W.-C. Hsueh, A. Nair, C. Koroglu, S. Kobes, W.C. Knowler, L.J. Baier, R.L. Hanson. NIDDK, Phoenix, AZ.

Apolipoprotein CIII (apoCIII) is an important component of major lipoproteins. We previously reported that an nsSNP (rs147210663, A43T) in *APOC3* has a large effect on lipid levels, particularly triglycerides ($\beta \approx 1$ SD unit per copy of the T allele) in Amerindians. Thus, we conducted genome-wide searches for genetic variants affecting circulating apoCIII and apoE, which correlates strongly with apoCIII ($r=0.64$). Both population-based linkage and association approaches were conducted in 1,910 nondiabetic Pima Indians. A custom SNP array (~ 0.5 m SNPs) was designed for Pima Indians. Empirical estimates of IBD among all pairs of subjects for linkage analysis were obtained with Beagle. The genome-wide significance threshold for linkage is $\text{LOD}=3.7$, and for association is $p < 1.2 \times 10^{-7}$ after correction for multiple testing. After adjusting for effects of sex, age and genetic ancestry, we identified 1 region with strong evidence for linkage for apoCIII, covering the *APOC3* coding region (on chr11q, $\text{LOD}=12.5$, $p=1.4 \times 10^{-14}$, explaining 7.4% variance, or σ^2). In GWAS, only 1 region on 11q with strong evidence for association was observed. The *APOC3* A43T SNP had the strongest association ($\text{MAF}=2.4\%$, $p=3.5 \times 10^{-23}$, $\beta=1.06 \pm 0.11$ SD unit, $\% \sigma^2$ explained=5.6%). For apoE, while only 1 strong linkage signal was observed (in the *APOE* region on 19q, $\text{LOD}=6.0$, $p=7.7 \times 10^{-8}$, explaining 8.1% σ^2), 3 regions with suggestive association were observed for apoE (top SNPs: on 11q, A43T in *APOC3*; on 15q, rs17746210 near *OCA2*; and on 19q, rs35136575 near *APOE*; $p=3.0 - 7.7 \times 10^{-7}$). Mediation analysis suggested that the effect of A43T SNP on apoE was entirely mediated through apoCIII (mediation=100%, $p=2.5 \times 10^{-24}$). In both coding regions, there appear to be more than 1 SNP contributing to association/linkage signals. After controlling for rs147210663, a SNP in *SIK3* (rs35190172) also showed significant association ($p=1.0 \times 10^{-6}$) with apoCIII. For apoE, rs71352239 and rs7259004 on 19q remained significantly associated with apoE ($p=3.9 \times 10^{-7}$) after controlling for SNPs with stronger associations. Our findings identify evidence for linkage and association for these 2 apolipoproteins in their coding regions, and suggest that multiple variants in the same genomic region may contribute to the genetic signals. Our findings also suggest that some genetic influences on apoE may be mediated via apoCIII. Fine-mapping studies based on sequencing data are ongoing with the aim at identifying functional variants that may explain these signals.

3120W

Placental growth factor locus is associated with decreased risk of coronary artery disease and incident major adverse cardiac events. P.

Huang^{1,2}, Y. Han^{1,2}, Q. Jia^{1,2}, N.C. Woodward^{1,2}, J. Gukasyan^{1,2}, W.H.W. Tang^{3,4}, S.L. Hazen^{3,4}, H. Allayee^{1,2}, J.A. Hartman^{1,2}. 1) Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA; 2) Department of Biochemistry & Molecular Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA; 3) Department of Cardiovascular Medicine, Cleveland Clinic, Cleveland, OH; 4) Department of Cellular & Molecular Medicine, Cleveland Clinic, Cleveland, OH.

Recent studies show that PIGF levels are elevated in coronary artery disease (CAD) patients and are associated with adverse cardiac events. Although placental growth factor (PIGF) is an important cytokine involved in angiogenesis, cardiac remodeling and inflammation, it is not known whether the association between PIGF levels and CAD outcomes is causal. To address this question, we carried out a 2-stage genome-wide association study (GWAS) to identify genetic determinants controlling PIGF levels and then evaluated association of the variants with CAD. In stage 1, we performed a GWAS with ~ 2.5 million single nucleotide polymorphisms (SNPs) in ~ 2000 subjects from the GeneBank cohort. Suggestive associations near *PGF* (rs11850328, $p=1.2 \times 10^{-7}$) on chr14q24.3 and in intergenic region on chr2p16.1 (rs6545625, $p=7.5 \times 10^{-7}$) from stage 1 were genotyped in an independent sample of ~ 2000 additional GeneBank subjects, of which only rs11850328 (G/A) showed consistent evidence for association in stage 2. A combined analysis with ~ 4000 subjects yielded a genome-wide significance signal with the *PGF* locus and serum PIGF levels ($p=2.8 \times 10^{-6}$). Interestingly, the minor allele A was also associated with decreased risk of incident major cardiac events (MACE) in $\sim 4,000$ GeneBank subjects ($\text{HR}=0.87$; 95% CI 0.76-0.99; $p=0.04$). Since PIGF levels had also been previously clinically associated with atherosclerotic plaque destabilization and cardiac repair, we tested whether history of myocardial infarction (HxMI) modified the association between the *PGF* variant and MACE. Notably, rs11850328 yielded a significant interaction with HxMI ($p=0.009$), prompting us to perform a stratified analysis. The A allele did not influence risk of MACE among those with HxMI ($p=0.70$) but decreased risk of future events 2-fold in subjects with no history of CAD/MI ($\text{HR} 0.50$, 95% CI 0.30-0.81, $p=0.005$). Consistent with these observations, the A allele also significantly decreased risk of CAD ($\text{OR}=0.98$, 95% CI 0.97-0.99; $p=2.9 \times 10^{-6}$) in an independent dataset of $>500,000$ Caucasian subjects. Our data indicate that genetic variants increasing PIGF levels show decreased risk of CAD, but these observations were somewhat unexpected since higher levels of PIGF are clinically associated with increased risk of CAD. Thus, additional studies will be needed to determine the biological mechanism for the opposite associations we observe between genetic determinants of serum PIGF and risk of CAD.

3121T

A multiplex pedigree illustrates the challenges of applying a Mendelian monogenic paradigm to the interpretation of rare variants in familial dilated cardiomyopathy. E. Jordan, A. Morales, D. Kinnamon, K. Zareba, J. Cowan, L. Salyer, S. Raman, R. Hersberger. The Ohio State University, Columbus, OH.

Familial dilated cardiomyopathy (DCM) has a genetic basis. Identifying genetic risk and detecting early clinical evidence of DCM in at-risk relatives holds promise to mitigate, and ultimately prevent, development of disease. We report a family in which a proband and 5 affected relatives harbor a rare *RBM20* variant in exon 9 encoding a premature stop of translation, NM_001134363.1:c.2062C>T (p.Arg688X). *RBM20*, ribonucleic acid (RNA)-binding protein 20, has roles in RNA binding and regulation of mRNA splicing of key genes involved in cardiac development. To date, the only proven pathogenic *RBM20* variants are missense variants present in the exon 9 hotspot (encompassing amino acid positions 634, 636, 637, 638). After expanding the pedigree to include phenotype and genotype data from 15 individuals, two-point parametric linkage analysis performed under an autosomal dominant disease model provided moderate evidence that the *RBM20* variant segregates with affection with left ventricular enlargement, systolic dysfunction, or DCM (LOD = 1.69 at $\theta = 0.001$). Moreover, cardiac magnetic resonance imaging (CMR) in 12 relatives (7 with and 5 without the *RBM20* variant) also showed greater left ventricular end diastolic dimensions (5.9 ± 1.0 vs. 5.2 ± 0.4 cm) and volumes (121 ± 45 vs. 87 ± 9 ml/m²), lower ejection fractions ($47 \pm 14\%$ vs. $58 \pm 4\%$), a larger extent of midwall fibrosis, and higher native T1 and T2 values in those with the variant. While these findings hint at biological relevance, this is a variant of uncertain significance (VUS) based on our variant classification standards adapted from the published *MYH7* ClinGen Cardiomyopathy Expert Panel approach. This classification is primarily due to the lack of evidence of loss of function as a pathogenic mechanism in *RBM20* and limited case data. In addition, CMR abnormalities suggesting early DCM were observed in 4 of 5 individuals without the *RBM20* variant, suggesting that additional unidentified genetic variation may contribute to DCM in this family. This observation is consistent with our preliminary data showing that 19% of DCM probands carry multiple variants, classified as VUS, likely pathogenic, or pathogenic, in established DCM genes. Our observations suggest that the genetic background of DCM may be more complex than previously appreciated. Further elucidation of the disease model will be necessary to enable the clinical application of genetic information in the prevention of this disease.

3122F

The transcriptomic signature and regulatory landscape of iPSC-derived megakaryocytes and platelets. K. Kammers¹, M.A. Taub², B. Rodriguez³, I. Rucinski⁴, L.R. Yanek⁴, A.D. Johnson³, N. Faraday⁴, L.C. Becker⁴, R.A. Mathias⁴. 1) Division of Biostatistics and Bioinformatics, Department of Oncology, Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; 3) NHLBI Population Sciences Branch, The Framingham Heart Study, Framingham, MA; 4) The GeneSTAR Research Program, Johns Hopkins University School of Medicine, Baltimore, MD.

GWAS studies have identified platelet aggregation loci, but because these are largely intronic/intergenic, their mechanistic link to platelet function is unclear. To address these gaps, we performed expression-quantitative trait locus (eQTL) analyses of RNA sequencing data on platelets and iPSC-derived megakaryocytes (MKs), to define the transcriptome signature and its genetic regulatory landscape in both cell types in African American (AA) and European American (EA) subjects. Using 1,003,451 SNPs from the Illumina 1M GWAS array, analyses were performed stratified by ancestry and cell type, with a 1Mb window around each gene and adjusting for relevant covariates with MatrixEQTL. Significance was defined as a q-value < 0.05. Genes with median FPKM ≤ 1 were excluded, yielding ~10,000 genes in the MKs and ~3,000 in the platelets, 94% of which are also expressed in the MKs. In MKs (N=107 EA, N=87 AA), we found 8,629 and 1,280 eQTLs in the two groups, respectively. For platelets (N=185 EA, N=112 AA), we identified 23,355 eQTLs in EAs and 5,740 in AAs. We found high tissue specificity compared to 44 tissues in Genotype-Tissue Expression project (GTEx): 34% of MK and 55% of platelet eQTLs are not reported in GTEx. The identified eQTLs reveal significant enrichment ($p < 10^{-100}$) in super-enhancers previously reported to be specific to MKs with similar patterns of enrichment for both sets of eQTLs, but stronger enrichment in platelets. In general across both cell types and for both ancestries, the highest enrichment was for promoter proximal H3K27ac and H3K4me3 bound promoter regions reported in MKs. Additionally, we found that the eQTLs that map to these known super enhancer regions are significantly enriched for several key pathways in platelet physiology: platelet degranulation, platelet activation/aggregation, and hemostasis. These results offer promising insights into the regulatory signature of platelets and MKs, with significant overlap in genes expressed, eQTLs detected, and enrichment within known super enhancers relevant to platelet biology, which in turn could shed light on these clinically and pharmacologically important tissues. We are currently extending these analyses to leverage whole genome sequence data generated through the NHLBI TOPMed Program to overcome limitations of poor coverage of the GWAS array in the AAs, and to better refine the regulatory signature without relying on the tagging approach of the GWAS.

3123W

Associations of genetically-inferred ancestry with hypertension traits.

J.M. Keaton^{1,2,3,4}, J.N. Hellwege^{1,2,3,4}, A. Giri^{2,3,4,5}, E.S. Torstenson^{1,4}, C.P. Kovesdy⁶, Y.V. Sun⁷, P.W.F. Wilson⁸, C.J. O'Donnell⁹, T.L. Edwards^{1,2,3,4}, A.M. Hung^{4,10}, D.R. Velez Edwards^{2,3,4,5} on behalf of the Million Veteran Program. 1) Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA; 2) Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN, USA; 3) Institute for Medicine and Public Health, Vanderbilt University Medical Center, Nashville, TN, USA; 4) Biomedical Laboratory Research and Development, Tennessee Valley Healthcare System (626)/Vanderbilt University, Nashville, TN, USA; 5) Department of Obstetrics and Gynecology, Vanderbilt University Medical Center, Nashville, TN, USA; 6) Nephrology Section, Memphis VA Medical Center, Memphis TN, USA; 7) Department of Epidemiology, Emory University Rollins School of Public Health, Department of Biomedical Informatics, Emory University School of Medicine, Atlanta, GA, USA; 8) Atlanta VAMC and Emory Clinical Cardiovascular Research Institute, Atlanta, GA, USA; 9) VA Boston Healthcare, Section of Cardiology and Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; 10) Division of Nephrology and Hypertension, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA.

Hypertension is a risk factor for cardiovascular disease, has considerable impact on global healthcare, and underlying blood pressure (BP) traits are substantially heritable. Ethnic disparities in hypertension prevalence are well documented, though the influence of genetic factors is unclear. Our goal was to find associations of BP traits with genetic ancestry to determine whether ancestral geographic origin could explain racial disparities. We analyzed genetically inferred ancestry proportions from five 1000 Genomes reference populations [British in England and Scotland (GBR), Peruvians from Lima, Peru (PEL), Yoruba in Ibadan, Nigeria (YRI), Han Chinese in Beijing, China (CHB), and Luhya in Webuye, Kenya (LWK)] with four continuous BP traits (systolic BP [SBP], diastolic BP [DBP], pulse pressure [PP], and mean arterial pressure [MAP]), and the binary outcomes hypertension and drug-resistant hypertension in 59,927 black, 220,495 white, and 21,273 Hispanic individuals from the Million Veteran Program using electronic health records linked to genotypes. Continuous traits and binary outcomes were modeled using linear and logistic regression, respectively, for each ancestry proportion adjusted for age, age², sex, and body mass index. Results were meta-analyzed to report effect estimates per 10% difference for a given inferred ancestry proportion in all samples. Percent GBR was negatively associated with BP ($P=2.13 \times 10^{-19}$, 7.92×10^{-8} , 4.41×10^{-11} , and 3.57×10^{-13} for SBP, DBP, PP, and MAP respectively; coefficient range -0.10 to -0.21), and was protective against hypertension ($P=2.59 \times 10^{-5}$, odds ratio [OR]=0.98) relative to other ancestries. YRI percentage was positively associated with BP ($P=1.63 \times 10^{-23}$, 1.94×10^{-26} , 0.012, and 3.26×10^{-29} for SBP, DBP, PP, and MAP respectively; coefficient range 0.06–0.32 mm Hg) and was positively associated with hypertension risk ($P=3.10 \times 10^{-11}$, OR=1.04) and drug-resistant hypertension risk ($P=1.86 \times 10^{-4}$, OR=1.04) compared with other ancestries. Percent PEL was positively associated with SBP ($P=6.68 \times 10^{-3}$, beta=0.09 mm Hg) and PP ($P=3.69 \times 10^{-15}$, beta=0.21 mm Hg), but inversely associated with DBP ($P=2.84 \times 10^{-5}$, beta=-0.11 mm Hg). These results demonstrate that genetic ancestry is a significant risk factor for BP traits, provide insight into the geographic origin of genetic factors underlying hypertension risk, and establish that a portion of BP trait racial disparities are due to genetic differences between continental groups.

3124T

Whole genome sequencing analysis of blood pressure phenotypes in the NHLBI Trans-Omics for Precision Medicine Program. T. Kelly¹, X. Sun¹, J. Brody², TOPMed Blood Pressure Working Group.

1) Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA; 2) Cardiovascular Health Research Unit, University of Washington, Seattle, WA.

Although genome-wide association studies (GWAS) have made important strides in localizing genomic regions associated with blood pressure (BP) phenotypes, the causal mechanisms underlying the vast majority of identified signals remain unknown. Whole genome sequencing (WGS) provides an opportunity for not only novel genomic discoveries but also high resolution refinement of previously identified GWAS signals. We pooled high-depth WGS and phenotype data from 32,263 participants from 12 cohorts from phases 1 and 2 of the Trans-Omics for Precision Medicine (TOPMed) program. These data were used to conduct WGS association analyses of systolic BP (SBP) and diastolic BP (DBP). Genotype quality filtering included the removal of variants lying in low complexity regions, along with those with a minor allele count (MAC)<10, Hardy Weinberg p-values< 1×10^{-9} , and missingness>5%. For participants using antihypertension medications, 10 mmHg and 5 mmHg were added to SBP and DBP, respectively. BP values were then regressed on important covariables including age, sex, body mass index (BMI), study, and the first 10 principal components within each ancestry group. The ancestry-specific, residualized BP values were inverse-normal transformed, rescaled by variance, and pooled for the analysis. These pooled residuals were used as outcomes in a linear mixed model with the same adjustments, but also including a variance component associated with the genetic relationship matrix plus separate residual variance components for each ancestry group. In an effort to minimize cost and computational time, the full mixed model was only used for variants with $P < 1 \times 10^{-4}$ in an initial scan that used a linear model. All analyses were supported by the Analysis Commons. Among 44 novel loci achieving $P < 1 \times 10^{-5}$, two rare variants achieved genome-wide significance, including one at *EIF3LP1* for SBP (rs191347507; MAC=11; $P=4.78 \times 10^{-9}$) and one at *RNGTT* for DBP (rs578046022; MAC=91; $P=2.73 \times 10^{-9}$). Twenty previously reported loci also achieved $P < 1 \times 10^{-5}$, including three attaining genome-wide significance, *PRDM8/FGF5* for both SBP and DBP (rs16998073; MAC=15,888; $P=7.22 \times 10^{-9}$ and 1.80×10^{-9} , respectively), *SOX6* for SBP (rs34610248; MAC=10,552; $P=3.62 \times 10^{-9}$) and a novel variant at *MRPL4/ICAM1* for SBP (rs563785784; MAC=33; $P=4.95 \times 10^{-9}$). These discovery analyses provide empirical evidence that WGS studies across diverse ancestry groups may yield new genomic signals for BP.

3125F

Genetic association study of venous thromboembolism based on whole-genome sequencing in the trans-omics for precision medicine (TOPMed) program. C.A. Laurie¹, B. Coombes², A.A. Seyerle³, D. Jain¹, W. Tang³, M.H. Chen^{4,5}, M. Puurunen^{4,6,7}, D. Ko⁶, A.D. Johnson^{4,5}, A. Morrison⁸, M. Cushman⁹, S.R. Heckbert¹⁰, M. deAndrade², E. Boerwinkle⁴, R.S. Vasan⁶, A.P. Reiner^{11,12}, B. McKnight¹, K. Rice¹, T. Sofer¹, C.C. Laurie¹, N. Pankratz¹³, J. Pankow⁴, C. Kooperberg¹², N.L. Smith^{11,14,15}. 1) Department of Biostatistics, University of Washington, Seattle, Washington, WA; 2) Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, Minnesota; 3) Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, Minnesota; 4) National Heart Lung and Blood Institute's and Boston University's The Framingham Heart Study, Framingham MA; 5) Population Sciences Branch, Division of Intramural Research, National Heart, Lung and Blood Institute, Bethesda, MD; 6) Boston University School of Medicine, Boston, MA; 7) Boston University School of Public Health, Boston University, Boston, MA; 8) Human Genetics Center, Department of Epidemiology, Human Genetics, and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, Texas, USA; 9) Department of Medicine, Larner College of Medicine at the University of Vermont, Burlington, VT; 10) Cardiovascular Health Research Unit and Department of Epidemiology, School of Public Health, University of Washington, Seattle, Washington; 11) Department of Epidemiology, University of Washington, Seattle, Washington; 12) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington; 13) Division of Molecular Biology and Genomics, University of Minnesota, Minneapolis, Minnesota; 14) Group Health Research Institute, Group Health Cooperative, Seattle, Washington; 15) Seattle Epidemiologic Research and Information Center, VA Office of Research and Development, Seattle, Washington.

Venous thromboembolism (VTE) is a heritable, life-threatening condition with known genetic predictors. The most recent agnostic search for genetic association focused on common variation (PMID 25772935). Here we studied rare variation using whole-genome sequence (WGS). In TOPMed, we performed a pooled case-control analysis using data from 3 cohort studies (nested case-control) and 3 case-only collections. Case-control matching was performed within sex-ancestry-study groups, based on age at diagnosis for cases and age at last follow-up for controls. Excess controls from one of the cohort studies were matched with cases from the case-only collections. The matched sample set consisted of 8,911 participants with an average ratio of 1.9 controls per case (5,816 controls and 3,095 cases) with 88% European-American (EA), 10% African-American (AA) and 2% other ancestries. Mixed-model logistic regression was used for single-variant and aggregate tests. Covariates were sex, sample set, age, and $\ln(\text{case/control ratio per matched set})$. Single-variant tests (minor allele count filter > 5) identified 6 loci known to be associated with VTE (PMID 25772935, PMID 20212171) and one novel locus. Conditioning on the known associations did not identify any new associations in previously known loci and the new locus remained significant. Significant variant rs10963226 ($p = 3 \times 10^{-8}$) at the novel locus had a MAF of 0.0001 in EA and 0.04 in AA participants. The high quality variant rs10963226 lies in the first intron of one of the *SH3GL2* transcripts. Chromatin capture data indicates that the region overlapping rs10963226 interacts with the promoter of *SH3GL2* suggesting rs10963226 may have a regulatory role. Furthermore, *SH3GL2* has been reported to interact with coagulation factor XIII A subunit [PMID: 21900206]. These observations suggest that the rare variant association in the *SH3GL2* locus is biologically plausible. Burden and SKAT tests were performed using variants with MAF < 0.01, aggregating by sliding-window (50 kb and 5kb) and also by genes after applying various genome annotation filters. Only the analysis that aggregated loss of function and missense variants in each gene had a significant result (< Bonferroni p -value 2×10^{-6}). The unit identified in this analysis was the gene *PROC*, a locus well known to be associated with VTE based on family studies. In summary, the WGS data confirmed previously known loci containing common and rare variants and identified a new locus.

3126W

The CTLA4 gene is associated with Kawasaki disease and coronary artery lesions: Case-control and family-based studies. Y. Lee^{1,2,4,5}, M. Chen¹, N. Chiu¹, H. Chi¹, F. Huang¹, K. Yang¹, L. Chang¹, H. Lee¹, S. Lin¹, W. Chen², C. Lin², H. Chan², W. Lin², S. Chnag², T. Chnag², H. Yang², M. Lin², L. Chang². 1) MacKay Children's Hospital, Taipei, Taiwan; 2) MacKay Memorial Hospital, Taipei, Taiwan; 3) National Taiwan University Hospital, Taipei, Taiwan; 4) Macky Medical College, New Taipei, Taiwan; 5) Taipei Medical University, Taipei, Taiwan.

Background: Kawasaki disease (KD) is a febrile disease of children complicated with coronary artery lesions (CALs). The cytotoxic T-lymphocyte-associated protein 4 (CTLA4) increases during the acute stage of KD. **Methods:** 720 unrelated KD children and 1058 controls were genotyped for SNP rs5742909:C>T, rs231775:A>G, and rs3087243:G>A of the CTLA4 gene. Hardy-Weinberg equilibrium test, linkage disequilibrium, and haplotype frequency were assessed by Haploview 4.2. The genotype, allele, carrier, and haplotype distributions of patients and controls were compared using χ^2 test. Family-based data were analysed using Transmission-disequilibrium test. **Results:** Genotype rs3087243A/G was significantly less frequent in KD patients than in controls, OR (95%CI) = 0.76 (0.61-0.96), $P_c = 0.046$. So was haplotype CAG, OR = 0.17 (0.06-0.48), $P_c = 0.0096$. Allele rs231775A was significantly less frequent in patients with CALs than in those without CALs, OR = 0.72 (0.54-0.95), $P_c = 0.04$ while allele rs231775G was significantly more frequent, OR = 1.39 (1.05-1.84), $P_c = 0.04$. Genotype rs3087243G/G was significantly more frequent in patients with CALs than in controls, OR = 1.55 (1.10-2.18), $P_c = 0.021$. Carrier A was significantly less frequent in patients with CALs than in controls, OR = 0.64 (0.46-0.91), $P_c = 0.021$. So was genotype A/G, OR = 0.64 (0.45-0.91), $P_c = 0.026$. Haplotype CGG was significantly more frequent in CAL patients than in controls, OR (95%CI) = 1.37 (1.07 - 1.76), $P_c = 0.048$. Haplotype CAG was not found in CAL patients, OR (95%CI) = 0.06 (0.00 - 0.94), $P_c = 0.026$. TDT confirmed the association. **Conclusions:** The CTLA4 gene is associated with KD and CALs.

3127T

Genetic liability for clinical risk factors differentially associates with heart failure subtypes. R.T. Levinson, E. Farber-Eger, B.D. Lowery, J.D. Mosley, Q.S. Wells. Vanderbilt University Medical Center, Nashville, TN.

Background: Heart failure (HF) is a heterogeneous syndrome associated with significant morbidity and mortality, and is commonly sub-grouped into HF with preserved ($\geq 50\%$) and reduced ($\leq 40\%$) ejection fraction (HFpEF and HFrEF). Genome-wide association studies (GWAS) of all-HF have identified few risk loci. No GWAS of HF subtypes have been reported and whether genetic risk for HF subtypes is shared is unknown. Patients with HFpEF and HFrEF have similar, yet distinct clinical profiles, and the relative contribution of established risk factors to HF subtypes is incompletely defined. We proposed to interrogate these relationships by assessing the association of HF subtypes with polygenic liability for clinical risk factors. **Methods:** HF subjects with genotyping data (Illumina MEGA Array) were identified from Vanderbilt University Medical Center's electronic health record (EHR) using a random forest classifier and assigned to a HF subtype by ejection fraction. HF cases with any EF $\leq 40\%$ were considered HFrEF (1,383 cases), and those with no EF $< 50\%$ were considered HFpEF (920 cases). Polygenic risk scores for coronary artery disease (CAD), body mass index (BMI), fasting insulin, high-density lipoprotein (HDL), low-density lipoprotein (LDL), and triglycerides were constructed using published GWAS and tested for association with HFpEF and HFrEF phenotypes. **Results:** Higher genetically predicted BMI was strongly associated with both HF subtypes, though the association was stronger with HFpEF ($p=1.8e-05$ vs $5.6e-04$). Genetic risk for CAD was strongly associated with HFrEF ($p=6.8e-05$) but not with HFpEF ($p=0.2$). There was a trend toward association between higher genetically predicted fasting insulin levels and HFpEF ($p=0.07$), but there was no association with HFrEF ($p=0.6$). Genetically predicted LDL and triglycerides were not associated with either HF subtype, while weak associations were seen between HDL and both HFpEF ($p=0.09$) and HFrEF ($p=0.1$). **Conclusions:** HF subtypes show different patterns of association with genetic risk for clinical risk factors. Genetic risk for CAD is strongly associated HFrEF, while HFpEF is most strongly associated with obesity and metabolic dysregulation. These data suggest the genetic architectures of HF subtype risk are distinct.

3128F

The roles of *SMYD4* in epigenetic regulation of cardiac development in zebrafish. D. Ma^{1,2}, D.Y. Xiao¹, H.Y. Wang², L.L. Hao¹, X. Guo², X.J. Ma², Y.Y. Qian², J. Ma^{1,2}, J. Zhang¹, W. Sheng², W.N. Shou³, G.Y. Huang². 1) Fudan University, Shanghai, China; 2) Children's Hospital of Fudan University, Shanghai, China; 3) Indiana University School of Medicine, Indianapolis, IN, USA.

SMYD4 belongs to a family of lysine methyltransferases. We analyzed the role of *smyd4* in zebrafish development by generating a *smyd4* mutant zebrafish line (*smyd4*^{L544Efs*1}) using the CRISPR/Cas9 technology. The maternal and zygotic *smyd4*^{L544Efs*1} mutants demonstrated severe cardiac malformations, including defects in left-right patterning and looping and hypoplastic ventricles, suggesting that *smyd4* was critical for heart development. Importantly, we identified two rare *SMYD4* genetic variants in a 208-patient cohort with congenital heart defects. Both biochemical and functional analyses indicated that *SMYD4*(G345D) was pathogenic. Our data suggested that *smyd4* functions as a histone methyltransferase and, by interacting with HDAC1, also serves as a potential modulator for histone acetylation. Transcriptome and bioinformatics analyses of *smyd4*^{L544Efs*1} and wild-type developing hearts suggested that *smyd4* is a key epigenetic regulator involved in regulating endoplasmic reticulum-mediated protein processing and several important metabolic pathways in developing zebrafish hearts.

3129W

Trans-ethnic meta-analysis of genome-wide association studies for coronary artery disease reveals genetic differences between Japanese and Europeans. H. Matsunaga^{1,2}, K. Ito², M. Akiyama³, S. Koyama², Y. Onouchi², K. Ozaki², H. Morita¹, M. Kubo³, H. Akazawa¹, Y. Kamatani³, I. Komuro¹. 1) The University of Tokyo, Department of Cardiovascular Medicine, Tokyo; 2) RIKEN Center for Integrative Medical Sciences, Laboratory for Cardiovascular Diseases, Kanagawa; 3) RIKEN Center for Integrative Medical Sciences, Laboratory for Statistical Analysis, Kanagawa.

Coronary artery disease (CAD) is a leading cause of death worldwide. Although genome-wide association studies (GWAS) have revealed more than 150 loci associated with CAD and have gradually elucidated complex traits, the causal genes and pathways are not fully understood. In addition, it has been reported that the prevalence of CAD was lower in Japanese than in Europeans, but the reasons are unclear. To evaluate the contribution of genetic factors for CAD and understand the genetic differences between Japanese and Europeans, we carried out Japanese GWAS, followed by trans-ethnic meta-analysis for CAD. First, we conducted a meta-analysis of two GWAS that included 12,494 cases and 28,879 controls, and 2,808 cases and 7,261 controls, respectively, in the Japanese population. Next, to further identify related loci for CAD, we performed trans-ethnic meta-analysis using the result of UK Biobank and CARDIoGRAMplusC4D. To prioritize candidate causal genes at significant loci, we searched quantitative trait loci (QTL) through GTEx, BIOS and BioBank Japan clinical data. In the Japanese GWAS, we found 17 loci significantly associated with CAD. Of which just 1 was new. In the trans-ethnic meta-analysis, we discovered 76 CAD risk loci, of which 4 were novel. Since the new locus in the Japanese meta-GWAS and one of the new loci in the trans-ethnic meta-analysis were located within the same region, we finally identified 4 new loci: on 1q21, 6p21, 10q26, 11q22. We found that these new loci were related to immune system, diabetes mellitus and dyslipidemia by QTL analyses. Moreover, we compared effect sizes for CAD at the 76 loci, which exceeded the genome-wide significance threshold in the trans-ethnic meta-analysis, to assess the genetic differences between Japanese and Europeans. Among the 76 loci, 60 loci had a larger effect in the first Japanese GWAS than in the European GWAS (78.9%, P for sign test = 3.85×10^{-4}). To verify this result, we also compared the second Japanese GWAS and the European GWAS. Of the 76 loci, 54 loci had a larger effect in Japanese than in Europeans (71.1%, P for sign test = 3.13×10^{-4}). Then, the heritability explained by the 76 loci in the Japanese population was larger than that of the European population. Accordingly, we suggested that Japanese were genetically liable to CAD compared to Europeans. Our results showed that there were genetic differences between Japanese and Europeans, which provided new insights into the pathogenesis of CAD.

3130T

A genetic predictor of thyroid stimulating hormone levels is inversely associated with atrial fibrillation risk. J.D. Mosley^{1,2}, Q.S. Wells¹, M.B. Shoemaker¹, C.M. Shaffer¹, P. Straub¹, L. Bastarache², G.P. Jarvik³, E.B. Larson^{3,4}, D.R. Velez Edwards⁵, T.L. Edwards¹, L.K. Davis¹, M.G. Hayes⁶, P.T. Ellinor⁷, D.M. Roden^{1,2,8}, J-E. Salem^{1,9}. 1) Vanderbilt University Medical Center, Nashville, TN; 2) Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN; 3) Departments of Medicine and Genome Sciences, University of Washington, Seattle, WA; 4) Kaiser Permanente Washington Health Research Institute, Seattle, WA; 5) Department of Obstetrics and Gynecology and Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN; 6) Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL; 7) Cardiovascular Research Center, Massachusetts General Hospital, Charlestown, MA; 8) Department of Pharmacology, Vanderbilt University Medical Center, Nashville, TN; 9) Department of Pharmacology, Sorbonne Universite, Paris, France.

Background: Thyroid hormones (thyroxine [T4] and triiodothyronine [T3]) are released from the thyroid gland and regulate metabolic activity in multiple tissues. Thyroid hormone levels are regulated by a feedback mechanism governed by thyroid stimulating hormone (TSH), which is inversely associated with T4 levels, and is measured clinically to ascertain thyroid function. Excess thyroid hormone (thyrotoxicosis) is a well-recognized cause of the cardiac arrhythmia atrial fibrillation (AF). We hypothesized that we could identify diseases associated with thyroid hormone regulation by performing phenome-wide scanning in an electronic health record data set using a genetic predictor of TSH levels. **Methods:** We used weightings for 24 SNPs associated with plasma TSH levels at genome-wide significance in a European ancestry (EA) cohort to compute genetically predicted TSH levels in an EHR cohort comprising 61,667 EA individuals with 1,216 measured clinical phenotypes with over 150 cases. The performance of the predictor was ascertained in 6,797 and 3,019 individuals with clinically measured TSH and free T4 (FT4) levels, respectively. Phenome-wide association testing was used to identify clinical phenotypes associated with genetically predicted TSH levels. The observed association with AF described below was validated in an independent GWAS of 17,931 AF cases and 115,142 controls using an inverse variance weighted average method in conjunction with summary statistics data. **Results:** Genetically predicted TSH levels accounted for 5.8% and 0.5% of the measured phenotypic variance of TSH and FT4 levels, respectively. There were 17 phenotypes associated with predicted TSH levels at FDR $q < 0.1$. Fifteen of these were thyroid diseases and had directionally expected associations. For instance, higher predicted TSH was associated with hypothyroidism (OR = 1.18 [95% CI 1.15-1.21], $p = 7.1 \times 10^{-38}$) per s.d. change in genetically predicted levels) and lower TSH was associated with thyrotoxicosis (OR = 0.82 [0.77-0.86], $p = 7.1 \times 10^{-12}$). Higher TSH was also associated with decreased AF risk (OR = 0.95 [0.93-0.98], $p = 1.3 \times 10^{-3}$). This AF association persisted after excluding individuals with thyroid disease (OR = 0.92 [0.89-0.95], $p = 3.4 \times 10^{-4}$). In the AF validation cohort, each s.d. increase in TSH levels was associated with a decreased risk of AF (OR = 0.86 [0.79-0.94], $p = 0.001$). **Conclusions:** Genetically predicted TSH levels are inversely associated with atrial fibrillation risk.

3131F

Genome-wide association study of carotid artery disease in the eMERGE Network. M.R. Palmer¹, D.S. Kim², D.R. Crosslin³, I.B. Stanaway², E.A. Rosenthal¹, D. Carrell⁴, I. Kullo⁵, M.S. Williams⁵, E. Larson⁴, G.P. Jarvik¹ on behalf of the eMERGE consortium. 1) Medical Genetics, University of Washington, Seattle, WA; 2) University of Michigan, Ann Arbor, MI; 3) Biomedical Informatics and Medical Education, University of Washington, Seattle, WA; 4) Kaiser Permanente Health Research Institute, Seattle, WA; 5) Mayo Clinic, Rochester, MN; 6) Genomic Medicine Institute, Geisinger Health System, Danville, PA.

Introduction: The prevalence of symptomatic carotid artery disease (CAAD) in the United States is estimated at 2-9%, with 5-9% prevalence in patients over 65 years old, and is a major risk factor for stroke. We aim to discover genetic variants associated with CAAD in the electronic Medical Records and Genomics (eMERGE) Network. **Methods:** We identified adult CAAD cases and controls from electronic health records at eight eMERGE sites using a phenotype algorithm, which included carotid stenosis measurements extracted from image reports, and ICD-9 and CPT codes. We used a novel approach to deploy a single natural language processing (NLP) algorithm across multiple eMERGE sites. Participants were genotyped on various platforms, and all imputed to the Haplotype Reference Consortium data. Logistic regression analyses were performed in PLINK 1.9. We performed pooled and stratified analyses by genetically determined race/ethnicity, and adjusted for age, sex, and ancestry principal components.

Table 1.		
Basic demographics	Case 1,794	Control 17,959
Female	667 (37%)	10,391 (58%)
Male	1,126 (63%)	7,568 (42%)
Median Age (IQR)	73 (66 - 78)	64 (53 - 75)
African American	61 (3%)	4,056 (23%)
Caucasian	1,665 (93%)	12,214 (68%)
Unknown/Other	68 (4.2%)	1689 (9.5%)

Results: Our phenotype algorithm identified 1,794 CAAD cases and 17,959 controls (Table 1). In the pooled analysis, two loci reached genome-wide significance, at chr. 6 in *LPA* (rs10455872-G, OR(CI)=1.50 (1.30-1.73), $p=2.1 \times 10^{-8}$, MAF=0.053) and at an intergenic SNV on chr. 7 (rs6952610-T, OR(CI)=1.25 (1.16-1.36), $p=4.3 \times 10^{-8}$, MAF=0.43). The chr. 7 association remained significant in the presence of the *LPA* SNV as a covariate. The *LPA* SNV is also associated with coronary heart disease (CHD; 4,199 cases, 11,679 controls) in this study (OR(CI)=1.27(1.13-1.43), $p=5 \times 10^{-8}$) but the chr7 SNV is not (OR(CI)=1.03(0.97-1.09), $p=0.37$). *LPA* variants associated with elevated levels of its product, lipoprotein(a) (Lp(a)), have previously been associated with CAAD and CHD risk, including rs10455872. **Conclusions:** We replicated the known association of Lp(a) level associated *LPA* SNVs with CAAD and CHD. We identified a novel locus associated with CAAD on chromosome 7 at 7q11.22, which is robust to adjustment for the associated *LPA* variant and was not associated with CHD.

3132W

Genetic sequencing and mass spectrometry biomarkers link diagnosis and treatment precision for cardiovascular disease. J.E. Posey¹, E. Venner², T. Jonsson³, S. Elsea¹, D. Muzny², R.A. Gibbs^{1,2}, S. Scherer^{1,2}, L. Guo³, C.T. Caskey¹. 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA; 3) Metabolon, Morrisville, North Carolina, USA.

Cardiovascular disease is the leading cause of death in the USA. We have used whole genome sequencing (WGS) and mass spectrometry analyte quantitation to improve diagnosis of disease risk and monitor therapy intervention. Our study serves as a model for additional precision medicine disease interventions. Plasma metabolite analysis allows a global assessment of cholesterol pathway function and therapeutic response. Consistent with the known mechanism of HMG CoA reductase inhibition, 21 individuals on statin therapy had a lower average cholesterol:3-hydroxy-3-methylglutarate (CHO:HMG) ratio (15.033) compared to 94 individuals not on statin therapy (26.479, two tailed $t < 0.0001$). Analysis of WGS data did not reveal any variants predicted to impact protein structure or function. However, in analysis of variants predicted to affect transcription of cholesterol pathway genes, we identified 26 variants in this cohort, including 6 in cases being treated for HLD, and 8 in cases predicted to be at increased risk of HLD based on CHO:HMG ratio. No potential molecular etiology was identified in individuals predicted to be protected from HLD based on CHO:HMG ratio. Interpretation of non-coding variants in WGS data remains challenging; we propose that at present, metabolic biomarkers can provide a more accurate assessment of the cholesterol pathway, including *de novo* synthesis and statin response, in a given individual. These observations will need to be demonstrated in a larger cohort. We also analyzed plasma metabolite data for the presence of 19 commonly prescribed or over-the-counter drugs. We found 13 drugs present in 32 subjects and cross-referenced their corresponding genetic variants observed in genes having a PharmGKB "Strength of Evidence" level of 2B or higher for appropriate gene-drug pairs. While we observed preliminary clear examples where alternative SSRIs may be considered in the treatment of depression, data for genetic variants and statins were less clear and further analyses are ongoing. We demonstrate that WGS and metabolome analysis can be a precision medicine approach to diagnosis and treatment of cardiovascular disease. Interpretation of the clinical impact of non-coding variants remains challenging, and plasma mass spectrometry analysis supports a true interrogation of the cholesterol pathway and therapeutic response, and will support future efforts to resolve questions variant pathogenicity for genes within the cholesterol pathway.

3133T

Investigating disease progression through single-cell transcriptional profiling of CD45+ immune cells at the atherosclerotic plaque. A. Sartori¹, K. Thanopoulou², C. Borel¹, E. Falconnet¹, L. Romano³, C. Howald⁴, I. Galani², E. Andreacos², E.T. Dermitzakis^{1,3,4}. 1) University of Geneva, Geneva, Switzerland; 2) Laboratory of Immunobiology, Center for Clinical, Experimental Surgery and Translational Research, Biomedical Research Foundation of the Academy of Athens, Athens, Greece; 3) iGE3 Institute of Genetics and Genomics of Geneva, Switzerland; 4) SIB, Swiss Institute of Bioinformatics, Geneva, Switzerland.

Atherosclerosis is the main cause of cardiovascular diseases, which in turn are the leading cause of mortality worldwide. It is also a chronic and progressive vascular disease of large- and medium-sized arteries characterized by the deposition of fat molecules into their inner wall. This is mostly due to accumulation of low-density lipoproteins (LDL) in the arteria intima which are trapped and turned into oxidized lipoproteins, leading to an innate inflammatory response which results in the formation of an atherosclerotic plaque. In this study, we aim to define the «resident» cell immune diversity at the atherosclerotic plaques level using single cell transcriptomics. In addition, we are investigating transcriptional differences due to atherosclerotic plaque cell composition in symptomatic individuals (i.e. who had plaque fissuring) compared to asymptomatic individuals to better understand the atherosclerosis progression process. We generated single-cell RNA-seq data from CD45+ sorted cells from carotid tissue from 5 symptomatic and 5 asymptomatic patients using drop-seq technology. We have sequenced 16225 single-cells in total, detecting a median of 1152 genes in each cell. Preliminary cluster analysis identified a distinctive pattern of co-expression among genes able to differentiate cells from symptomatic and asymptomatic subjects, demonstrating the presence of a distinctive transcriptional signature in the disease progression. Further work is aiming to identify subpopulations of CD45+ cells present in both types of atherosclerotic plaques, comparison of their signals with circulating immune cells and their possible implication in the transition to a symptomatic, more severe, stage.

3134F

A Mendelian randomization analysis of hemostatic factors in peripheral artery disease. A.M. Small¹, J.E. Huffman², D. Klarin³, M. Sabater-Lleal^{4,5}, J. Lynch⁶, T. Assimes⁷, Y. Sun⁸, D. Miller⁹, M. Freiberg¹⁰, D. Rader¹¹, S. Kathiresan¹², P. Wilson¹³, K. Cho¹⁴, P. Tsao¹⁵, K. Chang¹¹, N. Smith¹³, C.J. O'Donnell¹⁶, P.S. de Vries¹⁴, S.M. Damrauer^{11,15}. 1) Yale University, New Haven, CT; 2) Center for Population Genomics, MAVERIC, VA Boston Healthcare, Boston, MA; 3) Massachusetts General Hospital, Boston, MA; 4) Unit of Genomic of Complex Diseases, IIB Sant Pau, Research Institute of Hospital Sant Pau, Barcelona, Spain; 5) Cardiovascular Medicine Unit, Department of Medicine, Karolinska Institutet, Stockholm, Sweden; 6) VINCI, VA Salt Lake City Healthcare System, Salt Lake City, UT; 7) Stanford University, Palo Alto, CA; 8) Emory University, Atlanta, GA; 9) Boston University, Boston, MA; 10) Vanderbilt University, Nashville, TN; 11) Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 12) Boston VA Healthcare System, Boston, MA; 13) University of Washington, Seattle, WA; 14) University of Texas Health Science Center, Houston, TX; 15) Corporal Michael Crescenz VA Medical Center, Philadelphia, PA.

Introduction: Peripheral artery disease (PAD) affects over 200 million individuals worldwide. A recent GWAS identified novel genetic evidence for association of a prothrombotic coagulation protein variant (the factor V Leiden mutation) with risk of prevalent PAD, indicating a possible role of coagulation and thrombosis in PAD. **Objective:** We assessed the causal role of plasma levels of factor VII (FVII), factor VIII (FVIII), and von Willebrand factor (VWF) in the development of PAD using a 2-sample Mendelian randomization (MR) approach. **Methods:** We performed 2-sample MR using the inverse variance weighted (IVW) approach to calculate the causal odds ratio for each hemostatic factor with PAD among individuals of European ancestry (EA). We combined variants having genome-wide significant associations with plasma levels of FVII (9 loci), FVIII (16 loci), and VWF (5 loci) from GWAS performed by the Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium to construct a single genetic instrument for each hemostatic factor. We then tested the association of each of these three instruments with PAD using summary statistics from a VA Million Veteran Program PAD GWAS. We used Egger Regression and the weighted median approach as sensitivity analyses. A Bonferroni p-value cutoff was used to determine significance [$p < 0.016$ based on the number of hemostatic factors tested ($n=3$)]. **Results:** There were 23,434, 32,610, and 46,354 EA participants for the CHARGE FVII, FVIII, and VWF GWAS meta-analyses respectively. Among 227,817 EA individuals in MVP, there were 24,009 individuals with prevalent PAD. MR analysis using the IVW approach demonstrated statistically significant associations with PAD for FVIII (OR = 1.31, CI 95% 1.13-1.51, $p = 0.0003$) and vWf (OR = 1.24, CI 95% 1.07-1.43, $p = 0.003$), but not for FVII (OR = 1.07, CI 95% 0.89-1.28, $p = 0.45$). Egger regression and the weighted median approaches yielded similar effect estimates and confidence intervals. **Conclusions:** These data suggest a fundamental role of the component factors FVIII and/or VWF in the pathobiology of PAD and promote further study to characterize the role of thrombosis in this highly prevalent and morbid disease.

3135W

The TAGA Study: A new family study for the genetic analysis of aortic abdominal aneurysm. M. Vázquez-Santiago¹, J.M. Romero^{1,2}, J. Dilme^{1,2}, J. Giner³, A. Martínez-Perez¹, A. Buil⁴, O. Sibila³, V. Plaza³, M. Camacho¹, J.M. Soria¹, J.R. Escudero^{1,2}, M. Sabater-Lleal^{1,5}. 1) Unit of Genomics of Complex Diseases (IIB-Sant Pau), Barcelona, Spain; 2) Joint Services of Angiology, Vascular and Endovascular Surgery, University Hospital Sant Pau-Hospital Dos de Maig, Barcelona, Spain; 3) Department of Respiratory Medicine, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain; 4) Institute of Biological Psychiatry, Sankt Hans Hospital, Roskilde, Denmark; 5) Department of Medicine, Karolinska Institutet, Stockholm, Sweden.

Introduction: Aneurysms are characterized by structural deterioration of the vascular wall leading to progressive dilatation and, potentially, rupture of the aorta. Abdominal aortic aneurysms (AAA) are responsible for approximately 11,000 deaths/year in the US, with no current effective medical therapies, highlighting the need to better understand factors influencing pathogenesis and disease progression, and its relation to its most important comorbidities.

In this direction, we have designed and collected the TAGA Study (Triple-A Genetic Analysis), a family study collected through a *proband* with AAA. **Methods:** The TAGA Study comprises 405 individuals related in 12 large Spanish families. Cases were defined as those with dilation in abdominal aortic diameter (AAOD) higher than 30 mm in adults. We applied variance component methods to calculate heritabilities and phenotypic correlations (splitted into genetic and environmental correlations) between 60 quantitative traits and the AAOD, using R package *solaris* as interface for SOLAR-Eclipse software.

Results: Heritability of AAOD was 0.34 (CI=0.24-0.44). Among the main related arteries, the strongest genetic correlation was observed between AAOD and the diameter of left femoral artery ($r^2=0.79$; $p=3.9 \times 10^{-6}$). We also observed significant phenotypic correlations between AAOD and serum creatinine levels ($r^2=0.35$; $p=9.3 \times 10^{-13}$), and lung function tests, the strongest being with forced vital capacity observed (FVC) ($r^2=0.38$; $p=1.6 \times 10^{-12}$), forced expiratory volume in 1 second (FEV1) ($r^2=0.37$; $p=6.4 \times 10^{-13}$), and post dilator FVC ($r^2=0.32$; $p=8.0 \times 10^{-6}$). **Discussion:** Our results indicate that one third of the variability in AAOD is explained by genes, which suggests that well-design genetic studies may be a suitable tool to elucidate novel biological mechanisms underlying this phenotype. We also present phenotypic, genetic and environmental correlation results that indicate that common genetic factors regulate normal variation in the diameter of the main related arteries. While we found strong phenotypic correlations between markers of respiratory alterations and AAOD, these were mainly driven by non-genetic factors independent of smoking. In-depth study of novel intermediate phenotypes genetically correlated with AAOD can be a powerful tool to prioritize the main biological traits and mechanisms underlying AAA formation and development. **Grant:** Spanish Ministry of Health, ISCIII (Miguel Servet CP17/00142).

3136T

Significant enrichment of damaging recessive genotypes in cilia genes is associated with laterality defects in a cohort of 2391 trios with congenital heart disease. W.S. Watkins¹, E.J. Hernandez^{1,2}, B.W. Bisgrove³, B.L. Demerest⁴, B.D. Gelb⁴, D. Bernstein⁵, G. Porter⁶, J.W. Newburger⁷, C.E. Seidman⁸, W.K. Chung⁹, E. Goldmuntz¹⁰, L.B. Jorde¹, H.J. Yost⁸, M. Yandell^{1,2}, M. Tristani-Firouzi¹¹, Pediatric Cardiac Genomics Consortium. 1) Department of Human Genetics, University of Utah, Salt Lake City, UT; 2) USTAR Center for Genetic Discovery, University of Utah, Salt Lake City, UT; 3) Department of Neurobiology & Anatomy, University of Utah, Salt Lake City, UT; 4) Mindich Child Health and Development Institute, Departments of Pediatrics and Genetics & Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 5) Department of Pediatrics, Stanford University School of Medicine, Palo Alto, CA; 6) Department of Pediatrics, University of Rochester Medical Center, Rochester, NY; 7) Department of Cardiology, Boston Children's Hospital, Boston, MA; 8) Departments of Genetics and Medicine, and HHMI; Harvard Medical School; 9) Departments of Pediatrics and Medicine, Columbia University, New York, NY; 10) Division of Cardiology, Department of Pediatrics, Children's Hospital of Philadelphia, PA; 11) Department of Pediatrics, University of Utah, Salt Lake City, Utah.

To investigate the genetic contribution of cilia genes to human congenital heart disease (CHD), we used a burden test approach to assess 2391 trios from the Pediatric Cardiac Genomics Consortium (PCGC) for damaging variation in 3 cilia gene lists: a "gold-standard" list of genes encoding ciliary components, SysCilia, $n=302$; a broader list of cilia genes curated from model organisms, $n=669$; and a list of *FOXJ1*-responsive genes derived from zebrafish, $n=116$. Proband and parent exomes were joint-genotyped with GATK/VQSR and all trio relationships were verified. Damaging recessive and compound heterozygous genotypes were identified using the Variant Annotation, Analysis & Search Tool (VAASST3). Genotypes were defined as damaging by a VAASST3 burden p -value of < 0.005 , consistent with our previous study (PMID 28991257). These CHD probands are highly enriched for damaging compound heterozygous and recessive genotypes in SysCilia and cilia-related genes (p -value $< 1e-5$). *FOXJ1*-responsive genes also show modest enrichment (p -value < 0.026). In contrast, random genes containing a length-normalized equal number of rare variants (gnomAD, $maf \leq 0.005$) and house-keeping gene lists of equal size show no significant enrichment (p -value > 0.2). Cilia genes also demonstrate little enrichment for damaging *de novo* variants, in sharp contrast to chromatin-related genes ($n=163$) which are known to be highly-enriched in damaging *de novo* variation but only moderately enriched in damaging recessive genotypes. We also find that laterality defects are significantly more common in probands with damaged cilia genes than other cardiac phenotypes, such as conotruncal or left ventricular outflow tract obstructive disorders (p -value ≤ 0.015). These results indicate that segregating, rare variation in cilia and cilia-related genes contributes to human CHD, particularly laterality defects.

3137F

Assessing and characterising the causal association of glycine with risk of coronary heart disease. L.B.L. Wittemans¹, L.A. Lotta¹, C. Oliver-Williams², I.D. Stewart¹, P. Surendran², S. Karthikeyan², F.R. Day¹, A. Koulman^{1,3}, L. Zeng^{4,5}, J. Erdmann^{6,7,8}, F. Imamura¹, H. Schunkert^{4,5}, K.T. Khaw⁹, J.L. Griffin¹⁰, N. Forouhi¹, A.M. Wood², S. Burgess^{2,11}, R.A. Scott¹, J.M.M. Howson², J. Danesh^{2,12}, N.J. Wareham¹, A.S. Butterworth², C. Langenberg¹, EPIC-CVD.

1) MRC Epidemiology Unit, University of Cambridge, Cambridge, United Kingdom; 2) Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom; 3) NIHR BRC Nutritional Biomarker Laboratory, University of Cambridge, Cambridge, United Kingdom; 4) Deutsches Herzzentrum München, Technische Universität München, München, Germany; 5) DZHK, Partner Site München Heart Alliance, München, Germany; 6) Institute for Cardiogenetics, University of Lübeck, Lübeck, Germany; 7) DZHK, Partner Site Hamburg/Lübeck/Kiel, Lübeck, Germany; 8) University Heart Center Lübeck, Lübeck, Germany; 9) Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom; 10) Biochemistry Department, University of Cambridge, Cambridge, United Kingdom; 11) MRC Biostatistics Unit, University of Cambridge, Cambridge, United Kingdom; 12) Wellcome Trust Sanger Institute, Genome Campus, Hinxton, United Kingdom.

Background/aims: Observational and animal-based studies suggest an aetiological role of glycine in coronary heart disease (CHD). Five genetic loci have been reported for glycine, including *CPS1*, which is more strongly associated in women and for which a sex-specific cardioprotective effect has been proposed. We aimed to (1) identify additional genetic loci for plasma glycine levels, (2) investigate the causal role of glycine on risk of CHD in men and women, and (3) characterise the biological pathways underlying the glycine-CHD association. **Methods:** We performed a meta-GWAS for glycine in up to 80,003 participants from 3 novel studies and 2 publicly available summary-level GWAS datasets. Sex-specific effect sizes were estimated in 30,226 men and 31,957 women. Using two-sample Mendelian randomization (MR), we assessed the genetic association of glycine with CHD risk, based on 88,800 sex-combined cases and 9,852 female and 21,994 male cases for sex-specific analyses. We tested the genetic association of glycine with CHD risk factors, including systolic (SBP) and diastolic blood pressure (DBP), blood lipids and 13 potentially relevant blood cell traits. The role of CHD risk factors as potential mediators of the glycine-CHD association was assessed using multivariable MR. **Results:** We identified 27 loci, of which 6 are in genes encoding enzymes of glycine metabolism and 22 have not previously been linked to glycine. Genetically predicted glycine was significantly associated with CHD in sex-combined analyses (OR [95% CI] per SD genetically higher glycine=0.95 [0.92,0.98]). Based on a subset of the CHD cases, there was no evidence that this association differed by sex (women: OR = 0.95 [0.91,1.00]; men: OR = 0.93 [0.84,1.02]; *p* for sex difference=0.60). After taking into account the sex-specific effect on glycine, the genetic glycine-CHD association based on *CPS1* only did not differ by sex. Glycine was genetically associated with SBP and DBP but not with other CHD risk factors. Adjusting for genetically predicted SBP and DBP attenuated the association of glycine with CHD. **Conclusion:** Using substantially enhanced genetic predictors for glycine, we demonstrated that glycine is genetically associated with lower CHD risk, which suggests a cardio-protective role of glycine. Despite previous suggestions, there was no evidence that the effect of glycine was sex-specific. We also found that the inverse relationship between glycine and CHD may be explained by blood pressure.

3138W

The relationships between genotype and histopathological alterations in hypertrophic obstructive cardiomyopathy patients. KL. Yin, Y. Ma, CS. Zhu, Y. Sun, WZ. Zhou, H. Yang, WK. Li, BM. Han, XW. Liu, HY. Wang, SY. Wang, Z. Zhou. State Key Laboratory of Cardiovascular Disease, Fuwai Hospital, National Center for Cardiovascular Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, People's Republic of China.

Aims: Genotype has been related to specific phenotype of hypertrophic cardiomyopathy (HCM) patients, but its relationship with histopathological alterations remains uncertain. The purpose of the study is to explore whether there are differences in the relationship between the genotype and histopathological alterations under different genetic backgrounds. **Methods:** A 93-cardiomyopathy-related-gene panel was used to screen 193 Chinese patients with hypertrophic obstructive cardiomyopathy (HOCM), and all rare variants were further verified by Sanger sequencing. The pathogenicity of all variants was classified into 5 categories according to American College of Medical Genetics and Genomics (ACMG). In addition to variants of unknown significance (VUS), an additional tier, VUS-LP was used to define those VUS that lacked one piece of supporting evidence for classification as LP. The effects of different numbers or pathogenicities of 8 major sarcomeric variants on histopathological alterations and phenotypes were evaluated in HOCM patients using clinical parameters or 177 septal tissue sections with Masson trichromatic staining, respectively. **Results:** The overall diagnosis rate was 45.1%. 7 splicing variants predicted with *in silico* tools were validated in cardiac tissues, and all but the novel canonical variant in *ACTN2* (c.1515+1G>T) were confirmed to cause aberrant splicing. Notably, the patients with VUS-LP variants shared similar clinical and histopathological characteristics to those with LP/P variants. The existence of any variant or LP/P variants, regardless of number, was related to an earlier diagnosis and LV dysfunction. Moreover, compared with the patients with VUS and no variants, those with LP/P and VUS-LP variants exhibited relatively severe phenotypes and significantly worse histopathological alterations including hypertrophy, disarray, interstitial and replacement fibrosis. **Conclusions:** For the first time, the characteristic effects of genotype, including the numbers and pathogenicities of variants, on the histopathological alterations and certain phenotypes were revealed through large-scale comprehensive examinations of tissue sections and panel screening. This approach not only filled in the gaps between genotype and histopathological alterations in HCM patients but also had the potential to contribute to the prediction of clinical outcomes. The additional tier of VUS-LP had significance for the future pathogenicity interpretations.

3139T

A new African risk factor for ventricular arrhythmia. P. Guicheney¹, M. Blancard¹, A. Debiche¹, K. Kato¹, C. Cardin², G. Guichard³, E. Gandjbakhch^{1,4}, V. Probst⁵, M. Hocini⁶, F. Extramiana⁷, M. Haissaguerre⁶, O. Geoffroy⁶, A. Leenhardt⁷, J.S. Rougier⁸. 1) Inserm, Sorbonne University, UMRS1166, Paris, France; 2) University Hospital Rangueil, Toulouse, France; 3) University of Bern, Institute of Biochemistry and Molecular Medicine, Bern, Switzerland; 4) AP-HP, Hôpital Pitié-Salpêtrière, Département de Cardiologie, Paris, France; 5) CHU Nantes, L'institut du thorax, Service de Cardiologie, Nantes, France; 6) L'Institut de Rythmologie et Modélisation Cardiaque (LIRYC), Université de Bordeaux, Bordeaux, France; 7) AP-HP, Hôpital Bichat, Département de Cardiologie, Paris, France; 8) CHU Sud Réunion, La Réunion, France.

Objectives: Cardiac channelopathies are genetic disorders associated with an increased risk of ventricular arrhythmia and sudden death (SD) in young individuals with a structurally normal heart. An important aspect of such cardiac diseases is the precise calcium regulation. One of the most important effector is the voltage-gated calcium channel Ca_v1.2. Several variants in Ca_v1.2 have been implicated in various arrhythmic syndromes. We screened 19 patients with ventricular fibrillation in the absence of any cardiac structural abnormalities for variants in Cav1.2 subunit genes. All these patients presented short-coupled premature ventricular beats or torsade de pointes originating from Purkinje fibres. **Methods:** Ca_v1.2 related genes, *CACNA1C*, *CACNB2* and *CACNA2D1*, were screened in probands by exome sequencing. Missense variants were introduced in Ca_v1.2 alpha subunit plasmid by mutagenesis to assess their pathogenicity using patch clamp approaches. **Results:** The first symptoms occurred in the probands at a mean age of 32.9 ± 10.9 years, eight were male, and six had a family history of sudden death. Missense variants in *CACNA1C* encoding the Ca_v1.2, alpha subunit were identified in four individuals. Two variants were found in an inhabitant of La Reunion Island found, Ca_v1.2-T1787M and Ca_v1.2-C1992F. Interestingly, we previously identified Ca_v1.2-T1787M in another patient originating from Cameroon with syncope and early repolarisation pattern on ECG recording, in combination with the African pro-arrhythmic Na_v1.5-S1103Y variant of the cardiac sodium channel. Although the mean allele frequency of Ca_v1.2-T1787M was of superior to 1% (MAF=0.016) in the African population according to ExAC database, we performed patch-clamp analysis. It revealed that the C-terminal Ca_v1.2-T1787M; 1) reduces the calcium and barium currents probably by increasing the C-terminal auto-inhibition mechanism and 2) increases the voltage-dependent inhibition. The other variants did not induce major changes of the channel function. **Conclusion:** We identified a new loss-of-function variant, Ca_v1.2-T1787M, present in 0.8% of the African population, as a new potential risk factor for lethal-ventricular arrhythmia combined other risk factors.

3140F

Genetic variants of electrical storm in post myocardial infarction patients. A. Rangaraju¹, S. Krishnan², G. Aparna³, S. Shankaran³, A. Mannan³, H. Bhaganagar^{2,4}. 1) Krishna Institute of Medical Sciences and Research centre, Telangana, India; 2) Division of Pacing & Electrophysiology, Krishna Institute of Medical Sciences, Minister Road, Secunderabad, Telangana, India; 3) STRAND Life Sciences, Bangalore, India; 4) Arrhythmia Research & Training Society (ARTS), Hyderabad, India.

Sudden Cardiac Death (SCD) in patients with remote myocardial Infarction (MI) is due to the occurrence 'Ventricular Tachycardia' (VT). Few patients in this subset during their natural history develop Electrical storm (ES) which is defined as "Three or more distinct episodes of ventricular tachycardia (VT)/ventricular fibrillation (VF) within 24h, requiring the intervention of the defibrillator or ICD. Though a few clinical pointers exist, the occurrence of ES in a patient with remote myocardial infarction (MI) is generally unpredictable. It has a reported incidence of 10 -28% and an in-hospital mortality of 60 - 70%. Genetic markers for this entity have not been studied. Consecutive ten patients with Left ventricular dysfunction (LVEF< 35%), underlying remote myocardial infarction (>1 year), presented to our institute with ES and hemodynamically unstable monomorphic VT, were included in the study. To ensure homogeneity of the phenotype, we selected patients with a specific substrate presenting only with monomorphic unstable VT. After taking the informed consent, genetic analysis was performed in all patients by next generation sequencing (NGS) of SCD panel. 145 genes associated with arrhythmias and coronary artery disease were assessed which revealed 25 rare variants in 19 genes, predominantly in *RYR2*, *SCN5A*, *KCNJ11*, *KCNE1* and *KCNH2*, *CACNA1B*, *CACNA1C*, *CACNA1D* and desmosomal genes – *DSP* and *DSG2*. These are essentially cardiac ion channel genes and whose role is established in LQT and other arrhythmic disorders. Of the 25 rare variants, pVal125Leu of *SCN5A* was found to be pathogenic while p.Val30Met of *DSP* and p.Thr1107Met of *RYR2* revealed mixed interpretations of pathogenicity. We hypothesize that underlying pathophysiology of SCD due to ES of unstable VT may be similar to that of Ion channelopathies. Variations in genes coding for cardiac ion channels (*SCN5A*, *KCNE1*) that are currently being used to stratify arrhythmic risk in patients with inherited syndromes of SCD, may also be associated with occurrence of ES in patients with remote MI. The clinical relevance of this study is the possibility that identification of such genetic variants may better risk stratify high-risk MI subsets prone to electrical storm. However, studies on larger cohorts in various ethnic groups are required, as multiple genes among the panel seem to influence the function and a cumulative effect of various mutations play a role in the manifestation of the conditions.

3141W

Mouse and human genomic studies identify modifiers of blood pressure and supravalvular aortic stenosis in elastin insufficiency and Williams Beuren syndrome. B. Kozel¹*, P.C.R. Parrish¹*, R.H. Knutsen¹, A. Troia¹, A. Wardlaw¹, C.J. Billington¹, D. Liu¹, Y.P. Fu². 1) Laboratory of Vascular and Matrix Genetics, Translational Vascular Medicine Branch, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD; 2) Office of Biostatistics Research, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD.

In rare diseases of haploinsufficiency, phenotype severity typically varies among affected individuals. However, identification of genetic modifiers is challenging due to small sample size. Williams Beuren syndrome (WBS), a microdeletion disorder that causes a vascular disease notable for stenosis of the aorta (SVAS), occurs in approximately 1 in 10,000 births and shows significant phenotypic variation. To approach this power problem, we have used a two-pronged approach, combining modifier work in humans and in murine models to identify common genes and pathways that underlie the difference in vascular disease severity among individuals with elastin insufficiency. In our mouse studies, we used quantitative trait locus analysis in *Elm^{-/-}* mice of different genetic backgrounds, followed by congenic line development to identify modifiers of blood pressure and aortic diameter. This work identified differences in renin-angiotensin signaling and reactive oxygen species generation as potential modifiers. Those findings were then confirmed in a cohort of human patients with the condition. In our primary human studies, we performed exome sequencing in 104 well-phenotyped WBS patients and used SKAT-O and gene set enrichment analysis to identify genes and pathways with increased variation in patients with no vs. severe aortic disease. We tested several mechanisms to improve power for our studies including extreme phenotype modeling and gene-set/pathway-based approaches. While not meeting stringent criteria for statistical significance, pathway testing revealed that patients with more severe disease had more variation in connective tissue and VEGF signaling pathways. In addition, discovery based gene set enrichment analysis suggested a role for inflammation in SVAS pathogenesis. These primary findings are being replicated in a larger (>400 subject) cohort and we have successfully targeted relevant inflammatory pathways in our mouse models, showing complimentary results. Taken together, our studies in humans and mice have allowed us to detect important genomic variations that contribute to disease severity among individuals with WBS with small sample numbers. The use of complimentary models allows for the results of lower powered studies to be directly tested by secondary methods to assure replicability and biological significance. *PCRP is the first author for this work and BAK is the senior author but will present in PCRP's absence.

3142T

Whole genome sequencing approach to identify genetic variation influencing heart failure: A Singaporean perspective. C. Bellis¹, N. Bertin², CA. Lin¹, TH. Koh¹, C. Wang³, WW. Soon⁴, A. Wilm⁵, CC. Shih⁶, RS. Foo^{6,7}, AM. Richards⁸, CS. Lam^{9,10}, J. Liu¹, SG10K consortium. 1) Human Genetics 2, Genome Institute of Singapore, Agency for Science, Technology and Research of Singapore (A*STAR), Singapore; 2) Centre for Big Data and Integrative Genomics (c-BIG), Genome Institute of Singapore, Agency for Science, Technology and Research; 3) Computational and Systems Biology, Genome Institute of Singapore, Agency for Science, Technology and Research of Singapore, Singapore; 4) Next Generation Sequencing Platform, Genome Institute of Singapore, Agency for Science, Technology and Research of Singapore, Singapore; 5) Scientific and Research Computing, Genome Institute of Singapore, Agency for Science, Technology and Research of Singapore, Singapore; 6) Cardiovascular Research Institute, National University Health System, Singapore; 7) Human Genetics 3, Genome Institute of Singapore, Agency for Science, Technology and Research of Singapore, Singapore; 8) Duke-National University of Singapore, Singapore; 9) National Heart Centre Singapore, Singapore; 10) Singapore Bioimaging Consortium, Agency for Science, Technology and Research of Singapore, Singapore.

Heart failure (HF) is the final common pathway of cardiovascular diseases (CVD) and leading cause of CVD hospitalizations in Singapore and worldwide with a mortality rate that rivals most cancers (>50% mortality in 5 years). Also similar to cancer, HF is a staged disease where early detection and treatment are vital for prevention of disease progression from Stage A (risk factors) to Stage B (subclinical structural heart disease) and Stage C (symptomatic HF). Cardiac imaging is critical for the detection of early subclinical disease. Currently, there is a gap in the understanding of the mechanisms underlying disease progression and effective therapies are limited, particularly for the proportion of the HF population with preserved ejection fraction (HFPEF or "diastolic HF") for which there is no effective therapy to date (in contrast to HF with reduced ejection fraction [HFrEF] or "systolic HF"). The central aim of this project is to uncover disease predisposition by studying the full allelic spectrum of genetic determinants in carefully defined clinical cohorts (SHOP [The Singapore Heart Failure Outcomes and Phenotypes study] and ATTRaCT [The Asian neTwork for Translational Research and Cardiovascular Trials; <https://www.a-star.edu.sg/attract/About-Us/Overview>]) that have been recruited for this study. To achieve such an aim, a research strategy has been developed with the following approach; recent advancements in genomic technologies allow simultaneous performance of whole-genome sequencing (WGS) to uncover common genetic variants as well as rare (possible protein-coding) changes, extending us the opportunity to perform comprehensive genetic and genomic analyses. Genotypes will be systematically assessed for associations with either HF subtype. In our analyses, we will compare (a) HF vs non-HF controls; (b) HFrEF vs HFPEF. Currently, we have sequenced approximately 1,500 participants across the SHOP and ATTRaCT heart failure cohort studies, which include controls and cases (~70% HFrEF). Patients enrolled in the ATTRaCT cohort have been extensively/fine-grained phenotyped through detailed and careful imaging with cardiac magnetic resonance (CMR) and advanced molecular studies that we will integrate into our analysis for genotype association.

3143F

Genetic risk scores for cardiometabolic traits in sub-Saharan Africans. K. Ekoru, A.A. Adeyemo, G. Chen, A.P. Dourmatey, J. Zhou, A.R. Bentley, D. Shriner, C.N. Rotimi, the AADM study investigators. Center for Research on Genomics and Global Health, National Human Genome Research Institute, Bethesda, MD.

Background Genetic risk scores (GRS) have been shown to be robustly associated with several traits and are increasingly being proposed for use in clinical prediction. However, their evidence base derives mostly from studies in populations of European and/or Asian ancestry. The utility of GRS in cardiometabolic trait risk prediction has not been assessed in sub-Saharan African populations. We evaluated GRS for a set of cardiometabolic traits incorporating genetic variants from published GWAS and assess their potential for risk prediction in a large genetic epidemiology study. **Methods** Participants were African men and women aged 18 years or older (N=5231, 63% women) enrolled from Ghana, Nigeria and Kenya in the AADM Study. GRS were constructed by summing the number of risk alleles (0, 1, 2) over unique SNPs reported in the NHGRI-EBI Catalog for each of twelve traits, including body mass index (BMI), waist circumference (WC), hip circumference (HC), waist-to-hip ratio (WHR), systolic blood pressure (SBP), diastolic blood pressure (DBP), fasting plasma glucose (FPG), triglycerides (TG), total cholesterol (TC), LDL-cholesterol (LDL), HDL-cholesterol (HDL) and type 2 diabetes (T2D). The predictive potential of the GRS for each trait was assessed using the additional phenotypic variance explained over that explained by age, sex, BMI (where appropriate) and principal components of ancestry. **Results** GRS for BMI, TG, TC, LDL, HDL and WHR were significantly associated with the respective traits. For every additional risk allele, BMI increased by 0.02 kg/m², TG by 0.469 mg/dL, TC by 0.52 mg/dL, LDL by 0.557 mg/dL, HDL by 0.146 mg/dL and WHR by 0.0004. The additional variation explained by GRS for these traits ranged from 0.4% in WHR to 32.6% in HDL. GRS was significantly associated with T2D (OR=1.021, P<0.0001) and explained an additional 14.3% of variation in the probability of having T2D. GRS was not significantly associated with SBP, DBP, WC or HC. **Conclusion** The strength of association and predictive ability of GRS for cardiometabolic traits show marked variation in African ancestry populations, probably reflecting population differences in associated variants, LD and effect size, among others. More genome wide studies in African populations are likely to identify significant variants not currently included in our prediction models with the possibility of better optimization of GRS for clinical risk prediction of cardiometabolic traits in African ancestry populations.

3144W

Genomics to medicine - P5 FinHealth Study. K. Kristiansson¹, M. Kinnunen¹, P. Wurtz², M. Marttila¹, A. Haukkala³, M. Brunfeldt¹, H. Kääriäinen¹, M. Perola¹. 1) National Institute for Health and Welfare, Helsinki, Finland; 2) Nightingale Health Ltd, Helsinki, Finland; 3) University of Helsinki, Helsinki, Finland.

Introduction Research on omics, including genomics and metabolomics, has identified thousands of genomic regions and several blood metabolites linked to common human diseases and features. This new knowledge combined with large amounts of samples collected to biobanks could provide new tools for predictive, preventive, personalized, and participatory (P4) health-care. **Methods** P5 study, P4 + Population health, studies the value of returning genetic and metabolomic risk information on coronary heart disease, type 2 diabetes and venous thromboembolism. We have invited 6.000 FinHealth 2017 study participants in our study. Prospective clinical significance of the risk scores will be based on analysis and follow-up of 30.000 Finnish individuals. **Results** So far, clinical utilization of omics is fairly diminutive and, above all, unvalidated. We will analyze the genetic and metabolomic profile of the P5 participants and develop and test a protocol for returning them health related risk information. The impact of our intervention will be studied by following up the participants by questionnaires and through national health registers for five years. **Conclusions** We hypothesize that 1) combining genetic and metabolomic risk with traditional risk factors adds value to the personal risk assessment of these diseases, 2) such risk information can be provided online, and 3) receiving genetic and metabolomic risk information will have an effect on the health of the study participants.

3145T

Overview of the NHLBI Trans-Omics for Precision Medicine (TOPMed) program: Whole-genome sequencing of >100,000 deeply phenotyped individuals.

C. Laurie¹, T. Blackwell², G. Abecasis³, K. Rice⁴, J. Wilson⁵, D. Nickerson⁶, S. Gabriel⁷, R. Gibbs⁸, S. Dutcher⁹, S. Germer¹⁰, D. Arnett¹¹, A. Ashley-Koch¹², K. Barnes¹³, E. Boerwinkle¹⁴, S. Rich¹⁵, E. Silverman¹⁶, R. Beer¹⁷, J. Mikulla¹⁸, P. Srinivas¹⁹, W. Gan²⁰, G. Papanicolaou²¹, C. Jaquish²², NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium. 1) University of Washington, Seattle, WA, USA; 2) University of Michigan, Ann Arbor, MI, USA; 3) University of Mississippi, Jackson, MS, USA; 4) The Broad Institute, Massachusetts Institute of Technology, Cambridge, MA, USA; 5) Baylor College of Medicine Human Genome Sequencing Center, Houston, TX, USA; 6) Washington University, St. Louis, MO, USA; 7) New York Genome Center, New York, NY, USA; 8) University of Kentucky, Lexington, KY, USA; 9) Duke University, Durham, NC, USA; 10) University of Colorado at Denver, Denver, CO, USA; 11) University of Texas Health Science Center at Houston, Houston, TX, USA; 12) University of Virginia, Charlottesville, VA, USA; 13) Brigham and Women's Hospital, Boston, MA, USA; 14) National Heart, Lung and Blood Institute, Bethesda, MD, USA.

A primary goal of the NHLBI TOPMed program is to improve scientific understanding of the fundamental biological processes that underlie heart, lung, blood, and sleep disorders. TOPMed is providing deep whole-genome sequencing (WGS) to pre-existing 'parent' studies having large samples of human subjects with deep phenotypic characterization (biochemical, physiological, clinical, behavioral, and anatomical measures) and environmental exposure data. WGS was performed to an average depth of >30x. A support vector machine quality filter was trained with known variants and Mendelian-inconsistent variants. The average pairwise non-reference genotype discordance rate among 69 pairs of duplicate samples was estimated at 5×10^{-5} . Currently, >100,000 samples have completed sequencing, with ~700 million variants discovered. Genotypes are being called jointly over all studies to facilitate cross-study analyses, including the use of common controls. TOPMed studies include prospective cohorts; cross-sectional case-control studies; family-based studies; and 'synthetic cohorts' consisting of cases from multiple sources with common controls from other sources. The prospective cohorts provide large numbers of disease risk factors, subclinical disease measures, and incident disease cases, while the case-control studies provide large numbers of prevalent disease cases. Extended family structures can provide improved power to detect rare variant effects. Well-represented diseases include asthma, COPD, atrial fibrillation, stroke and venous thromboembolism. Among currently sequenced individuals, the approximate ancestral/ethnic composition is 43% European-American, 31% African-American, 19% Hispanic/Latino, 5% Asian and 3% other groups. Discovery of genotype-phenotype associations frequently includes pooled analysis across ancestry groups and studies, using statistical models that account for population structure and relatedness. The resources of TOPMed are being made available to the scientific community as a series of 'data freezes': genotypes and phenotypes via dbGaP; read alignments via the SRA; and variants via the Bravo variant server and dbSNP. Genotypes for a set of 18.5k samples have been released on dbGaP, another freeze of 55k is being released in Q2 2018, and a freeze of >100k is planned for Q1 2019. Genotypic data are contained in study-specific accessions with names containing "NHLBI TOPMed", while most phenotypic data are in parent study accessions.

3146F

Phenome wide scan of two LPA loss-of-function variants reveals insights for drug development. S.E. Ruotsalainen¹, S. Surakka^{1,2}, J. Koskela¹, V. Salomaa³, A. Palotie^{1,4}, M.J. Daly^{1,4}, S. Ripatti^{1,4,5}.

1) Institute for Molecular Medicine Finland, Helsinki, Finland; 2) Department of Internal Medicine, University of Michigan, Ann Arbor, MI, US; 3) National Institute for Health and Welfare, Helsinki, Finland; 4) Broad Institute of MIT and Harvard, Cambridge, Massachusetts, US; 5) Department of Public Health, University of Helsinki, Helsinki, Finland.

Lipoprotein(a), Lp(a), level is a highly heritable, independent causal risk factor for coronary heart disease. Genetic variation in and around *LPA* gene accounts for 75% of the heritability. Two known loss-of-function (LoF) variants in *LPA* that are enriched in the Finnish population (9.6 and 1.4-fold compared to Europeans) have previously been shown to have protective effect on cardiovascular diseases through decreased Lp(a) levels. Lp(a) levels have also been associated with risk of multiple diseases, including Type 2 diabetes (J Lipid Res. 2018 May;59(5):884-891) and vascular dementia (Jpn J Psychiatry Neurol. 1987 Dec;41(4):743-8.). We set to test the association between the *LPA* LoF variants and a wide range of quantitative biomarkers and disease endpoints through a phenome-wide association scan (PheWAS). We calculated *LPA* LoF dosage and tested its association with 620 disease endpoints derived from hospital discharge registry, prescription drug registry and causes of death registry with decades of follow-up and covering the whole nation and 108 biochemical and physiological quantitative measurements. We performed our study using data set of over 27,000 Finns from population-based FINRISK Study with genome-wide genotype data linked to nationwide health registries. We used logistic regression for endpoints, and linear regression for quantitative measurements to get the associations with *LPA* LoF dosage. In the PheWAS scan, the *LPA* LoF dosage is associated with lower risk for cardiovascular diseases, including vascular dementia (OR = 0.339, $p = 0.0013$) and arrhythmias (OR = 0.80, $p = 0.0028$). On the other hand, the *LPA* LoF dosage increases risk for multiple sun exposure related skin disorders, especially actinic keratosis (OR = 1.451, $p = 0.00044$). In our PheWAS for quantitative traits, we see a strong lowering effect for Lp(a) level as expected (-0.63 SD per LoF allele, $p = 2.07 \times 10^{-84}$), and also for multiple inflammatory markers including interleukins 4/6/12p70 and FGF (p -values < 0.007). In total, *LPA* LoF dosage was associated with 12 disease endpoints and 6 quantitative measures. Our results highlight the potential disease areas in addition to CHD where an intervention of lowering Lp(a) levels may provide protective effects, but also some of the potential side effects identified through elevated risk in our PheWAS scan. It also shows the power of rich electronic health record data in combination with enriched LoF variants in Finns.

3147W

MyGeneRank: A mobile app for polygenic risk estimation. *N.E. Wineinger, S. Liu, Y. Zhang, A. Torkamani.* Scripps Translational Science Institute, Scripps Research Institute, La Jolla, CA, USA.

Recent studies have demonstrated clinical and personal utility of genetic risk profiling for a subset of common adult-onset diseases in the form of polygenic risk (PR) scores. Coronary artery disease (CAD) has been a popular target of studies on PR as researchers have demonstrated preferential uptake and effectiveness of preventative actions and behaviors based on an individual's estimated PR. Despite general public interest, there exist many barriers which limit the adoption of PR utilization outside of the traditional scientific environment, including access, validity, and interpretability. In effort to study some of these factors, we developed a free, mobile app called MyGeneRank (MGR). MGR uses genetic data supplied by the user from direct-to-consumer genetic testing services (e.g., via access provided through 23andMe's API) and provides unbiased, ancestry-adjusted PR estimates of CAD to the user through the app. Within MGR, users electing to participate in the study are asked a short series of questionnaires assessing their health characteristics and behaviors related to CAD, and can access information aimed to guide the user in their understanding of their PR as well as determine how non-genetic, modifiable factors influence their overall risk. The iPhone version of the app was released in late 2017, and includes 540 unique study participants at the time of this writing with an Android version in internal testing. Nearly half of the study participants are over the age of 40, with one-third over 50. Approximately 75% of users have opted-in to allow us to access objective health metrics captured by Apple's HealthKit. Based upon genetically derived estimates, 75% of users are 90% or more European. But despite including individuals of non-European ancestry in our study, we have observed no evidence of bias in the PR estimation we provide to users ($p=0.36$). This presentation will also detail the backend and user-facing architecture of MGR, our methods for eliminating bias due to population stratification, and how we are leveraging the smartphone to identify relationships with new forms of phenotypic data.

3148T

Causal association between birth weight and adult diseases: Evidence from a Mendelian randomisation analysis. *X. Zhou¹, P. Zeng².* 1) University of Michigan, Ann Arbor, MI; 2) Xuzhou Medical University, Xuzhou, Jiangsu 221004, China.

Birth weight is a widely used surrogate measurement of early life development, which has long been hypothesized to have a profound long-term impact on individual's predisposition to various diseases at adulthood — a hypothesis commonly referred to as the fetal origins of adult diseases. To investigate the fetal origins of adult diseases hypothesis and determine the extent to which birth weight causally affects various adult diseases, we performed a series of comprehensive Mendelian randomisation analyses to examine the causal effects of birth weight on 21 adult diseases and 38 other complex traits. Our study relies on summary statistics obtained from 37 genome-wide association studies with sample sizes ranging from 4,798 (for rheumatoid arthritis) to 339,224 (for body mass index), thus representing the largest and most comprehensive Mendelian randomisation analysis performed on birth weight to date. The large sample size used in our study allows us to fully establish an inverse causal relationship between birth weight and three adult diseases that include CAD, myocardial infarction (MI) and T2D. These inferred causal relationships are robust with respect to the selection of instrumental variables and to the choice of statistical methods, and are carefully validated in the present study through various sensitive analyses. In addition, our analysis also suggests that the lack of causality evidence between birth weight and the other diseases may be partly due to a lack of statistical power resulting from relatively small sample sizes for the remaining diseases. Finally, we investigate the possibility that any of the analyzed 38 complex traits may mediate the causal effects of birth weight on CAD, MI or T2D. Overall, our study provides important causality evidence supporting the fetal origins hypothesis for three adult diseases and suggests that increasing sample size is likely needed to reveal causal effects of birth weight for the other disorders.

3149F

GWAS of resistant hypertension in over 70,000 participants from the Million Veteran Program identifies genes with altered expression in the adrenal gland. J.N. Hellwege^{1,2}, A. Giri^{2,3}, E.S. Torstenson^{1,2}, J.M. Luther⁴, J.M. Keaton^{1,2}, O. Wilson^{2,5}, C.P. Kovesdy⁶, C. Robinson-Cohen⁶, C.P. Chung^{4,7}, K.A. Birdwell^{8,9}, Y.V. Sun⁹, P.W.F. Wilson¹⁰, P.S. Tsao¹¹, C.J. O'Donnell¹², D.R. Velez Edwards^{2,3}, A.M. Hung^{2,5}, T.L. Edwards^{1,2} on behalf of the VA Million Veteran Program. 1) Division of Epidemiology, Department of Medicine, Institute for Medicine and Public Health, Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Tennessee Valley Healthcare System (626)/Vanderbilt University, Nashville, TN, USA; 2) Biomedical Laboratory Research and Development, Tennessee Valley Healthcare System (626)/Vanderbilt University, Nashville, TN, USA; 3) Department of Obstetrics & Gynecology, Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA; 4) Division of Clinical Pharmacology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA; 5) Division of Nephrology and Hypertension, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA; 6) Nephrology Section, Memphis VA Medical Center, Memphis TN, USA; 7) Division of Rheumatology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA; 8) Division of Nephrology, Department of Medicine, Nashville Veteran Affairs Hospital, Nashville, TN, USA; 9) Department of Epidemiology, Emory University Rollins School of Public Health, Department of Biomedical Informatics, Emory University School of Medicine, Atlanta, GA, USA; 10) Atlanta VAMC and Emory Clinical Cardiovascular Research Institute, Atlanta, GA, USA; 11) VA Palo Alto Health Care System; Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA, USA; 12) VA Boston Healthcare, Section of Cardiology and Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA.

Resistant hypertension (rHTN) is the failure to achieve outpatient blood pressure (BP) control to <140/90 mmHg, despite use of three antihypertensive drugs including a thiazide diuretic or use of four or more drugs regardless of achieving control. rHTN is associated with worse health outcomes than controlled hypertension (cHTN) and is more common in African ancestry individuals. Previous GWAS of rHTN with a smaller sample did not identify significant associations. We performed a large GWAS of rHTN using electronic health records linked to genotypes in the VA Million Veteran Program across European American (EA, 78%) and African American (AA, 22%) individuals. All participants had eGFR ≥ 60 ml/min/1.73m². rHTN vs. cHTN was regressed onto additively coded genotypes imputed to 1000 Genomes, adjusting for age, age², sex, body mass index, and 10 principal components within ethnicity. Inverse-variance-weighted fixed-effects meta-analysis across ethnicities included a total of 16,830 cases of rHTN and 53,376 cases of cHTN as controls. Eleven SNPs were genome-wide significant in trans-ethnic meta-analysis in or nearby *CASZ1*, *WNT2B*, *KCNK3*, *FBN2*, *HOXA13-HOTTIP*, *BLK*, *LSP1*, and *RXFP2*, and 3 intergenic SNPs (rs9398815, rs35444, and rs74429862). Two additional SNPs were significant in AA only (5,071 rHTN) near *CACNA1D* and *GUSBP10-MIR3147*, with lead SNPs too rare in EA to confirm associations. All loci, with the exception of *GUSBP10-MIR3147*, have also been associated with BP, and *CACNA1D* is targeted by calcium channel blockers often used in African ancestry patients. We also applied genetically-predicted gene expression association analysis using S-PrediXcan and 44 tissue eQTL references from the GTEx Project. Increasing expression of *RXFP2* and *FBN2* in adrenal tissue was associated with increased risk of rHTN (p-values = 9.48×10^{-17} and 1.89×10^{-7} , respectively). *RXFP2* is a relaxin receptor, which causes vasodilation, increases cardiac output and renal perfusion, and has been investigated as a treatment for heart failure. *FBN2* encodes fibrillin 2, which is a component of connective tissue microfibrils, and shares functions with *FBN1*, the Marfan syndrome gene. Inclusion of expression data from RNA sequencing of tubules and glomeruli isolated from human kidneys with S-PrediXcan analysis is in preparation. This is the first study to identify significant loci influencing rHTN risk, as well as genes with expression changes contributing to the pathophysiology of rHTN.

3150W

Common variants do not explain significant linkage peaks of carotid artery traits in exceptionally long-lived families. A.L. Kuipers¹, M.K. Wojczynski², E.J.M. Barinas-Mitchell³, R.L. Minster³, L. Wang², L. Christiansen⁴, A. Kulminski⁵, B. Thyagarajan⁶, M.A. Province², A.B. Newman¹ On behalf of the Long Life Family Study. 1) Epidemiology, University of Pittsburgh, Pittsburgh, PA; 2) Genetics, Washington University of St. Louis, St. Louis, MO; 3) Human Genetics, University of Pittsburgh, Pittsburgh, PA; 4) Epidemiology, Biostatistics and Biodemography, Danish Aging Research Center, Odense C, Denmark; 5) Social Science Research Institute, Duke University, Durham, NC; 6) Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN.

Dozens of common genetic variants with small effects on vascular disease have been identified. Yet, it is likely that rare private variants with large effects exist within families. In order to test this, we used data captured via B-mode ultrasonography of the carotid arteries in the Long Life Family Study (N=2020), a family-based study of exceptional longevity. We previously found strong evidence of genetic linkage to carotid artery interadventitial diameter and atherosclerotic plaque on chromosomes 3 and 17 (max LOD, peak cM[cM range]: 5.3,20[113–134] and 3.2,8[72–89]), respectively. Haplotypes generated from genotypes typed on the Human Omni Chip 2.5 v1 (Illumina, CA) were used as linkage markers, and multipoint identity-by-descent estimates were calculated by Loki. In the current study, we performed linkage analysis allowing for heterogeneity between families (HLOD) using SOLAR to refine the peaks. We also tested for genetic association between variants under the refined peaks with a minor allele frequency $\geq 1\%$ and carotid artery traits using lmeqin in R. We then performed linkage analyses conditioned on any suggestive or significant variant under the peak to estimate the variant's effect on the linkage signal. All models were adjusted for age, age², sex, and field center, and the association analyses were additionally adjusted for principal components of population substructure. By identifying the top 15% of families contributing to the linkage signals, we were able to narrow the peaks to <5 cM regions (chromosome, peak cM[cM range]: 3,120[119–123] and 17,82[80–83]). However, no common variants in these regions were significantly associated with the carotid artery trait of interest. One chromosome 3 SNP, rs13100154, was suggestive of an association with carotid artery interadventitial diameter ($P=4.2 \times 10^{-6}$). The magnitude of the linkage peak was only attenuated by ~5% after adjustment for this SNP, suggesting that neither it nor a variant in linkage disequilibrium with it was responsible for the significant linkage peak on chromosome 3. Therefore, common genetic variation does not explain the significant genetic linkage identified for carotid artery traits in these exceptionally long-lived families. We are currently sequencing these regions in families of interest in an attempt to identify rare variants with significant effects on carotid artery traits.

3151T

Assessment of anthropometric and lifestyle risk factors for the prediction of cardiovascular diseases in three generations cohort among North Indian population. *R. Kumar^{1,2}, B. Doza²*. 1) Govt. Sen. Sec. School, Aliwal, Punjab, India; 2) Dept. of Human Genetics, G.N.D.U. Amritsar, Punjab, India.

The present study has attempted to evaluate comparative three anthropometric indicators (BMI, waist circumference and WHR) and five socio-economic lifestyle factors (food habits, physical activity, exercise, smoking and alcohol) to identify the most distinctive indicators for pre-hypertension and hypertension for a specific community in Punjab. No such study has been performed in this region based on sensitivity, specificity and likelihood ratio with three generations. Among male parental generation the waist circumference and WHR for pre-hypertension; WHR and exercise for hypertension and waist circumference and BMI for pre-hypertension and hypertension among female parental generation were better indicators to assess CVD risk factors. High sensitivity suggested that waist circumference and BMI performed well in female parental generations. The results of the analysis of present data suggested waist circumference for almost all the generations with average high sensitivity upto 97% and 99% for female parental generation in the prediction of pre-hypertension and hypertension. BMI and WHR may be considered with respect to sensitivity in the second level of best predictor for both pre-hypertension and hypertension. If we compare positive likelihood ratio of the three generations with respect to all the indicators then it was observed that LR+ value of BMI has been consistently higher among all the generations as compared to waist circumference and WHR. Therefore, it is suggested that combination of waist circumference and BMI would be the better predictors to assess CVD risk.

3152F

Bilirubin and ischemic heart disease: A two-sample Mendelian randomization study. *J. Lee, C. Jeon, S. Jee*. Yonsei university, Seoul, Korea, Republic of.

Bilirubin is an endogenous antioxidant which protects cells against oxidative stress. Increased plasma levels of bilirubin have been associated with reduced risk of ischemic heart disease (IHD) in previous studies. Nonetheless, whether the associations reflect a true protective effect of bilirubin on IHD, rather than confounding or reverse causation remains unknown. Therefore, we applied the two-sample Mendelian randomization to evaluate the causal association that bilirubin levels and IHD risk among Korean populations. A total of 5 genetic variants (rs10490012, rs12993249, rs2119503, rs4149014, and rs73233620) were selected as genetic instruments for serum bilirubin levels using a community-based cohort, the Korean Genome and Epidemiology Study (KoGES) comprising 33,598 subjects. Then we evaluated their impact on ischemic heart disease using the Korean Cancer Prevention Study-II (KCPS-II) cohort. We found no significant association between serum bilirubin levels and IHD. Sensitivity analyses demonstrated consistent association, suggesting robustness of our observation. Further studies to conform the observed interaction among other ethnicities are warranted.

3153W

Cohort profile of the EPOZ study: 40 years of follow-up. *N. Terzikhan, M.A. Ikram, H.H.H. Adams.* Erasmus medical centre, The Netherlands, Zuid Holland, Netherlands.

EPOZ (Epidemiologisch Preventief Onderzoek Zoetermeer, translated: Epidemiological Preventive Study Zoetermeer), is a population-based prospective cohort study in two districts of Zoetermeer, a suburban residential community that is near The Hague in the Netherlands. The study included 10,361 persons aged 5 to 91 years at the baseline year (1975) and was originally concerned with the prevalence of various chronic diseases. All participants were included between 1975 and 1978 (78% response). Currently, it has over 40 years of follow-up and includes focus cohorts with extensive repeated measurements of cardiovascular, neurological and bone health. For the first focus cohort, families with children between 5 to 19 years of age were invited to participate in a focus study on cardiovascular risk factors. In total, 4,649 (82% response) took part of this study. From this group a random sample of 596 children was selected for annual follow-up between 1975 and 1993. In 2002, additional measurements of cardiovascular risk factors were collected. The median number of visits was 15 (range 2 to 19). For the second focus cohort, 909 subjects aged between 60 and 90 years were randomly selected to assess eligibility for MRI brain scan. Of these 909 individuals, 819 subjects were eligible. Among the eligible participants, 514 (63%) agreed to have an MRI brain scan and each individual received up to three scans during 1995-2009. Recently, we performed genome-wide genotyping in the focus cohort and are in the process of acquiring additional omics data, including gene expression in multiple tissues and methylation, for both cohorts. While EPOZ has already been part of ongoing collaborative efforts, we are eager to promote this unique and rich dataset among the international community.

3154T

Role of segregation in multiplex families for variant discovery in cardiac birth defects. *L.J. Martin¹, V. Pilipenko¹, D.W. Benson².* 1) Pediatrics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 2) Pediatrics, Medical College of Wisconsin, Milwaukee, WI.

Introduction. Cardiovascular malformations (CVM) are the most common birth defects with an incidence of 2 to 5 per 100 live births. Although CVM heritability provides strong evidence of a genetic basis, in most cases the cause remains unknown. Exome data for CVM case series has generated large numbers of potential variants but definitive evidence of causality has remained elusive except in a small percentage of cases. Given the strong heritability, we hypothesized that segregation of variants in a kindred will aid in novel gene discovery. **Methods.** Whole exome sequencing was performed on 52 individuals from 4 families ascertained by a proband with hypoplastic left heart syndrome (HLHS). Each family had at least 3 additional individuals with CVM. Analysis of exomes from probands, identified rare variants with informatics support (RVIS, minor allele frequency ≤ 0.01 and Combined Annotation Dependent Depletion score ≥ 20). Variants which did not meet these criteria were defined as non-RVIS. Family specific two point logarithm of the odds (LOD) scores were used to identify segregating with CVM using a dominant model and 80% penetrance. **Results.** A total of 702 RVIS variants in 668 genes were identified in the 4 probands. Additional analysis included RVIS sharing among probands and higher evidence of functionality, but no strong candidates were identified. However, when looking at linkage evidence in these families only 1 RVIS, rs200183228, a non-synonymous variant in *SORCS2* had a LOD score ≥ 1 (LOD = 1.3). On the other hand, there were 109 non-RVIS variants with LOD scores ≥ 1.0 , including 10 variants with LOD ≥ 1.5 , and one LOD > 2.0 (rs34053053 in *CTTN*). Since 97% of segregating variants were common, these results suggest that although a single variant is unlikely to explain CVM multiple common segregating variants contribute to CVM in these families. **Conclusions.** These results suggest that for CVM, conventional identification of candidate variants using allele frequency and predicted informatic functionality may miss a large proportion of variants which segregate with CVM. For birth defects such as CVM which exhibit strong familial transmission, yet for which single variants are not sufficient, utilization of segregation could be a powerful tool to identify novel candidates missed by current strategies.

3155F

Association between genetic variants of functional candidate genes and its interaction with smoking on the risk of coronary acute syndrome in an admixed population.

L. Franco¹, N. Gallego-Lopera¹, C. Velarde¹, C. Vergara², D. Valencia¹, K. Betancur^{3,4}, J. García⁵, N. González⁶, H. Cuervo^{3,4}, C. Saldarriaga^{3,4}, C. Martínez⁴, K. Marisancen¹, P. Parra¹, S. Carvalho¹, L. Parra¹, E. Jiménez¹, JP. Bedoya¹, I. Ortiz¹, G. Bedoya⁶, A. Valencia-Duarte^{1,2}. 1) Universidad Pontificia Bolivariana, Medellín, Colombia; 2) Division of Infection Diseases, John Hopkins University, Baltimore, United States; 3) Facultad de Medicina, Universidad Pontificia Bolivariana, Medellín, Colombia; 4) Clínica CardioVID, Medellín, Colombia; 5) Facultad de Medicina, Universidad de Antioquia, Medellín, Colombia; 6) Grupo Genética Molecular, Universidad de Antioquia, Medellín, Colombia.

Cardiovascular diseases are the main cause of mortality in the world and the acute coronary syndrome (ACS) is the most important clinical manifestation. The development of ACS events is a consequence of the interaction between multiple genes and environmental factors. This study evaluated the association between genetic variants related to lipid metabolism, inflammation and platelet aggregation and its interaction with smoking on the risk to develop ACS in young adults from Antioquia (Colombia). The study included 332 patients with diagnosis of early ACS and 432 healthy subjects. Polymorphisms in the genes *PCSK9*(rs11206510), *SORT1* (rs646776, rs599839), *LPA*(rs3798220, rs10455872), *ABCB1*(rs1045642), *ZPR1-APOA5-A4-C3-A1*(rs964184), *LDLR*(rs6511720), *APOE*(rs429358, rs7412), *APOE-APOC1*(rs4420638), *IL10*(rs1800872, rs1800896), *TNF-α*(rs1799964), *IFN-γ*(rs1861493), *ABO*(rs579459), *CXCL12*(rs501120, rs1746048), *PAI-1*(rs2227631) and *ITGB3*(rs5918) were genotyped using RTPCR. Association of each variant was evaluated using logistic regression models stratified by smoking, and adjusted by conventional risk factors and genetic ancestry. Rs10455872 in *LPA* (G allele, OR = 2.68; IC 95%=1.64–4.38) and rs599839 in *SORT1* (G allele, OR = 0.74; IC 95%=0.57–0.95) were associated with ACS independently of smoking status. Smoking significantly (interaction term p-value<0.05) modified the effect of the genetic variants *ZPR1-APOA5-A4-C3-A1*(rs964184), *LDLR*(rs6511720) and *PAI-1*(rs2227631) on ACS; so, allele G of rs964184 (OR = 1.90; CI 95% = 1.33-2.71), allele T of rs6511720 (OR = 0.41; CI 95% = 0.19-0.90) and allele A of rs2227631 (OR = 0.67; CI 95% = 0.46 - 0.98) were associated with ACS only in non-smokers, although p-values of interaction were not significant and this was not due to a lower power in the group of smokers. In conclusion, we found evidence that smoking status modifies the risk of SNPs in candidate genes for ACS, highlighting the importance of evaluating the role of environmental factors in genetic associations.

3156W

Variants affecting menopausal timing associate with CVD, stroke and blood pressure in >26,000 Finns. A. Joensuu, H. Mattsson, V. Salomaa, M. Perola, K. Auro. Department of Public Health Solutions, National Institute for Health and Welfare, Helsinki, Finland.

Background & Aims: Age at natural menopause (ANM) associates with the prevalence of several diseases, including cardiovascular diseases (CVD) and breast cancer. Whether this is due to common genetic background or to the effects of menopause itself remains unclear. Our aim was to study the combined effect of genomic variants affecting the timing of menopause by creating a polygenic risk score (PRS) and studying its association with disease endpoints in both females and males. **Materials & Methods:** Our study sample consists of 26,832 subjects aged 25-74 at baseline (14,039 females and 12,793 males) from five independent population-based Finnish FINRISK studies (baseline years 1992, 1997, 2002, 2007 and 2012) with follow-up of disease endpoints from national registers up to the year 2015. A meta-analysis by Day et al (Nat Genet. 2015) detected 54 variants associated with ANM - our study was able to use 50 of these (of which 6 genotyped and 44 imputed with a mean information quality score of 0.98) to form the PRS. The PRS was calculated by multiplying the amount of effective alleles per variant with the effect size reported in Day et al. and summing the effects of all variants together. We studied the association of the PRS with CVD endpoints, diabetes and breast cancer with logistic regression and with CVD risk factors with linear regression. We corrected the analyses for baseline study year, study area, ever smoking, socioeconomic status, as well as baseline age in risk factor analysis and age at end of follow-up in endpoint analysis. We restricted the logistic analysis to individuals with follow up to at least 40 years of age (mean follow-up end age 65.2 years). **Results:** The PRS (range -12.6 - -6.2) was validated by confirming its correlation with age at last period in postmenopausal women ($p=1.8 \times 10^{-7}$, $N=2170$). In men, the PRS did not associate with any disease endpoint, whereas in women the PRS associated ($p<0.05$) with CVD and stroke, with a lower menopausal age genetic score increasing the risk of these diseases. Breast cancer was not associated ($p=0.21$) with the PRS. Of CVD risk factors, the PRS associated only with systolic blood pressure in men ($\beta=0.43$, $p=0.028$). **Conclusions:** Our results suggest that menopausal timing and CVD risk may have a common genetic background, which is seen also in males through elevated blood pressure. Interestingly, no association was seen with breast cancer and the menopausal age PRS.

3157T

Characterizing proof-of-concept genes for familial hypercholesterolemia *in vivo* using zebrafish model systems. B.V.V. Manoj Kumar^{1,2}, A. Emmanouilidou^{1,2}, P. Ranefall³, B. von der Heyde^{1,2}, T. Kingstrom^{1,2,4}, J. Ledin^{2,4}, A. Larsson⁵, H. Brooke⁶, C. Wahlby^{2,3}, E. Ingelsson^{5,7}, M. den Hoed^{1,2}. 1) Department of Immunology, Genetics and Pathology, Uppsala University, Sweden; 2) Science for Life Laboratory, Uppsala University, Sweden; 3) Department of Information Technology, Uppsala University, Sweden; 4) Department of Organismal Biology, Uppsala University, Sweden; 5) Department of Medical Sciences, Uppsala University, Sweden; 6) Department of Public Health and Caring Sciences, Uppsala University, Sweden; 7) Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford.

Small-scale studies suggest that zebrafish larvae represent a promising model system for genetic screens in dyslipidemia and early-stage atherosclerosis, the main mechanisms underlying coronary artery disease (CAD). We expanded the phenotypic pipeline, increased the throughput, and validated the model system using a large screen for familial hypercholesterolemia proof-of-concept genes. We used a multiplexed CRISPR-Cas9 approach to target zebrafish orthologues of *APOE*, *APOB* and *LDLR*. Offspring (n=384) of founder mutants were overfed on a cholesterol supplemented diet from 5 to 9 days post fertilization (dpf). At 10 dpf, vascular atherogenic traits and body size were visualized using an automated positioning system and fluorescence microscope, and quantified using custom-written scripts. After imaging, whole-body lipid fractions and glucose levels were assessed using enzymatic assays and normalized for protein content. CRISPR-Cas9-induced mutations were identified using paired-end sequencing (2x250 bp) and annotated using Ensembl's VEP. Association analyses were performed using hierarchical linear models or zero-inflated negative binomial regression, adjusting for body size, time of day and batch. A risk score comprising the number of mutated alleles across the four zebrafish orthologues of *APOE*, *APOB* and *LDLR* for which we observed sufficient genetic variation was associated with 0.094±0.035 SD higher LDLc levels and 0.138±0.070 SD more vascular lipid deposition. At the level of individual genes, each additional mutated allele in *apoea* was associated with higher whole-body LDLc levels and with more vascular co-localization of lipids and neutrophils; each additional mutated allele in *apobb.1* resulted in more vascular co-localization of lipids and neutrophils; and each additional mutated allele in *ldlra* resulted in more vascular infiltration by lipids, and in less vascular co-localization of macrophages and neutrophils. Our results suggest that systematically characterizing candidate genes across robustly associated loci identified in large-scale genetic association studies in humans using zebrafish model systems will likely help increase our understanding of disease etiology and may help identify novel targets that can be translated into efficient medication for prevention and treatment.

3158F

Association between titin loss-of-function variants and early-onset atrial fibrillation. S. Choi¹, L. Weng^{1,2}, C. Roselli^{1,2}, H. Lin^{3,4}, C. Haggerty⁵, M. Shoemaker⁶, J. Barnard⁷, D. Arking⁸, D. Chasman^{1,9}, C. Albert¹⁰, M. Chaffin¹¹, N. Tucker^{1,2}, J. Smith¹¹, N. Gupta¹, S. Gabriel¹, E. Boerwinkle¹², T. Blackwell¹³, G. Abecasis¹³, B. Psaty^{14,15,16,17}, S. Kathiresan¹, D. Darbar¹⁸, A. Alonso¹⁹, S. Heckbert^{14,16}, M. Chung²⁰, D. Roden⁶, E. Benjamin^{3,21,22}, M. Murray²³, K. Lunetta^{3,24}, S. Lubitz^{1,2,25}, P. Ellinor^{1,2,25}, the DiscovEHR study and the NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium. 1) Broad Institute of MIT and Harvard, Cambridge, MA; 2) Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA; 3) NHLBI and Boston University's Framingham Heart Study, Framingham, MA; 4) Section of Computational Biomedicine, Department of Medicine, Boston University School of Medicine, Boston, MA; 5) Department of Imaging Science and Innovation, Geisinger, Danville, PA; 6) Departments of Medicine, Pharmacology, and Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN; 7) Departments of Quantitative Health Sciences, Cleveland Clinic, Cleveland, OH; 8) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 9) Divisions of Preventive Medicine and Genetics, Brigham and Women's Hospital & Harvard Medical School, Boston; 10) Divisions of Preventive and Cardiovascular Medicine, Brigham and Women's Hospital & Harvard Medical School, Boston; 11) Departments of Cellular and Molecular Medicine, Cleveland Clinic, Cleveland, OH; 12) Human Genetics Center, Department of Epidemiology, Human Genetics and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX; 13) Center for Statistical Genetics, Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, MI; 14) Department of Epidemiology and Cardiovascular Health Research Unit, University of Washington, Seattle, WA; 15) Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA; 16) Kaiser Permanente Washington Health Research Institute, Seattle, WA; 17) Department of Health Services, University of Washington, Seattle, WA; 18) Division of Cardiology, Department of Medicine, University of Illinois, Chicago, IL; 19) Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA; 20) Departments of Cardiovascular Medicine, Cleveland Clinic, Cleveland, OH; 21) Department of Medicine, Boston University School of Medicine, Boston, MA; 22) Department of Epidemiology, Boston University School of Public Health, Boston; 23) Genomic Medicine Institute, Geisinger, Danville, PA; 24) Department of Biostatistics, Boston University School of Public Health, Boston, MA; 25) Cardiac Arrhythmia Service, Massachusetts General Hospital, Boston, MA.

Atrial fibrillation (AF) is the most common arrhythmia affecting over 33 million individuals worldwide. Although AF is typically considered a primary electrical disorder, the mechanisms underlying AF remain incompletely understood. The National Heart Lung and Blood Institute's Trans-Omics for Precision Medicine (TOPMed) Program includes longitudinal and cohort studies in which high depth, whole genome sequencing was performed in over 18,000 individuals between 2014 and 2017. In this case-control study, we included 2,781 early onset AF cases from 9 studies and identified 4,959 controls of genetically determined European ancestry from the remaining TOPMed participants. Early onset AF was defined as AF onset <66 years of age. Rare loss of function variants in genes at AF loci and common genetic variation across the whole genome were tested for association with early onset AF. Replication of common and rare variant associations were conducted with 346,546 unrelated UK Biobank participants of European ancestry and 42,783 participants in the MyCode study, respectively. Among early-onset AF cases, 72.1% were male; the mean age of AF onset was 48.7±10.2 years. Samples were sequenced at a mean depth of 37.8 fold and mean genome coverage of 99.1%. In an analysis of common variants, we observed an association with 5 known loci and 1 recently identified locus (*NAV2*). This *NAV2* locus was replicated in the UK Biobank (OR, 1.11; 95% CI, 1.07-1.15; $P=9.70 \times 10^{-10}$). In an analysis of rare variants, we found at least one loss-of-function variant in *TTN*, the gene encoding the sarcomeric protein titin, in 2.1% of cases and in 1.1% in controls. The proportion of individuals with early-onset AF who carried a loss-of-function variant in *TTN* was higher in those with an earlier age of AF onset (P value for trend 4.92×10^{-4}); 6.5% of individuals with AF onset before age 30 carried a *TTN* loss-of-function variant (OR, 5.94; 95% CI, 2.64-13.35; $P=1.65 \times 10^{-5}$). The association between *TTN* loss of function variants and early onset AF was replicated in MyCode, comprised of 1,582 early-onset AF cases and 41,200 controls (OR, 2.16; 95% CI, 1.19-3.92; $P=0.01$). In sum, we discovered that *TTN* is a disease susceptibility gene for early onset AF. Our findings expand the role for *TTN* in cardiovascular disease.

3159W

Gene-based genome-wide association analysis implicates novel genes involved in venous thromboembolism in African, Asian, and European populations. W. Hernandez¹, M. Pividori², A. Barreira², R. Bonazzola², J. Liao³, H. Im². 1) Section of Cardiology, Section of Genetic Medicine, University of Chicago, Chicago, IL; 2) Section of Genetic Medicine, University of Chicago, Chicago, IL; 3) Section of Cardiology, University of Chicago, Chicago, IL.

Venous thromboembolism (VTE) is a chronic multifactorial disease encompassing deep vein thrombosis (DVT) and pulmonary embolism (PE) and the third most common life-threatening cardiovascular condition world-wide. Epidemiological studies have shown that the prevalence of VTE varies significantly across ethnic/racial groups, yet the genetic basis for these differences remains largely undefined. Our aim was to integrate genomic and transcriptomic data to elucidate molecular mechanisms and examine the genetic heterogeneity underlying VTE risk among different ethnic populations. We applied a novel gene-based statistical method, MultiXcan, to VTE cases and controls of African (cases=141; controls=6,450), Asian (cases=127; controls=8,356), and European (cases=8,053; controls=318,007) ancestry populations from the UK Biobank to identify genes associated with VTE risk (Bonferroni correction was 2.9×10^{-6}). Among our European cohort — in addition to the well-established associations *F5*, *F11*, *SLC19A2* ($p \leq 5.00E-32$), we found *EDEM2* and *PROCR* to be associated with VTE risk ($p \leq 5.57E-17$). Loci within or near this region have been previously associated with circulating Protein C concentration, one of the most important endogenous anticoagulants. We also found *EXOSC4*, part of the exosome complex, to be associated with VTE risk in African and European populations ($p=2.58E-07$ and $p=2.57E-09$, respectively). The exosome complex recently has emerged as an important mediator in vascular remodeling and consequently vascular diseases. The most significant finding among our Asian cohort was *ELMO1* ($p=7.56E-12$), although a novel finding in humans, *ELMO1* deficient mice have been shown to have enhanced thrombus formation. By harnessing the large scale data from the UK Biobank and the latest advances in bioinformatics and genomics our study provides new molecular insight into the underlying mechanism that may regulate VTE susceptibility in diverse populations and highlights the importance of conducting ethnic-specific research.

3160T

Trans-ethnic genetic analysis of ischemic stroke in the VA Million Veteran Program. J.E. Huffman¹, D. Posner^{2,3}, J. Honerlaw², J.L. Vassy^{2,4}, Y.-L. Ho², S.M. Damrauer^{5,6,7}, K.-M. Chang^{5,6}, T.L. Assimes^{8,9}, P.S. Tsao^{8,9}, J.M. Gaziano^{2,4}, S. Seshadri^{10,11}, D.R. Gagnon^{2,3}, C.J. O'Donnell^{1,4,12}, K. Cho^{2,4}, P.W.F. Wilson^{13,14} on behalf of the VA Million Veteran Program. 1) Center for Population Genomics, MAVERIC, VA Boston Healthcare System, Boston, MA; 2) Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC), VA Cooperative Studies Program, VA Boston Healthcare System, Boston, MA; 3) Department of Biostatistics, Boston University School of Public Health, Boston, MA; 4) Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 5) Corporal Michael J. Crescenz VA Medical Center, Philadelphia, PA; 6) Perlman School of Medicine, University of Pennsylvania, Philadelphia, PA; 7) Hospital of the University of Pennsylvania, Philadelphia, PA; 8) Department of Medicine, Stanford University School of Medicine, Stanford, CA; 9) VA Palo Alto Health Care System, Palo Alto, CA; 10) Framingham Heart Study, Framingham, MA; 11) Glenn Biggs Institute for Alzheimer's and Neurodegenerative Diseases, University of Texas Health Sciences Center, San Antonio, TX; 12) Cardiology Section, Department of Medicine, VA Boston Healthcare System, Boston, MA; 13) Atlanta VA Medical Center, Decatur, GA; 14) Emory University School of Medicine, Atlanta, GA.

Stroke is a leading cause of mortality and morbidity worldwide. Stroke subtypes have different risk factors and clinical presentations, with ischemic stroke (IS) accounting for 85% of all events. Several large studies have been undertaken to identify genetic predictors of IS, but the identified loci do not explain a large proportion of the genetic risk. Additionally, most have been undertaken only in Europeans (EUR) or have not reported ancestry-stratified results for non-EUR samples. We aimed to discover novel genetic loci for IS, within and across ancestral groups, in the Million Veteran Program (MVP) cohort. Prevalent IS was ascertained within MVP using both the Tirschwell algorithm and more sensitive algorithm developed within MVP. Preliminary analyses focused on EUR and African American (AFR) veterans with 20,686 Tirschwell cases [5,038 AFR; 15,648 EUR] & 6,993 algorithm cases [1,784 AFR; 5,209 EUR] identified in the initial ~350,000 individuals genotyped. GWAS was performed using 1000G imputed dosages, within each sex and ancestry stratum, adjusting for age and first 10 ancestry-specific PCs, then combined in an inverse-weighted meta-analysis. No loci reached genome-wide significance ($p < 5E-08$) in initial MVP analyses, regardless of phenotype algorithm or ancestry stratum. Of the 18 SNPs reported by the recent MEGASTROKE trans-ethnic meta-analysis for cases (N=60,341) of any ischemic stroke (AIS), *PITX2*, *FOXF2*, *CDK6*, *MMP12*, and *ZCCHC14* replicated in MVP using the Tirschwell criteria, and *MMP12* and *ZCCHC14* using the MVP algorithm (Bonferroni threshold for 20 loci = $p < 0.0025$). Of the 8 AIS loci from the MEGASTROKE EUR-only analysis (N=34,217 cases), *FOXF2* and *HDAC9-TWIST1* replicated in MVP (EUR-only). Most loci that did not replicate did achieve a p -value < 0.05 and had MVP odds ratios in the expected directions (with overlapping 95% CI). Failure to replicate may be due to the smaller number of IS cases within MVP and the differential population make-up of the respective trans-ethnic analyses, as 29% of MEGASTROKE AIS cases are East Asian. Further analyses are underway using an additional 110K samples from MVP and including Hispanic and Asian American individuals. As well, full summary results from the MEGASTROKE analyses have been obtained to complete the largest IS GWAS to date, a mega-analysis of MVP + MEGASTROKE, in time for presentation at this meeting.

3161F

Genome-wide meta-analysis in >30,000 Caucasian subjects identify 7 novel loci for glycine.

Q. Jia^{1,2}, Y. Han^{1,2}, P. Huang^{1,2}, N.C. Woodward^{1,2}, J. Gukasyan^{1,2}, J. Kettunen^{3,4}, M. Ala-Korpela^{3,5,6,7,8,9}, O. Anufrieva³, Q. Wang^{3,5}, M. Perola^{4,10,11}, V. Salomaa⁴, O. Raitakari^{12,13}, T. Lehtimäki¹⁴, J. Viikari¹⁵, M.-R. Järvelin^{3,16,17,18}, M. Boehnke¹⁹, M. Laakso²⁰, K.L. Mohlke²¹, Z. Wang²², W.H.W. Tang^{22,23}, S.L. Hazen^{22,23}, J.A. Hartiala^{1,2}, H. Allayee^{1,2}. 1) Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA; 2) Department of Biochemistry & Molecular Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA; 3) Computational Medicine, Faculty of Medicine, University of Oulu and Biocenter Oulu, Oulu, Finland; 4) National Institute for Health and Welfare, Helsinki, Finland; 5) Systems Epidemiology, Baker Heart and Diabetes Institute, Melbourne, Victoria, Australia; 6) NMR Metabolomics Laboratory, School of Pharmacy, University of Eastern Finland, Kuopio, Finland; 7) Population Health Science, Bristol Medical School, University of Bristol, Bristol, UK; 8) Medical Research Council Integrative Epidemiology Unit at the University of Bristol, Bristol, UK; 9) Department of Epidemiology and Preventive Medicine, School of Public Health and Preventive Medicine, Faculty of Medicine, Nursing and Health Sciences, The Alfred Hospital, Monash University, Melbourne, Victoria, Australia; 10) Estonian Genome Center, University of Tartu, Tartu, Estonia; 11) Institute for Molecular Medicine (FIMM), University of Helsinki, Helsinki, Finland; 12) Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland; 13) Department of Clinical Physiology, Turku, Finland; 14) Department of Clinical Chemistry, Fimlab Laboratories and Finnish Cardiovascular Research Center-Tampere, Faculty of Medicine and Life Sciences, University of Tampere; 15) Department of Medicine, University of Turku, Turku, Finland; 16) Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, London, UK; 17) Center for Life Course and Systems Epidemiology, University of Oulu, Oulu, Finland; 18) Unit of Primary Care, Oulu University Hospital, Oulu, Finland; 19) Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan; 20) University of Eastern Finland, School of Medicine, Kuopio, Finland; 21) Department of Genetics, University of North Carolina, Chapel Hill, NC; 22) Department of Cellular & Molecular Medicine, Cleveland Clinic, Cleveland, OH; 23) Department of Cardiovascular Medicine, Cleveland Clinic, Cleveland, OH.

Abstract We previously carried out a large-scale genetic analysis of the choline-derived metabolite betaine, whose levels are associated with development of coronary artery disease (CAD). These efforts identified a variant of *CPS1* (carbamoyl-phosphate synthase 1), the rate limiting enzyme in the urea cycle, as being strongly associated with decreased levels of betaine, increased glycine levels, and decreased risk of CAD in women but not men. Based on these data, we sought to determine whether plasma glycine levels represent a novel causal biomarker for risk of CAD using a genetics approach. We first carried out a meta-analysis of GWAS summary level data for plasma glycine levels in up to 30,118 European ancestry subjects and identified five previously known (*CPS1*, *ALDH1L1*-*SLC41A3*, *PPP1R3B*-*LOC157273*, *GLDC*, *GCSH-PKD1L2*) and seven novel (*ACADM*, *PHGDH*, *COX18-ADAMTS3*, *PSPH*, *TRIB1*, *PTPRD*, *ABO*) genome-wide significant ($p < 5 \times 10^{-8}$) loci. Although stratified meta-analyses for glycine levels in men ($n \sim 19,000$) and women ($n \sim 11,000$) separately did not reveal any additional loci beyond those in the combined analysis, the significance level at the *PHGDH* and *PTPRD* loci was one order of magnitude stronger in females despite a smaller sample size ($p < 10^{-6}$ in females; $p < 10^{-5}$ in males, respectively). Of the seven novel loci, glycine-raising variants at the *PSPH*, *TRIB1* and *ABO* loci were also significantly (Bonferroni-adjusted $p < 7.1 \times 10^{-3}$) associated with decreased risk of CAD based on GWAS summary level data in a dataset of >500,000 cases and controls from UK Biobank and the CARDIoGRAM-C4D Consortium. A combined analysis further demonstrated that the joint effects of the seven glycine-raising alleles were modestly but significantly associated with decreased risk of CAD (OR=0.98, 95% CI 0.98-0.99, $p = 2.1 \times 10^{-12}$). In summary, our analysis identified seven novel loci for plasma glycine levels, but it was not possible to determine whether the association of some of these loci with CAD is causally mediated through glycine since several of these loci have previously been associated with other known CAD risk factors, such as plasma LDL and total cholesterol levels. Thus, additional studies will be required to determine if there is a causal link between glycine and risk of CAD, and whether any of the novel loci exhibit sex-specific associations with CAD outcomes.

3162W

Genetic determinants of plasma bioactive lipids, potential novel biomarkers of cardiovascular disease.

K. McGurk¹, A. Nicolaou², B. Keavney¹. 1) Division of Cardiovascular Sciences, University of Manchester, Manchester, United Kingdom; 2) Division of Pharmacy & Optometry, University of Manchester, Manchester, United Kingdom.

Bioactive lipids have key roles in cardiovascular disease (CVD). Although some lipid classes have been associated with CVD, genetics studies to date have focused on few of the many species now measurable. Signalling lipids of the eicosanoid, endocannabinoid and ceramide classes have potent roles in inflammation, vasodilation and cell death. Heritability and family-based GWAS analyses may identify particularly heritable lipid biomarkers for prioritisation in large scale Mendelian randomisation studies to determine causal relationships. We analysed 175 eicosanoid, endocannabinoid and sphingolipid species in 204 plasma samples from 31 British Caucasian families, using liquid chromatography coupled to tandem mass spectrometry. Baseline heritability estimates (QTDT and GCTA) were determined and GWAS (FaST-LMM and GCTA) was performed. Certain species were found significantly heritable, particularly from the endocannabinoid and ceramide classes (20%-70%; corrected $p < 0.05$). DHET species, formed in a pathway involving CYP450 enzymes, were the most heritable eicosanoid species measured (47% - 58%, corrected $p = < 0.05$). Baseline GWAS analyses identified suggestive associations between lipid concentrations and SNPs in genes known to be involved in the metabolism of these lipids (e.g. CYP450 enzymes). A top associated SNP with an endocannabinoid congener was also identified in previously published GWAS for coronary artery disease, indicating a novel causal role for these bioactive lipids. Many ceramide species correlated with cardiovascular phenotypes (e.g. hypertension) previously measured in the cohort. Although a small number of eicosanoids were also found to be significantly heritable, the baseline associations at GWAS were not as promising as those for endocannabinoids and ceramides. GWAS analyses of a further 800 samples, from 170 families, will be presented at the meeting. We demonstrate for the first time estimates of heritability and GWAS analyses for this array of bioactive lipids. Although found at pg/ml concentrations much lower than cholesterol, many of the heritability estimates are similar to plasma genetic biomarkers of CVD (e.g. cholesterol h^2 30%-50%). The results from the extended cohort will be used to identify causal metabolic pathways, novel diagnostics and drug targets for CVD intervention.

3163T

Genome-wide association study of the QRS complex. K. Norland¹, G. Sveinbjornsson¹, R.B. Thorolfsdottir¹, O.B. Davidsson¹, V. Tragante¹, D.O. Amar^{1,2,3}, R. Danielsen³, S.S. Stephensen^{4,5}, G. Oskarsson⁶, G. Thorgeirsson^{1,2}, J.T. Sverrisson⁷, E.L. Sigurdsson^{8,9}, K. Andersen^{2,3}, A. Helgadottir¹, S. Gre-tarsdottir¹, P. Sulem¹, U. Thorsteinsdottir^{1,2}, H. Holm¹, D.F. Gudbjartsson^{1,10}, K. Stefansson^{1,2}. 1) deCode genetics/Amgen Inc., Reykjavik, Iceland; 2) Faculty of Medicine, University of Iceland, Reykjavik, Iceland; 3) Department of Medicine, Landspítali National University Hospital of Iceland, Reykjavik, Iceland; 4) Department of Pediatric Cardiology, Skane University Hospital, Lund, Sweden; 5) Department of Clinical Sciences Lund, Clinical Physiology, Lund University, Lund, Sweden; 6) Childrens Hospital, Landspítali National University Hospital of Iceland, Reykjavik, Iceland; 7) Department of Internal Medicine, Akureyri Hospital, Akureyri, Iceland; 8) Department of Family Medicine, University of Iceland, Reykjavik, Iceland; 9) Department of Development, Primary Health Care of the Capital Area, Reykjavik, Iceland; 10) School of Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland.

Features of the electrocardiogram's QRS complex, reflecting ventricular depolarization, associate with adverse outcomes. Previous genome-wide association studies have tested four QRS measures and found variants at 58 loci, mostly common. We tested 32.5 million variants for association with ten QRS measures in 12 ECG leads, using 434,000 electrocardiograms from 81,303 Icelanders. We found 190 associations at 130 loci, 123 at 86 new QRS loci, of which 19 are associations with rare or low-frequency coding variants with large effects. Assessment of genes highly expressed in heart identified additional 13 rare coding variants at 10 QRS loci, four of which are new loci. We tested the QRS variants for association with other cardiovascular phenotypes and found 23 new disease associations. We demonstrate the substantial advantage of analysing individual components of the QRS complex in conjunction with other cardiovascular phenotype information to further the understanding of the genetic basis of cardiac conduction and disease.

3164F

Genomic loci associated with lipid traits in Latinos. I. Kullo¹, G. Schaibi², K. Ding¹, E. DeFilippis², M. de Andrade¹, L. Mandarino¹, S. Thibodeau¹. 1) Mayo Clinic Minnesota, Rochester, MN; 2) Mayo Clinic Arizona, Scottsdale, AZ.

Background. To identify novel genetic loci and replicate previously identified loci influencing inter-individual variation in lipid traits in Latinos, we conducted a genome-wide association study (GWAS) in individuals participating in the Sangre Por Salud biobank. **Methods.** Participants (n=999, age 38.9 ± 12.3 yrs, 72% female) underwent genotyping of ~1.2 million SNPs on the Illumina Infinium Multiethnic array. Total and HDL cholesterol and triglycerides were measured using standard methods and LDL cholesterol was estimated by the Friedewald formula. Association analyses were performed adjusting for age, sex, BMI, and three principal components. **Results.** Mean levels of total, HDL and LDL cholesterol were: 182.2 ± 36.0, 50.2 ± 14.1, and 105.3 ± 29.2 (mg/dL). Mean triglycerides levels were 138.1 ± 110.0 (mg/dL). Several loci were associated with lipid traits at a genome-wide significant level (Table). Three novel loci, including intergenic variants near *PCDH7* (p=2.5E-07) and *ZFAT* (p=5.6E-08), were associated with log-transformed triglycerides. *PCDH7* has been previously reported to modify the effect of waist-to-hip ratio on total cholesterol (Surakka et al., 2015, *PLoS Genetics*), and *ZFAT* was associated with hip circumference adjusted for BMI (Shungin et al. 2015, *Nature*). Among 95 loci identified previously as being associated with lipid traits in a large meta-analysis (n~100,000), ~80% were replicated at a FDR < 0.05. **Conclusion.** We identified three novel loci associated with triglyceride levels in a Latino cohort. The majority of loci previously associated with lipid traits in European ancestry individuals were replicated at FDR < 0.05.

Trait	Chr.Pos	MAF	Beta	SE	Class	Gene	P
LDL cholesterol	1:109815252 (I)	0.24	8.29	1.52	intron	SORT1	6.00E-08
HDL cholesterol	11:116740482 (I)	0.08	-7.36	1.07	5' upstream	APOA1	9.30E-12
HDL cholesterol	16:56989590 (G)	0.31	-3.98	0.64	intergenic	CETP	9.80E-10
Triglycerides	11:116648917 (G)	0.33	-0.13	0.02	3' downstream	APOA1	2.40E-08
Triglycerides	4:30325531 (I)	0.08	-0.21	0.04	intergenic	PCDH7	2.50E-07
Triglycerides	8:135409420 (G)	0.26	0.14	0.03	intergenic	ZFAT	5.60E-08
Triglycerides	22:49539956 (I)	0.26	-0.13	0.03	intergenic	C22orf34	3.80E-07

3165W

A comprehensive evaluation of the genetic architecture of sudden cardiac arrest. *R.N. Mitchell¹, F.N. Ashar¹, D.E. Arking¹, N. Sotoodehnia², SCD working group of the CHARGE Consortium.* 1) Institute of Genetic Medicine, Johns Hopkins, Baltimore, MD; 2) Cardiovascular Health Research Unit, Division of Cardiology, Departments of Medicine and Epidemiology, University of Washington, Seattle, WA.

Background. Sudden cardiac arrest (SCA) accounts for 10% of adult mortality in Western populations. While several risk factors are observationally associated with SCA, the genetic architecture of SCA in the general population remains unknown. Furthermore, understanding which risk factors are causal may help target prevention strategies. **Methods.** We carried out a large genome-wide association study (GWAS) for SCA ($n=3,939$ cases, 25,989 non-cases) to examine common variation genome-wide and in candidate arrhythmia genes. We also exploited Mendelian randomization methods using cross-trait multi-variant genetic risk score associations (GRSA) to assess causal relationships of 18 risk factors with SCA. **Results.** No variants were associated with SCA at genome-wide significance, nor were common variants in candidate arrhythmia genes associated with SCA at nominal significance. Using cross-trait GRSA, we established genetic correlation between SCA and (1) coronary artery disease (CAD) and traditional CAD risk factors (blood pressure, lipids, and diabetes), (2) height and BMI, and (3) electrical instability traits (QT and atrial fibrillation), suggesting etiologic roles for these traits in SCA risk. **Conclusions.** Our findings show that a comprehensive approach to the genetic architecture of SCA can shed light on the determinants of a complex life-threatening condition with multiple influencing factors in the general population. The results of this genetic analysis, both positive and negative findings, have implications for evaluating the genetic architecture of patients with a family history of SCA, and for efforts to prevent SCA in high-risk populations and the general community.

3166T

Metabolomic consequences of PCSK9 inhibition compared with statin therapy. *P. Würtz.* Nightingale Health Ltd, Helsinki, Finland.

Abstract—Both statins and PCSK9 inhibitors lower blood low-density lipoprotein cholesterol (LDL-C) levels to reduce risk of cardiovascular events. We performed detailed metabolic profiling of a large randomized statin trial, and compared the results with the effects of genetic inhibition of PCSK9, acting as a naturally occurring trial. Altogether 228 circulating metabolic measures were quantified by nuclear magnetic resonance spectroscopy, including lipoprotein subclass concentrations and their lipid composition, fatty acids, and amino acids, for 5,359 individuals in the PROSPER trial. The corresponding metabolic measures were analyzed in eight cohorts ($N=72,185$) using *PCSK9* rs11591147 to mimic the therapeutic effects of PCSK9 inhibitors. Scaled to an equivalent lowering of LDL-C, the effects of genetic inhibition of PCSK9 on 228 metabolic markers were generally consistent with those of statin therapy ($R^2=0.88$). However, discrepancies were observed for very-low-density lipoprotein (VLDL) lipid measures. For instance, genetic inhibition of PCSK9 had weaker effects on lowering of VLDL-cholesterol compared with statin therapy (54% vs. 77% reduction, relative to the lowering effect on LDL-C). Genetic inhibition of PCSK9 showed no significant effects on amino acids, ketones, or a marker of inflammation, whereas statin treatment weakly lowered GlycA levels. Genetic inhibition of PCSK9 had similar metabolic effects as statin therapy on detailed lipid and metabolite profiles. However, PCSK9 inhibitors are predicted to have weaker effects on VLDL lipids compared with statins for an equivalent lowering of LDL-C, which potentially translate into smaller reductions in cardiovascular disease risk. **Keywords-** metabolic profiling, cholesterol lowering, Mendelian randomization, lipoproteins, drug development.

3167F

Towards a multi-modal health assessment: Gaining insights into cardio-metabolic conditions. I. Shomorony¹, E. Cirulli¹, L. Huang¹, W. Li¹, C. Swisher¹, N. Schenker-Ahmed¹, M. Hicks¹, L. Napier¹, C. Hou¹, J. Yu¹, R. Heister¹, T. Caskey¹, D. Karow¹, J.C. Venter², E. Kirkness¹, N. Shah¹. 1) Human Longevity, Inc, San Diego, CA; 2) J. Craig Venter Institute, San Diego, CA.

Despite the enormous U.S. healthcare spending of \$3.3 trillion in 2016, one in three individuals aged 50-75 die prematurely from major age-related chronic diseases. While cardio-metabolic factors remain one of the major risk contributors, the molecular underpinning of their effects is not yet fully understood. We introduce a unique platform of deep quantitative multimodal phenotyping that seeks to provide a comprehensive, predictive, preventative, and personalized assessment of an individual's health status. The offered multimodal assays include whole genome sequencing, advanced imaging, metagenomic sequencing, metabolome, electrocardiogram, and clinical labs. This platform provides critical data not only to identify previously undiagnosed disease states but also to identify early disease biomarkers. We present an analysis of the multimodal datasets that were collected for 1,253 seemingly healthy adults (median age 53; 63% male). To the best of our knowledge, this is the largest cohort with such a wide range of data modalities analyzed to date. To study disease risk and patterns of comorbidity in this high-dimensional dataset, we took an "unsupervised" approach and analyzed the cross-modality association network. More precisely, we identified pairs of variables across the different data modalities with statistically significant associations and used these pairs to construct an association network. The analysis of the cluster structure in this network reveals signatures and biomarkers for specific traits and conditions. In particular, the set of markers making up the "cardio-metabolic" cluster allowed us to identify individuals sharing similar molecular signatures and determine risk groups within the cardiac- and metabolic- disease conditions. This shows how the underlying structures in the association network can be used to define a clinically relevant health profile, thus providing a comprehensive view into an individual's health status.

3168W

Untargeted metabolomics analysis identifies and replicates serum metabolites associated with kidney function among Bogalusa Heart Study and Multi-Ethnic Study of Atherosclerosis participants. J.L. Nierenberg¹, J. He¹, C. Li², X. Gu^{1,3}, X. Mi¹, S. Li¹, L.A. Bazzano¹, W. Chen¹, J. Kinchen⁴, C.M. Rebholz^{5,6}, J. Coresh^{5,6}, A.S. Levey⁷, L.A. Inker⁸, M.G. Shlipak⁸, T.N. Kelly¹. 1) Department of Epidemiology, Tulane University, New Orleans, LA, USA; 2) Department of Epidemiology & Biostatistics, College of Public Health, University of Georgia, Athens, GA, USA; 3) Department of Epidemiology, Fuwai Hospital, National Center for Cardiovascular Diseases, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China; 4) Metabolon, Inc., Durham, NC, USA; 5) Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA; 6) Welch Center for Prevention, Epidemiology, and Clinical Research, Johns Hopkins University, Baltimore, MD, USA; 7) Division of Nephrology, Tufts Medical Center, Boston, MA, USA; 8) Department of General Internal Medicine, University of California San Francisco, San Francisco, CA, USA.

Background: Biological pathways related to reduced kidney function remain largely unknown. **Methods:** Untargeted ultrahigh performance liquid chromatography-tandem mass spectroscopy was utilized to detect and quantify serum metabolites. We used multivariate adjusted linear regression to examine associations between single metabolites and estimated glomerular filtration rate (eGFR, calculated with the creatinine based CKD-EPI equation) among 1,243 Bogalusa Heart Study participants (median eGFR: 94.4, 5th-95th percentile: 66.0-119.6 mL/min/1.73 m²). Replication (determined by statistical significance and consistent effect direction) was tested using gold standard measured glomerular filtration rate (mGFR) among 260 Multi-Ethnic Study of Atherosclerosis participants (median mGFR: 72.0, 5th-95th percentile: 43.5-105.0 mL/min/1.73 m²). All analyses used Bonferroni-corrected alpha-thresholds. **Results:** Twelve associated novel metabolites were comprised of carbohydrate (advanced glycation end-product), cofactor and vitamin (ascorbate & aldarate, nicotinate & nicotinamide, tocopherol & vitamin A), lipid (mevalonate and fatty acid & BCAA), nucleotide (pyrimidine metabolism, thymine containing), and xenobiotic (drug and bacterial/fungal) sub-pathways, along with two without fully characterized pathways (**Table**). **Discussion:** This analysis identified novel metabolites associated with kidney function, providing insight into pathway alterations in this complex phenotype.

Metabolite	BHS		MESA	
	Beta (SE)	P	Beta (SE)	P
Carbohydrates				
N6-carboxymethyllysine	-2.95 (0.56)	1.50×10 ⁻³⁸	-7.41 (1.49)	8.63×10 ⁻⁷
Cofactors and vitamins				
quilonate*	-5.27 (0.64)	2.60×10 ⁻⁴²	-6.06 (1.53)	2.38×10 ⁻¹⁴
quinolinate	-12.68 (1.45)	1.10×10 ⁻³⁰	-16.4 (3.52)	1.17×10 ⁻¹⁰
gamma-CEHC glucuronide*	-11.86 (1.09)	2.20×10 ⁻²⁸	-0.19 (1.20)	9.04×10 ⁻⁹
retinol (Vitamin A)	-5.48 (0.46)	5.90×10 ⁻¹⁸	-6.34 (0.94)	5.17×10 ⁻⁶
Lipids				
3-hydroxy-3-methylglutarate	-2.94 (0.60)	1.00×10 ⁻⁴²	-3.26 (1.55)	7.36×10 ⁻¹⁰
methylmalonate	-2.4 (0.39)	1.40×10 ⁻⁷	-1.68 (0.48)	1.25×10 ⁻⁶
Nucleotides				
3-aminoisobutyrate	-2.14 (0.19)	5.70×10 ⁻¹⁶	-2.70 (0.61)	9.29×10 ⁻⁵
Xenobiotics				
hydroquinone sulfate	-5.24 (0.93)	2.60×10 ⁻²⁷	-1.52 (1.09)	1.54×10 ⁻⁵
N-methylpipercolate	-4.01 (0.50)	5.80×10 ⁻²⁹	-1.25 (0.56)	1.34×10 ⁻⁴
Partially characterized				
glycine conjugate of	-3.37 (0.56)	9.00×10 ⁻⁵⁰	-3.69 (1.09)	1.60×10 ⁻⁸
C10H12O2*				
glycine conjugate of	-4.25 (0.32)	2.80×10 ⁻³⁰	-5.59 (1.11)	4.55×10 ⁻⁶
C10H14O2 (1)*				

TableBHS=Bogalusa Heart Study, MESA=Multi-Ethnic Study of Atherosclerosis, SE=standard error. .

3169T

The Hutchinson-Gilford progeria syndrome mutation, *LMNA* c.1824C>T, is present as a somatic mutation in chronic kidney disease. *H.T. Helgadottir¹, N. Viceconte¹, A. Witasp², E. Wallén Arzt¹, A. Sola Carvajal¹, G. Revéchon¹, A. Thorell³, D. McGuinness⁴, P.G. Shields¹, A. Wernerson², P. Stenvinkel³, M. Eriksson¹.* 1) Karolinska Institutet, Department of Biosciences and Nutrition, Center for Innovative Medicine, Huddinge, Sweden; 2) Karolinska Institutet, Department of Clinical Science, Intervention and Technology, Division of Renal Medicine, Stockholm, Sweden; 3) Karolinska Institutet, Department of Clinical Sciences, Danderyds Hospital, Department of Surgery, Ersta Hospital, Stockholm, Sweden; 4) University of Glasgow, Institute of Cancer Sciences, Glasgow, United Kingdom.

Premature vascular aging is a hallmark of chronic kidney disease (CKD). The uremic phenotype include arteriosclerosis, loss of vascular smooth muscle cells, calcification and increased arterial stiffness. A similar vascular phenotype is also observed in children with the very rare premature aging disease Hutchinson-Gilford progeria syndrome (HGPS). While the risk factors for vascular disease in CKD include hypertension, dyslipidemia, hyperphosphatemia and diabetes, HGPS is almost exclusively caused by a single-nucleotide mutation in the *LMNA* gene, c.1824C>T. The mutation leads to increased production of a truncated lamin A protein (named progerin), resulting in dysregulation of cell functions. Progerin has also been found at low levels in fibroblasts, skin and arteries from non-progeroid individuals, but its underlying mechanism is unknown. Here we analyzed the presence of progerin expression in epigastric arteries obtained from CKD stage 5 patients (GFR <15 ml/min) and healthy controls, and its contribution to premature vascular aging. Immunofluorescence of uremic epigastric arteries showed progerin positive cells in 70% of the arteries (46±2 yrs, n=40). Progerin positive cells were almost exclusively detected in the tunica media, up to 7.4% of cells. The number of progerin positive cells correlated with number of years with the disease ($P=0.03$, $r = 0.34$). The arteries with higher frequency of progerin positive cells had an increased activation of DNA repair ($P=0.005$, $r = 0.63$) and increased senescence that correlated with increased vascular calcification. Somatic mutation analysis of DNA extracted from progerin stained sections from CKD arteries (n=27), using digital droplet PCR, identified the *LMNA* c.1824C>T mutation in 55% of the sections. The same somatic mutation was also detected as a rare event in DNA from blood of all CKD patients tested (n=26), and was more common than in the healthy controls ($P=0.04$). Our results suggest that rare somatic mutations may become a risk when increased tissue regeneration is induced, and provide novel mechanistic insight to the underlying phenotype of premature vascular aging.

3170F

Genome-wide association study of angioedema induced by agents acting on the angiotensin system. *M. Wadelius¹, E. Rye Rasmussen², E.V. Baranova³, N. Eriksson^{1,4}, M. Karawajczyk⁵, C. Johansson^{1,6}, A. Veluchamy⁷, M. Cavalli⁸, L. Nordang⁹, N. Bylund⁹, G. Islander¹⁰, S. Hugosson¹¹, I. Terreehorst¹², F.W. Asselbergs^{13,14,15}, P. Norling¹⁶, H-E. Johansson¹⁷, A-C. Syvänen¹⁸, P.K.E. Magnusson¹⁹, Q-Y. Yue²⁰, C. Wadelius⁸, A. Bygum²¹, C. von Buchwald², A. Alfrevic²², A.H. Maitland-van der Zee²³, C.N.A. Palmer¹, P. Hallberg¹, PREDICTION-ADR.* 1) Medical Sciences, Clin Pharmacol and SciLifeLab, Uppsala University, Uppsala, Sweden; 2) Otorhinolaryngology, Head & Neck Surgery and Audiology, Rigshospitalet, University of Copenhagen, Denmark; 3) Pharmaceutical Sciences, Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht University, Utrecht, the Netherlands; 4) Uppsala Clinical Research Center, Uppsala, Sweden; 5) Medical Sciences, Clinical Chemistry, Uppsala University, Uppsala, Sweden; 6) Q-Linea AB, Uppsala, Sweden; 7) Medical Research Institute, College of Medicine Dentistry and Nursing, University of Dundee, Ninewells Hospital Dundee, UK; 8) Immunology, Genetics and Pathology, Medical Genetics and Genomics and SciLifeLab, Uppsala University, Uppsala, Sweden; 9) Surgical Sciences, Otorhinolaryngology, and Head & Neck Surgery, Uppsala University, Uppsala, Sweden; 10) Intensive and Perioperative Care, Skåne University Hospital, Lund, Sweden; 11) Otorhinolaryngology, Årebro University Hospital and Årebro University, Årebro, Sweden; 12) Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands; 13) Cardiology, Division Heart and Lungs, University Medical Center Utrecht, Utrecht, the Netherlands; 14) Durrer Center for Cardiogenetic Research, ICIN-Netherlands Heart Institute, Utrecht, the Netherlands; 15) Cardiovascular Science, Faculty of Population Health Sciences, University College London, London, UK; 16) Sickla Health Centre, Nacka, Sweden; 17) Public Health and Caring Sciences/Geriatrics, Uppsala University, Uppsala, Sweden; 18) Medical Sciences, Molecular Medicine and SciLifeLab, Uppsala University, Uppsala, Sweden; 19) Swedish Twin Registry, Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 20) Medical Products Agency, Uppsala, Sweden; 21) Dermatology and Allergy Centre, Odense University Hospital, Odense, Denmark; 22) Molecular & Clinical Pharmacology, University of Liverpool, Liverpool, UK; 23) Respiratory Diseases, Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands.

Background: Angiotensin-converting enzyme inhibitor (ACEi) and angiotensin receptor blocker (ARB) treatment is commonly used in hypertension, diabetic nephropathy and heart failure. A rare but potentially severe adverse reaction is angioedema, which is a swelling in the skin or mucosa. When located in the head and neck region it can be life-threatening due to asphyxiation. The objective of this genome-wide association study (GWAS) was to test whether any common genetic variant predisposes to angioedema induced by an ACEi or ARB. **Method:** 173 cases recruited in Sweden were genotyped on Illumina HumanOmni2.5 arrays, and compared with 4890 population controls genotyped on Illumina HumanOmniExpress 700K arrays. The merged genotyped dataset was imputed up to 7.6 million SNPs, and corrected for principal components 1-4. The genome-wide significance p-value threshold was set to $p < 5 \times 10^{-8}$. For the replication, 156 new cases and 58 treated controls from Sweden and Denmark were genotyped using TaqMan SNP genotyping assays, and $p < 0.05$ was considered statistically significant. A meta-analysis was performed using the present GWAS, the Swedish/Danish replication cohort, plus 175 cases of ACEi-induced angioedema and 489 ACEi-treated controls from a previous GWAS in USA.¹ Functional annotations were obtained by testing top hits for linkage disequilibrium (LD) with variants in transcription factor binding sites and chromatin state models.^{2,3} **Results:** SNPs in intron 1 of the *Calcium-activated potassium channel subunit alpha-1 (KCNMA1)* gene on chromosome 10 were significantly associated with angioedema (OR=2.47, 95% CI=1.79-3.42, $p=4.31 \times 10^{-6}$). The subsequent replication and meta-analysis showed a tendency in the same direction (OR=1.51, 95% CI=0.56-4.09, $p=0.42$, and OR=1.89, CI=1.47-2.44, $p=7.57 \times 10^{-7}$, respectively). The associated *KCNMA1* SNPs lacked evidence of functional activity, but were in LD with variants located in transcription factor binding sites in relevant tissues. **Conclusion:** Our data suggest that common intronic variants of the *KCNMA1* gene may be associated with risk of angioedema induced by ACEi or ARB treatment. Future studies based on whole exome sequencing will show whether rare coding variants in *KCNMA1* or other genes contribute to angioedema risk. **References:** 1. Paré G, et al. Pharmacogenet Genomics 2013;23:470-8. 2. Encode Project Consortium. Nature 2012;489:57-74. 3. Roadmap Epigenomics Consortium. Nature 2015;518:317-30.

3171W

Independent effects of lifestyle and genetic factors on risk for coronary artery disease in a large and deeply-phenotyped cohort. A.T. Magis¹, X. Xu¹, O. Manor¹, M.P. Conomos¹, J.E. Rohwer¹, E. Sheng¹, S.M. Speck², G.S. Omenn^{3,4}, N.D. Price^{1,3}, J.C. Lovejoy^{1,3}, L. Hood^{1,3,5}. 1) Arivale, Inc, Seattle, WA; 2) Internal Medicine, University of Washington Medical Center, Seattle, WA; 3) Institute for Systems Biology, Seattle, WA; 4) Departments of Computational Medicine and Bioinformatics, Internal Medicine, Human Genetics, and School of Public Health, University of Michigan, Ann Arbor, MI; 5) Providence St. Joseph Health, Seattle, WA.

Genetic factors as well as unhealthy behaviors such as poor diet, lack of physical activity, and smoking are well-understood as risk factors for coronary artery disease (CAD), the leading cause of death worldwide. A recent seminal study showed that lifestyle and genetic risk factors are independently associated with long-term CAD outcomes but did not investigate their impacts on intermediate phenotypes. The goal of the current study was to characterize the independent impacts of genetic risk as well as lifestyle factors on risk for CAD using a large cohort of deeply-phenotyped individuals. Our study included 2,735 individuals participating in a commercial lifestyle coaching program (Arivale). Data collected included whole genome sequencing, 70 clinical tests, 273 plasma proteins, 961 metabolites, and 16S rRNA sequencing of the gut microbiome. Genetic risk for CAD was estimated using a validated polygenic risk score. Lifestyle risk factors examined were weekly physical activity, dietary pattern, and smoking status. Statistical associations between genetic risk, lifestyle risk factors, and phenotypes were identified using generalized estimating equations (GEE) adjusted for confounders including age, sex, race, observation month, BMI, and medication use. Our results show that, controlling for lifestyle, genetic risk for CAD is most significantly (FDR<0.05) associated with higher LDL cholesterol and LDL-P, lower absolute monocytes count, and higher levels of the protein granulin and the metabolite campesterol. In contrast, controlling for genetics and the other lifestyle risk factors, unhealthy diet is most significantly associated with higher insulin, hs-CRP, and triglycerides, lower omega-3 fatty acids, higher levels of the metabolites carotene diol and beta-cryptoxanthin, and several microbiome changes. Lack of physical activity is most significantly associated with lower HDL cholesterol and HDL-P, higher levels of the proteins leptin and fatty acid binding protein 4, lower levels of the metabolite phenyllactate, and decreased gut microbiome diversity. Smoking is most significantly associated with increased white cell count and increased levels of the metabolites cotinine and o-cresol sulfate, as well as decreased abundance of many microbiome taxa. Interestingly, relatively few phenotypes are associated with multiple risk factors. Our study improves the systemic understanding of CAD risk factors and provides a basis for personalized strategic interventions.

3172T

A phenome-wide association study of genes involved in genetic syndrome reveals pleiotropic effects of common and rare alleles. C. Tcheandjieu^{1,2,3}, P. Saha^{1,2,3}, P. Potiny^{1,2,3}, S. Gustafsson^{1,3}, E. Inglesson^{1,3}, J. Priest^{1,2,3}. 1) Stanford Cardiovascular Medicine, Stanford, CA; 2) Department of pediatric Cardiology; 3) Stanford school of Medicine.

Genetic syndromes (GS) are characterized by a defined set of symptoms or phenotypes which may be linked to one or more distinct loci. The clinical evaluation of a genetic syndrome relies upon recognition of a characteristic pattern of signs or symptoms to guide targeted genetic testing for confirmation of the diagnosis. However, individuals displaying component phenotypes of a complex syndrome may not meet criteria for clinical diagnosis or genetic testing. Here, we present a phenome-wide association study approach exploring pleiotropy of common and rare alleles in genes involved in GS. Using the UK Biobank, a large prospective cohort study of ~500,000 individuals with genotyping and health-related information, we systematically identified 60 phenotypes related to Alagille, Marfan, Usher, and Noonan syndromes by mapping human phenotype ontology (HPO) terms of each syndrome to ICD10 code diseases reported in the hospital admission record of each participant. We performed, in individuals from European ancestry, a logistic regression adjusted on age, sex, and the first 5 genetic principal components for each phenotype and each SNP in the target genes (*JAG1*, *FBN1*, *PTPN11*, *NOTCH2*, and *MAP2K1*). SKAT-o test was performed at the gene and FDR was used for multiple testing corrections. Several SNPs in *JAG1* (for Alagille syndrome) and *PTPN11* (for Noonan syndrome) were associated with diastolic BP, systolic BP, and urine microalbumin level ($p < 10^{-6}$). In addition, SNPs in *PTPN11* were associated with mitral regurgitation ($p < 10^{-9}$) hypothyroidism ($p < 10^{-6}$) and body mass index ($p < 10^{-6}$). Consistent with this, at a gene level, *JAG1* was associated with systolic BP ($p = 3 \times 10^{-6}$), diastolic BP ($p = 1 \times 10^{-7}$), chronic bronchitis ($p = 1 \times 10^{-4}$) and, cataract ($p = 5 \times 10^{-4}$); while *PTPN11* was associated with hypothyroidism ($p = 8.8 \times 10^{-6}$), diastolic BP ($p = 9.4 \times 10^{-6}$), urine micro albumin level ($p = 6 \times 10^{-3}$) and umbilical hernia ($p = 6 \times 10^{-3}$). Our preliminary observations clearly relate polymorphisms in *JAG1* and *PTPN11* in the general population with phenotypes observed in Alagille and Noonan syndromes. Increased urine microalbumin is likely a secondary phenotype related to kidney injury from elevated blood pressure. Our results suggest that SNPs in syndromic genes are causative of individual component phenotypes present in a general population. Further research is needed to characterize the pleiotropic effect of alleles in genes in persons without the corresponding genetic syndrome.

3173F

Replication of a single nucleotide polymorphism variant in *CETP* gene associated with HDL level with tiled regression in the *ClinSeq*® Study.

H. Sung¹, J.A. Sabourin¹, K. Lewis², J.C. Mullikin^{3,4}, L.G. Biesecker², A.F. Wilson¹, NISC Comparative Sequencing Program. 1) Genomics section, Computational and Statistical Genomics Branch, NHGRI, NIH, Baltimore, MD, USA; 2) Medical Genomics and Metabolic Genetics Branch, NHGRI, NIH, Bethesda, MD, USA; 3) Comparative Genomics Analysis Unit, Cancer Genetics and Comparative Genomics Branch, NHGRI, NIH, Bethesda, MD, USA; 4) National Institutes of Health Intramural Sequencing Center (NISC), NHGRI, NIH, Bethesda, MD, USA.

ClinSeq® is a large-scale medical sequencing study designed to investigate associations of sequence variants with traits related to Coronary Artery Disease (CAD). The study currently includes more than 1000 non-smoking participants between the ages of 45 to 65 with normal to severe coronary artery calcification scores. About 200 CAD-related traits were measured at the NIH Clinical Research Center in Bethesda, MD. Exome sequencing was performed with four distinct capture kits at the NIH Intramural Sequencing Center (NISC): Agilent SureSelect with 38 Mb and 50 Mb (Index and ICGC) capture kits for 396 and 320 individuals, respectively, TruSeq (V1 and V2) for 285 individuals, and SeqCap EX Exom+UTR for 18 individuals respectively. Data cleaning was performed within each capture kit data set and then across data sets after merging within-cleaned capture kits. Only single nucleotide variants (SNVs) common to all capture kits were merged yielding 590,441 SNVs in 886 unrelated European Americans (EAs) for further analysis. Tests of association of each SNV with High Density Lipoprotein (HDL) level was performed on untransformed, log-transformed and rank-inverse transformed HDL adjusting for age, sex, BMI, and the usage of cholesterol lowering medication with simple linear regression and tiled regression. Using simple linear regression, rs1532625 in the intron of *CETP* gene was found to have the most significant association for untransformed and transformed HDL after Bonferroni correction, replicating the study by Reilly et al. [2013]. For untransformed and transformed HDL, one additional SNV, rs7205804 was found to be significant in the intron of the *CETP* at a significance level of $1e-06$. Using tiled regression to identify variants that jointly and independently predict HDL over all variants tested, rs1532625, the most significant SNV found with simple linear regression, was selected in the final multiple-regression model. However, rs7205804, the second most significant variant in the simple linear regression analysis was not included in the final tiled regression model because of its high correlation with rs1532625. Rather than test variants individually, which can lead to increased type I error because of unrecognized correlation, testing variants in a multiple regression context reduces type I error and helps in prioritizing variants for additional follow-up.

3174W

A pedigree-based study in Mexican Americans identifies altered plasma lipidome profiles in relation to ANGPTL3 levels and cardiovascular disease.

N.B. Blackburn¹, L.F. Michael², J.M. Peralta^{3,4}, M. Kowala², S. Kumar¹, A.C. Leandro¹, M. Almeida¹, T.D. Dyer¹, M.C. Mahaney¹, M. Mosior², S. McAhren², M. Bellinger², S. Williams-Blangero¹, H.H.H. Göring¹, J.L. VandeBerg¹, R. Duggirala¹, D.C. Glahn^{4,5}, P.J. Meikle², J. Blangero¹, J.E. Curran¹. 1) STDO and Department of Human Genetics, School of Medicine, University of Texas Rio Grande Valley, Brownsville, TX, USA; 2) Eli Lilly and Company, Indianapolis, IN, USA; 3) Menzies Institute for Medical Research, University of Tasmania, Hobart, TAS, Australia; 4) Olin Neuropsychiatric Research Center, Institute of Living, Hartford Hospital, Hartford, CT, USA; 5) Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA; 6) Baker Heart and Diabetes Institute, Melbourne, VIC, Australia.

Cardiovascular disease (CVD) is complex group of disorders that together are the leading cause of death worldwide. Identification of drug targets for preventative treatment is of significant clinical interest. Angiopoietin-like 3 (ANGPTL3) is one such CVD target, with clinical trials of inhibition of this protein currently underway. ANGPTL3 is a liver protein that has roles in lipid metabolism through the inhibition of lipoprotein lipase and endothelial lipase. Studies of ANGPTL3 in humans have focused on the classical clinical lipid parameters (HDL, LDL, TG, cholesterol). Given the potential for ANGPTL3 as a CVD drug target, deeper examination of clinically relevant phenotypes and disease outcomes is warranted. ANGPTL3 levels were measured in 1048 individuals from the San Antonio Family Heart Study. Using variance component modelling in SOLAR we found ANGPTL3 levels to be significantly heritable, ($h^2=0.33$; $P=1.19 \times 10^{-16}$). Given the recognized importance of ANGPTL3 levels in CVD we assessed the CVD relationship with protein levels in this cohort. We found that CVD was significantly associated with ANGPTL3 levels, with higher levels of ANGPTL3 associated with CVD, suggesting that lower ANGPTL3 levels are cardioprotective, ($\beta=-0.89$ SDU, $P=0.001$). Coding variants in *ANGPTL3* were tested for association with protein levels, identifying an unequivocal association of decreased ANGPTL3 levels ($\beta=-1.66$ SDU, $P=9.25 \times 10^{-13}$) with a 5 bp loss-of-function deletion, rs569107562. Previously identified by others as causing familial hypolipidemia in homozygous carriers, we identified 22 heterozygotes with this variant in a single Mexican American family. With lipidome profiling of 319 lipid species available in our cohort, we determined that rs569107562 carriers had significant reductions in levels of several phosphatidylinositol (PI) species including PI(36:2) ($\beta=-1.38$ SDU, $P=1.57 \times 10^{-7}$). ANGPTL3 levels overall also showed a strong positive association with PI levels (PI(36:2), $\beta=0.77$ SDU, $P=1.40 \times 10^{-6}$). The relationship between individual lipids and the therapeutic potential of reducing ANGPTL3 levels warrants further exploration in light of current research that suggests that statin based therapies work through PI reduction. This pedigree based study has used lipidome profiles and CVD outcomes to identify deeper phenotypic associations with ANGPTL3 levels than previously explored. Together this highlights the importance of decreasing ANGPTL3 levels for cardioprotection.

3175T

Next-generation sequencing aligned to high-resolution nuclear magnetic resonance (NMR) measurements reveal role of rare variation in circulating metabolic biomarkers. F. Riveros McKay Aguilera¹, C. Oliver Williams^{2,4}, S. Karthikeyan², K. Walter¹, K. Kundu^{1,5}, J. Danesh^{1,2,3}, E. Wheeler¹, E. Zeggini¹, A. Butterworth^{2,3}, I. Barroso^{1,6}. 1) Wellcome Sanger Institute, Cambridge, United Kingdom; 2) MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; 3) The National Institute for Health Research Blood and Transplant Unit (NIHR BTRU) in Donor Health and Genomics, University of Cambridge, Cambridge, UK; 4) Homerton College, Hills Road, Cambridge, CB2 8PH, UK; 5) Department of Haematology, University of Cambridge, Cambridge Biomedical Campus, Long Road, Cambridge CB2 0PT, UK; 6) University of Cambridge Metabolic Research Laboratories and NIHR Cambridge Biomedical Research Centre, Wellcome Trust-MRC Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK.

Circulating metabolite levels can be used as biomarkers, or risk factors, for cardiovascular disease (CVD). To establish the role of rare coding variation in circulating metabolic biomarker levels, we measured 226 serum lipoproteins, lipids and other amino acids and proteins in INTERVAL participants by NMR. Gene-based association analyses included whole-exome sequence (WES) data from 3,741 participants (discovery), and 3,401 whole-genome sequences (WGS) from independent participants (validation). To increase power at the validation stage, we used correlated NMR traits as covariates in the analysis. Compared to unadjusted tests, this approach increased power in known gene-trait associations (e.g. 11 additional gene-trait associations for *PCSK9*), and identified *ACSL1*, *MYCN*, *FBXO36* and *B4GALNT3* as novel gene-trait associations ($P < 2.5 \times 10^{-6}$). Gene-set analyses identified novel association of loss-of-function (LoF) variants in the regulation of the pyruvate dehydrogenase (PDH) complex pathway, with lipoproteins in three broad categories intermediate density lipoproteins (IDL), low density lipoproteins (LDL) and circulating cholesterol ($P_{\text{META-SKAT}} < 2.41 \times 10^{-9}$). This pathway is crucial for metabolic flexibility. In addition, genes near established HDL-associated loci were enriched for likely deleterious missense and LoF variants ($P_{\text{Bonferroni}} < 0.005$) in 18 HDL related traits, suggesting this gene set is enriched for effector transcripts. Exploring phenotypic tails of 49 lipoprotein and lipid measurements, we found enrichment ($P_{\text{permutation}} < 0.00037$) of likely deleterious rare variation in lipoprotein disorder and metabolism gene sets (e.g. hyperlipidaemia and HDL remodelling) in the lower tails of four measurements which are CVD risk factors (e.g. cholesterol in small VLDL and concentration of small HDL particles), demonstrating that rare "protective" variation is prevalent in the phenotypic tails of a healthy population. These findings provide novel biological insights into lipoprotein and lipid metabolism through rare variant analyses.

3176F

Influence of a coronary-artery-disease-associated genetic variant on *FURIN* expression and effect of furin on macrophage behavior. S. Ye^{1,2}, G. Zhao², W. Yang², J. Wu², B. Chen², X. Yang², J. Chen³, D.G. McVey¹, C. Andreadi¹, P. Gong¹, T.R. Webb¹, N.J. Samani¹. 1) University of Leicester, Leicester, United Kingdom; 2) Shantou University Medical College, Shantou, China; 3) Beijing University Shenzhen Hospital, Shenzhen, China.

Objective—Genome-wide association studies have revealed a robust association between genetic variation on chromosome 15q26.1 and coronary artery disease (CAD) susceptibility; however, the underlying biological mechanism is still unknown. The lead CAD-associated genetic variant (rs17514846) at this locus resides in the *FURIN* gene. In advanced atherosclerotic plaques, furin is expressed primarily in macrophages. We investigated whether this CAD-associated variant alters *FURIN* expression and whether furin affects monocyte/macrophage behavior. **Approach and Results**—A quantitative RT-PCR analysis showed that leukocytes from individuals carrying the CAD risk allele of rs17514846 had increased *FURIN* expression. A chromatin immunoprecipitation assay revealed higher RNA polymerase II occupancy in the *FURIN* gene in mononuclear cells of individuals carrying this allele. A reporter gene assay in transiently transfected monocytes/macrophages indicated that the CAD risk allele had higher transcriptional activity than the non-risk allele. An analysis of isogenic monocyte cell lines created by CRISPR-mediated genome editing showed that isogenic cells homozygous for the risk allele had higher *FURIN* expression levels than isogenic cells homozygous for the non-risk allele. An electrophoretic mobility shift assay exhibited preferential binding of a nuclear protein to the risk allele. Studies of monocytes/macrophages with lentivirus-mediated furin overexpression or shRNA-induced furin knockdown showed that furin overexpression promoted monocyte/macrophage migration, increased proliferation and reduced apoptosis, whereas furin knockdown had the opposite effect. **Conclusion**—Our study shows that the CAD-associated genetic variant increases *FURIN* expression and that furin promotes monocyte/macrophage migration and proliferation while inhibiting apoptosis, providing a mechanistic explanation for the association between variation at the chromosome 15q26.1 locus and CAD risk.

3177W

Modifiable risk factors and heart failure: A Mendelian randomization study.

S. Grover¹, S. Rapp², P.S. Wild², I.R. Koenig¹. 1) Institut für Medizinische Biometrie und Statistik, Universität zu Lübeck, Universitätsklinikum Schleswig-Holstein, Campus Lübeck, Lübeck, Germany; 2) Präventive Kardiologie und Medizinische Prävention, Zentrum für Kardiologie, Universitätsmedizin der Johannes-Gutenberg-Universität Mainz, Mainz, Germany.

Heart failure (HF) is a highly heterogeneous syndrome, and its diagnosis and treatment continues to pose serious challenges. Availability of improved imaging techniques has further lead to revised diagnostic groups with systolic (HF with reduced ejection fraction, HFREF) and diastolic HF (HF with preserved ejection fraction, HFPEF) contributing equally to increased morbidity and mortality rates. These subtypes of HF are assumed to be heterogeneous with regard to genetic aetiology and risk factors. Observational studies have reported obesity (OBS), hypertension (HTN) and diabetes (DBS) as likely risk factors associated with cardiovascular diseases (CVD). We perform a two sample Mendelian randomization (MR) study evaluating the causal effect of OBS, HTN and DBS on HF, HFREF and HFPEF in a European population of 3000 HF cases. To identify candidate instrument variables (IV) for judging the effect of risk factors on HF risk, we use glycated hemoglobin (HbA1c) and diagnosis of diabetes as the measures of diabetes, BMI and WHR for obesity, SBP and DBP for hypertension. We prioritize genetic loci that had reached genome-wide significant associations ($P < 5 \times 10^{-8}$) in the largest available GWAS studies in European populations to date for all the risk factors. Inverse variance weighted (IVW) method, MR Egger and Weighted Median method are used to judge the causal effect. In addition to detection of pleiotropic variants through the heterogeneity Q test and the I² index, MR Egger method and funnel shaped distribution of individual SNP level causal effect estimates are also used. Furthermore, sensitivity analyses are conducted on the basis of instruments comprising a subgroup of SNPs with known biological function in predisposition to risk factors. In the absence of potential pleiotropic variants, MR approach overcomes the issues of confounding and reverse causality inherent in observational studies with an increased power in the presence of multiple SNPs. This approach may help us to identify causal markers for specific HF subtypes for potential intervention and to define underlying biological pathways.

3178T

Utilizing admixture mapping to identify genomic regions associated with coronary artery calcification in African Americans.

Z. Liu¹, N.F. Hansen¹, J.C. Mullikin^{1,2}. 1) Comparative Genomics and Cancer Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 2) NIH Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD.

Admixture mapping is a powerful technique to identify genetic loci for heritable traits. We set out to map loci for coronary artery calcification (CAC), which occurs in higher rates in European Americans than African Americans. African Americans have varying proportions of African ancestry, due to the admixture of European and African ancestral populations. Proceeding on the assumption that causal variants are more frequently found on DNA segments that were inherited from the ancestral population with the higher disease risk, we employed admixture mapping to identify genetic loci with excess local European ancestry in African Americans displaying our phenotype, and thereafter targeted these loci to map disease genes. The ClinSeq@A2 cohort includes 502 participants of self-reported African descent, aged 45-65 at enrollment, each of whom was exome sequenced and measured for CAC-related phenotypes. After QC, 475 individuals were retained. To infer local ancestry, we performed SEQMIX analyses using European and African datasets from the 1000 Genomes Project as references. We then averaged the inferred local ancestry across the entire genome for all individuals, and results showed that our study participants had an average of 23% European and 77% African ancestry. For global ancestry, we used LASER to calculate principal components (PCs), which explained variations among individual genotypes. The PC estimation of global ancestry showed high correlation with genome-wide averaged local ancestry ($r < -0.97$). Based on the Agatston CAC scoring, we divided our samples into cases (CAC score > 0 ; $N = 135$) vs. controls (CAC score = 0; $N = 340$), and performed admixture mapping between local European ancestry and CAC, adjusting for global European ancestry, age, sex, BMI and lipid statin. Results showed that chromosomal region 9p13 had the top association signal between local European ancestry and the presence of CAC (min $P = 1.39 \times 10^{-3}$, $\beta = 1.58$), however, our results did not reach a genome-wide significant threshold of 7.48×10^{-4} after multiple-testing correction. These results suggested that this region may harbor CAC related genes that lead to prevalence differences in racial-ethnic groups. We will perform a gene-based test within this region to further identify genes and verify them in an independent dataset. Admixture mapping is complementary to genetic association studies in admixed populations when the trait has prevalence deviations in racial-ethnic groups.

3179F

Investigating pleiotropic architecture of 92 plasma proteins related to cardiovascular disease. *L. Repetto¹, Z. Ning², A. Bretherick³, J.F. Wilson^{1,3}, X. Shen^{1,2}.*

1) Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Teviot Place, Edinburgh, United Kingdom; 2) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 3) MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, United Kingdom.

Motivation: Multivariate genome-wide association analyses have been proven to provide higher power than standard univariate analyses in terms of discovering novel pleiotropic loci for a range of complex traits, e.g. omics phenotypes. This provides an opportunity to further understand the genotype-phenotype map between the genome and complex disease through omics measurements. **Results:** We measured a panel of 92 cardiovascular disease (CVD)-related plasma proteins in an isolated Scottish population (ORCADES) and conducted a 92-trait combined multivariate genome scan, using summary association statistics from 92 single-trait scans. We discovered 27 pleiotropic pQTL, and we replicate 17 multi-protein associations and examine the pleiotropic genetic effects using an independent cohort from Croatia (VIS). The replicated loci include the well-known ABO locus for blood types, whose association with CVD has been reported but not fully understood. Here, we provide a new multi-protein insight into the ABO-CVD connection. **Conclusion:** Our study provides new insights into shared genetic architecture across a set of human plasma proteins, as mediators of CVD. The analysis pipeline emphasizes the importance of secondary multivariate analysis of summary association statistics, as one can gain substantial statistical power and genetic interpretation that are difficult to achieve using conventional analyses.

3180W

Toward precision medicine: LpPLA2 activity and a LpPLA2 GRS show an effect on cardiovascular outcomes stratified by glycaemic control. *M.K. Siddiqui¹, G.S. Smith¹, C. Gao², L. Wallentin³, H. White⁴, D. Waterworth⁵, C.N.A. Palmer¹.*

1) University of Dundee, Dundee, United Kingdom; 2) Parexel International, Durham, NC, USA; 3) Department of Medical Sciences and Uppsala Clinical Research Center, Uppsala University, Uppsala, Sweden; 4) Green Lane Cardiovascular Service, Auckland City Hospital, and Auckland University, Auckland New Zealand; 5) GlaxoSmithKline, King of Prussia, PA, USA.

Lipoprotein-associated phospholipase A2 (LpPLA2) activity has a well-established biological and statistical association with major adverse cardiovascular events (MACE), independent of other risk factors. However, clinical trials for an LpPLA2 inhibitor darapladib has failed to show a significant reduction in MACE. Recently, it has been demonstrated that there is increased expression of LpPLA2 coding *PLA2G7* gene in adipose tissues of individuals with type 2 diabetes (T2D). We examined the impact of glycemic control (GC) on the association between LpPLA2 and MACE and demonstrate consistent results in the STABILITY trial of darapladib. Further, we developed a genetic risk score (GRS) for this biomarker and performed a Mendelian randomization (MR) experiment to confirm our findings. Discovery analyses were performed in the GoDARTS cohort of T2D (n=7420). Subjects were classified as having good (HbA1c<48 mmol/mol, n=1979) or poor (HbA1c≥48 mmol/mol, n=5441) GC. The association was analysed using survival analyses with 10 years of follow-up from the time of LpPLA2 measurement. In the STABILITY trial, we would expect that darapladib would confer a protective effect against major coronary events (MCE) for diabetics (n=15,828). Therefore, participants were stratified by T2D status. 33.6 % of those on darapladib and 34% on placebo had type 2 diabetes. MR was conducted in the GoDARTS cohort by selecting SNPs from previous GWAS studies of LpPLA2 activity. In the GoDARTS study, amongst individuals with good GC, the highest quartile of LpPLA2 (Q4) had 1.2 times the hazards of MACE compared to the lowest quartile (Q1) (*P*-value=0.04), whereas for individuals with poor GC the hazards were 1.4 (*P*-value<1×10⁻³), an increase of 20%. In STABILITY, amongst participants with T2D, those randomized to receive darapladib were significantly protected against having a MCE, HR 0.72 (*P*-value=4 ×10⁻³) compared to placebo. MR analysis showed that the GRS was associated with MACE, the highest quartile showed hazards of 1.26 (*P*-value=9×10⁻³) compared to the lowest. However, when stratified, the effect was only evident in the sub-group with poor GC. These results suggest that high LpPLA2 activity might be especially important in patients with poorly controlled T2D. Therefore further investigation of the effects of darapladib in relation to T2D seems warranted. Further, the preliminary GRS suggests that personalised risk could be predicted without the need to directly measure enzyme levels.

3181T

Investigation of the role of Growth Differentiation Factor 15 (GDF15) in human metabolic syndrome through population genetics. *E.M. Wigmore¹, S.M. Lemmela², J. Merino³, A. Platt¹, R. March¹, A. Palotie^{2,3}, M.J. Daly^{2,3}, M. Kinnunen², A. Matakidou¹. 1) Centre for Genomics Research, Precision Medicine and Genomics, IMED Biotech Unit, AstraZeneca, Cambridge, UK; 2) Institute of Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; 3) Broad Institute of MIT and Harvard, Cambridge, USA.*

GDF15 is a stress responsive cytokine known to be upregulated in type 2 diabetes and cardiovascular disease. Recent data from animal models have reported that knock-out of *GDF15* results in increased body weight, whereas overexpression leads to reductions in body weight and fat mass and improved metabolic parameters. These studies have identified GDF15 as a potential target for the treatment of obesity and related diseases (Xiong et al., 2017). However, data supporting its causal role in human disease are currently lacking. Here, we utilise a well characterised, Finnish, population-based cohort with baseline serum protein levels ($n = 5,817$) to examine GDF15 phenotypic and genetic associations. Associations of serum GDF15 levels with clinical phenotypes have to-date been analysed for 166 clinical endpoints and 18 quantitative traits ($n = 184$, multiplicity adjusted significance $p < 2.7 \times 10^{-4}$). The most significant associations observed were with key components of metabolic syndrome: type 2 diabetes ($OR = 1.5$, $P = 7.0 \times 10^{-12}$), serum insulin ($\beta = 0.02$, $P = 8.4 \times 10^{-12}$), serum triglycerides ($\beta = 0.14$, $P = 5.8 \times 10^{-23}$) and waist-to-hip ratio ($\beta = 0.066$, $P = 3.8 \times 10^{-10}$). Associations remained significant after correction for type 2 diabetes and body mass index (BMI). Consistent with previous reports we also observed associations with major cardiovascular events ($OR = 1.2$, $P = 1.1 \times 10^{-4}$), hypertension ($OR = 1.2$, $P = 1.1 \times 10^{-4}$), heart failure ($OR = 1.4$, $P = 5.4 \times 10^{-5}$), chronic kidney disease ($OR = 2.3$, $P = 8.5 \times 10^{-7}$) and cancer ($OR = 1.2$, $P = 4.8 \times 10^{-5}$) but not with diseases of the respiratory, neurological or gastrointestinal systems. These data support the previously reported role of GDF15 secretion as an adaptive response specific to cardiometabolic disease. To identify genetic drivers of serum GDF15, a genome-wide association analysis (GWAS) was performed. We identified a 100kb locus around the *GDF15* gene that strongly associated with protein levels ($P = 4.6 \times 10^{-83}$) and fine mapping analysis identified four putative causal variants (rs16982345, rs1054221, rs3787023, rs189593084). These data provide the genetic instruments for Mendelian Randomisation methodologies that allow assessment of causality between GDF15 and body mass index and adiposity and clarification of its likely clinical utility as a human therapeutic target.

3182F

Mutation spectrum in a large cohort of hypertrophic cardiomyopathy patients in Korea. *D. Yang¹, MW. Seong^{1,2}, SC. Lee³, CS. Ki⁴, SH. Seo⁵, M. Kim¹, SI. Cho¹, DY. Kang⁶, SP. Lee⁷, HK. Kim⁷, DW. Sohn⁷, S. Park¹.* 1) Seoul National Univ Department of Laboratory Medicine, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, Korea; 2) Biomedical Research Institute, Seoul National University Hospital, Seoul, Korea; 3) Department of Internal Medicine, Samsung Medical Center, Sungkunkwan University College of Medicine, Seoul, Korea; 4) Department of Laboratory Medicine, Samsung Medical Center, Sungkunkwan University College of Medicine, Seoul, Korea; 5) Department of Laboratory Medicine, Seoul National University Bundang Hospital, Seoul National University College of Medicine, Seoul, Korea; 6) Department of Internal Medicine, Asan Medical Center, Seoul, Korea; 7) Department of Internal Medicine, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, Korea.

Purpose: Hypertrophic cardiomyopathy (HCM) is characterized by left ventricular hypertrophy, particularly involving the interventricular septum, increased risk of sudden cardiac death, atrial fibrillation, and heart failure. Although progression in field of molecular genetics afford us insights into complex mechanism of the HCM, still identifying de novo and inherited variant is important. Here, we aimed to determine the mutational spectrum of the 36 HCM-related genes in a large cohort comprising 381 HCM patients referred to our laboratory. **Methods:** 36 hypertrophic cardiomyopathy-related genes were analyzed among 381 patients referred to our laboratory for genetic testing from 16 different hospitals throughout the country during 2013-2016. **Results:** Among 381 patients, 164 (43.0%) had single or multiple likely pathogenic (LP)/pathogenic (P) variants. Of 164 patients harboring one or more LP/P variants, 14 (8.5%) had multiple heterozygous LP/P variants, and 10 (6.1%) had genetic abnormalities in genes associated with systemic diseases accompanying HCM. **Conclusions:** To our knowledge, this is the largest nationwide study to reveal the mutational spectrum in Korean HCM patients using a comprehensive gene panel. Our results indicate that a comprehensive gene panel including non-sarcomere as well as sarcomere genes might be useful to diagnose HCM genetically.

3183W

Sodium, potassium and cardiovascular disease in the UK Biobank: Observational analyses, GWAS and Mendelian randomization. *D. Zanetti^{1,2}, H. Bergman³, T.L. Assimes², E. Ingelsson^{1,2}*. 1) Stanford Cardiovascular Institute, Stanford University, Stanford, CA; 2) Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA; 3) Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Sweden.

Background Urinary sodium and potassium have been suggested to be associated with cardiovascular diseases (CVD) and type 2 diabetes mellitus (T2D), but it is unknown whether these associations are causal. Mendelian randomization (MR) is a new method to address this important question. **Methods** We performed multivariable regression models and multivariable-adjusted Cox proportional hazards models to assess associations of urinary sodium and potassium with several cardiovascular risk factors and outcomes in 483,063 participants of the UK Biobank. Further, we performed genome-wide association studies (GWAS) of sodium and potassium and we assessed the causal relationship of the above associations using the two-sample MR approach combining analyses in UK Biobank with publicly available data from GWAS. **Results** In our observational analyses, urinary sodium and potassium showed significant positive associations with body fat percentage, body mass index (BMI), and waist-to-hip ratio (WHR) (β , 0.05-0.17; per SD change in exposure and outcome) after Bonferroni correction ($P \leq 2.1 \times 10^{-3}$). High sodium was associated with increased systolic blood pressure (SBP) (β , 0.02; per SD change in exposure and outcome); whereas high potassium was associated with lower SBP and diastolic blood pressure (DBP) (β , -0.11 and -0.08; per SD change in exposure and outcome, respectively). Urinary sodium showed significant inverse associations with atrial fibrillation (hazard ratio; 0.88), lipid-lowering medication and T2D (odds ratio [OR]; 0.96 and 0.97; respectively), whereas potassium showed significant positive association with lipid-lowering medication (OR; 1.03). We identified 38 genome-wide significant independent loci associated with sodium, and 12 for potassium. The MR analyses showed no evidence of causal associations of urinary sodium and potassium with SBP and DBP, BMI and WHR, or lipid levels (high and low-density lipoprotein cholesterol and triglycerides). The statistical power of our MR analyses to detect a causal effect of 0.15 SD change in outcome per 1-SD change in exposure was $\geq 80\%$. Sensitivity analyses provided consistent results and indicated no significant directional horizontal pleiotropy. **Conclusion** Our study indicates that although urinary sodium and potassium, used as a proxy for sodium and potassium intake, are associated with several cardiometabolic outcomes in observational analyses; the associations with blood pressure, obesity and lipid levels are unlikely to be causal.

3184T

Genome-wide analysis of ex vivo coagulation biomarkers provides new insights into the genetic risk for venous thromboembolism. *J.D. Backman¹, S.E. McCarthy¹, C. O'Dushlaine¹, A. Marcketta¹, A.H. Li¹, J.H. Chung¹, D.J. Carey², D.H. Ledbetter², O. Gottesman¹, J.D. Overton¹, J.G. Reid¹, A. Baras¹, F.E. Dewey¹, T.M. Teslovich¹*. 1) Regeneron Genetics Center, Tarrytown, NY; 2) Geisinger Health System, Danville, PA.

Venous thromboembolism (VTE) is a chronic, debilitating, and potentially life-threatening disease in which blood clots form abnormally and disrupt blood flow in the deep veins (Deep Vein Thrombosis), oftentimes embolizing and becoming lodged in the lungs (Pulmonary Embolism). The etiology of VTE involves a complex interplay of environmental risk factors and genetic predisposition, with hypercoagulability, inflammation, and reduced blood flow believed to be the triad of conditions that ultimately culminate in a thrombotic event. However, despite a strong genetic role and the apparent involvement of multiple tissues and biological pathways, insights from genetic studies of VTE have been limited, pointing primarily to variation in coagulation-related genes as the most prominent drivers of genetic risk. We hypothesized that the genetic analysis of coagulation-relevant phenotypes such as activated partial thromboplastin time (aPTT) and international normalized ratio of prothrombin time (INR), which measure the intrinsic and extrinsic coagulation pathways, respectively, can provide new insights into the etiology of VTE. We conducted the largest genome-wide analysis (GWAS) of aPTT and INR reported to date, identifying 5 novel association signals for aPTT and 11 novel association signals for INR in the DiscovEHR cohort (aPTT $N=23,325$; INR $N=32,301$). To evaluate clinical implications, we tested each variant for association with VTE risk in a total of 18,502 cases and 470,396 controls from the DiscovEHR and UK Biobank cohorts. We observed significant associations with VTE risk for 8 of the 23 independent variants we analyzed, including 3 variants in *F5* that associate with VTE independently of *F5* Leiden. As a complementary approach to our discovery analyses, we also queried the GWAS catalog for variants with links to 17 different hemostatic traits and performed follow-up analyses to determine whether these markers are associated with VTE risk. Of the 274 independent signals we mined from published studies, 72 variants were significantly associated with VTE risk after multiple-testing correction ($P \leq 1.8 \times 10^{-4}$). Several of these signals map to genes with minimal prior evidence for association with VTE, including *SCARA5*, *TC2N*, and *CLEC4M*. While further vetting and functional follow-up is required, these findings offer up potential new therapeutic targets for VTE treatment and provide novel insight into the genetic variation that predisposes to thrombotic event risk.

3185F

Heritability of atrial fibrillation across a spectrum of clinical risk factors.

L. Weng^{1,2}, S. Choi², A. Hall^{1,2}, C. Roselli², M. Chaffin², P. Ellinor^{1,2}, S. Lubitz^{1,2}.

1) Massachusetts General Hospital, Boston, MA; 2) The Broad Institute of Harvard and MIT, Cambridge, MA.

Background: Atrial fibrillation (AF) is a common and morbid arrhythmia. Although AF is heritable (SNP-heritability [h^2] 22%), the lifetime risk of AF is also influenced by clinical risk factors. The contributions of genetic predisposition to AF heritability across the spectrum of clinical risk factor burden has not been assessed. **Methods:** Our study included 345,373 unrelated individuals of European ancestry (AF-free at baseline) from the UK Biobank. We defined clinical risk factors for AF including obesity, smoking status, type 2 diabetes, hypertension, and coronary artery disease based on baseline interviews and medical data. Incident AF was ascertained through medical data or death records during follow-up. We assessed h^2 of AF using a variance components method (BOLT-REML) on ~650K genotyped common/low-frequency variants (minor allele frequencies $\geq 0.5\%$) across a spectrum of clinical risk factor burden (0, 1, and ≥ 2 clinical risk factors). All analyses were adjusted for baseline age, sex, array, and AF-related principal components of ancestry. **Results:** A total 9,337 individuals developed AF during follow-up. The baseline age was higher among those who developed incident AF than in those who did not (mean 62.7 vs. 57.1 years), and more were male (65% vs 46%). The overall h^2 for incident AF was 17.4% (95%CI 15.2%-19.7%), and 15.5% (9.7%-21.4%) in individuals without any clinical risk factors. Similarly, the h^2 for individuals with 1 or ≥ 2 clinical risks factors was 14.5% (8.6%-20.4%) and 17.2% (9.33 %-25.0%), respectively. **Conclusion:** The additive genetic heritability of incident AF is about 17% and does not vary substantially by clinical risk factor burden. Thus, genetic variation is a significant contributor to AF susceptibility regardless of an individual's clinical risk factor profile.

3186W

Identification of novel structural variations affecting common and complex disease risks with >16,000 whole genome sequences from ARIC and HCHS/SOL.

G. Jun¹, F.J. Sedlazeck², H. Chen¹, B. Yu¹, Q. Qi³, O. Krasheninnikova², A. Carroll⁴, X. Liu¹, A. Mansfield², S. Zarate⁴, G. Metcalf², D. Muzny², S. Lindstrom^{5,6}, E. Selvin⁷, R. Kaplan³, W.J. Salerno², R. Gibbs², E. Boerwinkle^{1,2}.

1) Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 3) Department of Epidemiology & Population Health, Albert Einstein College of Medicine, Bronx, NY; 4) DNAnexus, Mountain View, CA; 5) Fred Hutchinson Cancer Research Center, Seattle, WA; 6) Department of Epidemiology, School of Public Health, University of Washington, Seattle, WA; 7) Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University Baltimore, MD.

Despite their clear impact on Mendelian diseases, the contribution of structural variation (SV) to common complex phenotypes has been relatively understudied. A key missing opportunity has been the lack of available high-quality SV data in large deeply-phenotyped cohorts. We analyzed whole genome sequencing (WGS) data from 16,424 individuals: 6,369 European Americans (EAs) and 2,282 African Americans (AAs) from the Atherosclerosis Risk in Communities (ARIC) study and 7,772 individuals from Hispanic Community Health Study/Study of Latinos (HCHS/SOL) sequenced by NHGRI Centers for Common Disease Genomics (CCDG) and NHLBI Trans-Omics for Precision Medicine (TOPMed) projects. We developed a massively parallel SV discovery and genotyping pipeline, Parliament2, that integrates discovery from multiple methods across a large number of samples to maximize discovery sensitivity. We ran the Parliament SV discovery pipeline followed by muCNV multi-sample genotyper to obtain 163,292 deletions and 71,815 duplications. Most of SVs were rare with median MAF of 0.06%. We observed relatively more common SVs in AAs (median MAF 0.13%) than in EAs (0.05%) and in Hispanic/Latinos (0.03%). We assessed the contribution of individual SVs to multiple cardiometabolic, hematology and liver function quantitative traits (e.g. glucose, HbA1c, red blood cell indices, blood pressure, cholesterol, AST, ALT, and GGT). We identified a 2400-bp deletion overlapping *HBA2* in human alpha globin cluster to be significantly associated with lower hemoglobin ($p=6.85 \times 10^{-9}$), hematocrit ($p=6.38 \times 10^{-9}$), and higher HbA1c ($p=1.74 \times 10^{-12}$) levels. MAF was 18% in AAs, 0.5% in EAs, and 5.0% in Hispanic/Latinos. We also discovered an 800-bp deletion in *BCRP3* at 22q11.23 that was significantly associated with higher gamma-glutamyltransferase (GGT) levels ($p=5.7 \times 10^{-13}$), a clinical marker of liver and cardiometabolic diseases, located at approximately 9 kilobase downstream of *GGT1* that encodes type I GGT enzyme. This deletion was common (MAF 49%) in AAs, rare in EAs (MAF 0.26%), and intermediate (MAF 11%) in Hispanic/Latinos. In AAs, this variant explained 1.4% of variability of GGT levels. Our results suggest that structural variations may partially explain racial differences in the distributions of several biomarkers. These results establish a precedent for analysis of SV data for comprehensive assessment of the genetic architecture of common complex phenotypes.

3187T

Exome sequencing reveals variants in F5 are associated with ACE

inhibitor and ARB induced angioedema. C. Maroteau¹, M.K. Siddiqui¹, A. Veluchamy¹, F. Carr¹, M. White¹, A.J. Cassidy¹, E.V. Baranova², P. Hallberg³, E.R. Rasmussen³, N. Eriksson⁴, K. Bloch⁵, A. Alfrevic⁶, A.H. Maitland-van der Zee^{3,8}, M. Wadelius⁴, C.N.A. Palmer¹ on behalf of PREDICTION-ADR. 1) Division of Molecular and Clinical Medicine, University of Dundee, Ninewells Hospital and Medical School, Dundee, Angus, United Kingdom; 2) Tayside Centre for genomic Analysis, School of Medicine, University of Dundee, Dundee, U.K.; 3) Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht University, Utrecht, The Netherlands; 4) Department of Medical Sciences, clinical Pharmacology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 5) Department of Otorhinolaryngology, Head & Neck Surgery and Audiology, Rigshospitalet, University of Copenhagen, Denmark; 6) Uppsala Clinical Research Centre, Uppsala, Sweden; 7) Department of Molecular and Clinical Pharmacology, University of Liverpool, Liverpool, U.K.; 8) Department of Respiratory Medicine, Academic Medical Centre, University of Amsterdam, The Netherlands.

Aims: Angioedema occurring in the head and neck region is a rare extreme life-threatening reaction to angiotensin converting enzyme inhibitors (ACEI) and angiotensin receptor blockers (ARB). The European consortium PREDICTION-ADR was launched in 2013 to investigate the genetic variation responsible for extreme adverse drug reactions (ADRs) induced by ACEI or ARB and statins, using ascertained cases. **Methods and results:** Adjudicated cases of extreme ACEI or ARB-induced angioedema (ACEI-AE, ARB-AE) and controls had been recruited in 4 different centres: Uppsala (Sweden), Dundee (Scotland), Utrecht (The Netherlands), Liverpool (England). Sequencing of 1066 samples (408 ACEI and ARB cases, 658 controls) was performed using exome-enriched sequence data (SureSelect QXT, XT and XT2 reagents, Agilent Technologies, Wokingham, UK). After QC control, 1033 individuals (387 cases, 646 controls) with 128,157 SNPs (MAF<1%) and 50,838 SNPs (MAF>1%) were analysed. Common and rare *F5* variants seemed to be associated with ACEI-AE and ARB-AE phenotype. Common variant rs6025 (MAF=0.05) was found significantly associated with both ACEI and ARB angioedema [OR: 2.85, 95% confidence interval (CI): 1.89-4.25] where those homozygous for Arg534 had increased odds of developing ACEI-AE or ARB-AE. A group of 5 rare variants (MAF=5x10⁻⁴) in the *F5* gene (rs140530655, CHR1:169529903, rs200157005, rs149389480, rs143509841), was found to trend toward significance, p-value=2.09x10⁻³ (corrected threshold for SKAT-0 analysis: 10⁻⁴). A combined genetic risk score (GRS) of these variants and rs6020 showed individuals carrying at least one of these variants had 2.21 (95% CI: 1.49-3.27) times the odds of having ACEI-AE or ARB-AE. **Conclusion:** This study suggests a role of the anticoagulation system in angioedema. Carrying an alternative variant of *F5* (rs6025, rs6020, rs200157005, rs149389480, rs143509841, CHR1:169529903 and CHR1:169529903) increases the odds of developing angioedema when treated with ACEI or ARB.

3188F

Natural genetic variation of the human atrial transcriptome, proteome

and metabolome. I. Assum^{1,2}, J. Krause^{3,4}, M. Scheinhardt⁵, T. Zeller^{3,4}, R. Schnabel^{3,4}, M. Heinig^{1,2}. 1) Institute of Computational Biology, Helmholtz Center Munich, Neuherberg, Germany; 2) Department of Informatics, Technical University of Munich, Garching, Germany; 3) University Heart Center Hamburg, University Medical Center Hamburg-Eppendorf, Hamburg, Germany; 4) DZHK (German Center for Cardiovascular Research), partner site Hamburg/Kiel/Lübeck, Hamburg, Germany; 5) Institute of Medical Biometry and Statistics, University of Lübeck, University Hospital of Schleswig-Holstein, Campus Lübeck, Lübeck, Germany.

Genome-wide Association Studies (GWAS) have uncovered thousands of disease associated genetic loci. A common approach to identify underlying causal variants and genes is to consider tissue specific, cis-acting expression quantitative trait loci (cis eQTL). However, little is known about the downstream consequences on protein and metabolite levels. Here, we present a comprehensive systems genetics approach in human heart atrial appendage tissue samples from patients undergoing bypass surgery (N=75). This is the first survey that combines proteomics in disease relevant tissue with transcriptomics and metabolomics measurements. This unique data set allowed for the identification of widespread effects of genetic variants on cis regulatory elements for transcript (eQTL), protein (pQTL) and metabolite (mQTL) levels, as well as for an integrated analysis across molecular levels. We focused on a subclass of pQTLs, where the genetic variation in protein concentrations cannot be explained by the regulation of their corresponding mRNA and is most likely the result of posttranscriptional regulation or missense mutations. Our findings represent a powerful resource to follow up and understand genetic regulatory mechanisms, particularly for protein and metabolite variations related to cardiovascular disease. As an example we identified highly significant eQTL as well as pQTL for *MYOZ1* - a GWAS hit for atrial fibrillation, a disease directly related to atrial tissue.

3189W

Linkage evidence and TOPMed whole genome sequencing identify multiple genes on chromosome 1 with rare variants associated with blood pressure. K.Y. He¹, H. Wang², K. Schwander³, J.A. Smith⁴, T.N. Kelly⁵, X. Zhu¹, D.K. Arnett⁶, TOPMed Blood Pressure Working Group. 1) Case Western Reserve University, Cleveland, OH; 2) Brigham and Women's Hospital, Boston, MA; 3) Washington University in St. Louis, St. Louis, MO; 4) University of Michigan School of Public Health, Ann Arbor, MI; 5) Tulane University School of Public Health and Tropical Medicine, New Orleans, LA; 6) University of Kentucky College of Public Health, Lexington, KY.

Elevated blood pressure or hypertension is a major risk factor for coronary heart disease and cerebrovascular disease. The prevalence of hypertension in African Americans is among the highest in the world. We followed up a 33 Mb region with a linkage peak on 1q31 associated with pulse pressure (MLOD = 3.8) that was previously identified with 4,394 African American subjects (1,802 families) in the Family Blood Pressure Program, consisting of HyperGEN, GENOA, and GenNet. Here, we analyzed 2,806 African-American subjects in HyperGEN and GENOA that were whole-genome sequenced by TOPMed as a discovery cohort by conducting gene-based and single SNP association analysis in the linkage region for 3 blood pressure (BP) traits: SBP, DBP, and PP. We hypothesized that the linkage peak could be partially explained by low frequency and rare variants that are segregating in the families contributing to the linkage peak. Our analysis was restricted to variants that segregate at least twice within families that passed a variable threshold for family-specific LOD scores. We used functional annotation to focus on 174 protein-coding genes and analyzed functional coding variants and non-coding variants (MAF < 1%; CADD > 10) separately. Our analysis of selected functional coding and non-coding variants resulted in 46 unique genes with either burden or SKAT $p < 0.05$ for at least 1 BP trait in the discovery cohort. These 46 genes were carried forward for verification analysis in 9 independent TOPMed studies consisting of 27,214 individuals of African, European, and Asian descent. In the multiethnic meta-analysis of functional coding variants, we observed 2 genes significantly associated with SBP and passed Bonferroni correction for multiple testing (*NAV1*, $p = 0.0016$ for SKAT; *PTPN7*, $p = 0.0012$ for burden test). For non-coding variants, there was 1 gene significantly associated with PP and passed the multiple testing threshold (*MR1*, $p = 1.37 \times 10^{-8}$ for burden test). *MR1* is especially noteworthy because we identified 5 non-coding rare variants significantly associated with PP in the meta-analysis of African Americans ($p = 0.0087$ for burden test) and passed the Bonferroni correction in the meta-analysis of European Americans ($p = 1.7 \times 10^{-4}$ for burden test). Additional tissue-specific functional annotation analysis and gene expression association analysis are currently underway.

3190T

Gender and age possibly affect the moderation of *BDNF* Val66Met genotype on psychosocial stress-related cardiovascular disease risk factors in a large harmonized sample? R. Jiang¹, M.A. Babyak¹, A. Singh¹, B.H. Brummett¹, E.R. Hauser^{1,2}, I.C. Siegler¹, W.E. Kraus¹, K.M. Huffman², R.B. Williams¹. 1) Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, NC; 2) Duke Molecular Physiology Institute, Duke University Medical Center, Durham, NC.

Our previous studies have shown that the *BDNF* Val66Met genotype moderates the association between psychosocial stress and risk factors of cardiovascular disease (CVD), namely a significant gene×stress interaction effect on the outcomes. However, the effect is inconsistent in the literature, perhaps due in part to small sample sizes with non-standardized measurements in individual studies, as well as failure to account for possible age and gender effect. This study examines whether the three-way interaction, age×gene×stress or gender×gene×stress, is associated with depressive symptoms and metabolic traits using a large harmonized dataset (N=33030) from 10 studies consisting of Duke in-house and dbGap public access datasets. Stress was scored using a validated algorithm based on specific stress components in each study. Depressive symptoms were assessed by available Center for Epidemiological Studies Depression Scale (CES-D) or proxy depression measures. Stress and depression measures were standardized within each study by transforming to z-scores. The three-way interaction was tested in a random intercepts model adjustment for age/gender. All analyses were stratified by race. Study source was specified as a random effect. No significant gender×gene×stress interaction was observed for any phenotype, but a marginally significant effect of age×gene×stress was observed on total cholesterol levels ($P = 0.089$) and LDL ($P = 0.044$) in whites. At older ages, the Val/Val carriers had higher total cholesterol (TCHOL) and LDL with increased stress levels compared with Met carriers, while at younger ages, the Val/Val carriers had lower TCHOL and LDL with increased stress levels compared with Met carriers. The further two-way interaction of gene×stress was not significant when stratified by race or gender. The association between stress and phenotypes was strong and consistent. Overall there were no significant three-way or two-way interactions after multiple testing correction, and no age or gender or genotype effect was associated stress-related risk factors in this large harmonized sample. The results of this study suggest that in a large individual participant data meta-analysis, previously observed relationships of the *BDNF* Val66Met were not replicated.

3191F

Mega-analysis of the *EBF1* GxE association: Evaluation of differences in gene-by-stress interaction across race, sex, and age in harmonized datasets of over 27,000 individuals. A. Singh^{1,2,3}, M.A. Babyak^{1,2}, B.H. Brummett^{1,2}, W.E. Kraus^{3,4}, I.C. Siegler^{1,2}, E.R. Hauser^{3,5}, R.B. Williams^{1,2}. 1) Behavioral Medicine Research Center, Duke University School of Medicine, Durham, NC, USA; 2) Department of Psychiatry and Behavioral Sciences, Duke University School of Medicine, Durham, NC, USA; 3) Duke Molecular Physiology Institute, Duke University School of Medicine, Durham, NC, USA; 4) Department of Medicine, Duke University School of Medicine, Durham, NC, USA; 5) Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, NC, USA.

Developing robust estimates of the interactions among individual's genetic makeup and complex physiological aspects of race, sex, and age is critical in a precision medicine framework. In our prior work (Singh et al., *EJHG*, 2015) we identified a novel CVD-risk gene *EBF1* in the Multi-Ethnic Study of Atherosclerosis (MESA) White samples, where a common variant (rs4704963) contributed to inter-individual differences in central obesity (i.e., hip circumference, HIPCM) in the presence of chronic psychosocial stress. We replicated this association in multiple studies using synthetic stress measure derived by our algorithm based on proxy indicators of stress (Singh et al., *Gen. Epi*, 2015). Further, we observed this association only in female and predominantly in older White samples, suggesting race, sex, and age related differences. We now extend this prior work by testing for the *EBF1* SNPxSTRESS interaction with respect to HIPCM in a mega-analysis (i.e., joint analysis of the pooled samples), using harmonized data from 27,249 participants derived from nine studies, with a focus on formal testing of differences across race, sex, and age. We tested the 4-ways interaction term including SNP, stress, sex and race, which was not statistically significant at $p < 0.05$. We next fit the 3-ways interactions SNPxSTRESSxSEX, SNPxSTRESSxRACE, and SNPxSTRESSxAGE in regression models. Out of all these 3-ways interactions, only SNPxSTRESSxSEX resulted in a statistically significant association with HIPCM ($p = 9.48 \times 10^{-4}$). In order to better describe the differences among race, sex, and age with respect to the *EBF1* GxE association, we stratified the analysis by race, sex, and age (using bins age ≤ 44 and age > 44 , based on the change in distribution of HIPCM over age). The SNPxSTRESS term p -values for all Whites (2.47×10^{-7}) and all Blacks (0.01), White Females (1.33×10^{-7}), Black Females (0.036), White Females of age > 44 (1.26×10^{-7}), and Black Females of age > 44 (0.046) were statistically significant. The p -values of the same interaction term for White Males (0.21), Black Males (0.20), White Females of age ≤ 44 (0.28), and Black Females of age ≤ 44 (0.96) were not statistically significant. We also observed the sex difference in stress main effect. Our work provides additional confirmation that common variation in *EBF1* contributes to inter-individual differences in human obesity in the presence of stress, and that the SNPxSTRESS interaction differs depending on the sex of participants.

3192W

Multi-ancestry genome-wide meta-analysis incorporating gene-sleep duration interactions in 62,970 individuals identified potential novel loci for blood pressure. H. Wang^{1,2}, J. Lee^{1,2}, R. Noordam³, D. Mook-Kanamori^{4,5}, S. Redline^{1,2}, CHARGE Gene-Lifestyle Interactions working group. 1) Division of Sleep and Circadian Disorders, Brigham and Women's Hospital, Boston, Massachusetts, USA; 2) Division of Sleep Medicine, Harvard Medical School, Boston, Massachusetts, USA; 3) Department of Internal Medicine, Section Gerontology and Geriatrics, Leiden University Medical Center, Leiden, the Netherlands; 4) Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, the Netherlands; 5) Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, the Netherlands.

Long and short sleep duration are associated with elevated blood pressure (BP), possibly through effects on molecular pathways that influence neuro-endocrine and vascular systems. We hypothesized that differences in sleep duration may modify the effect of genetic factors on BP. Therefore, to gain new insights into the genetic basis of sleep-related BP variation, we conducted multi-ancestry genome-wide gene by sleep duration interaction analyses on four BP traits (systolic BP, diastolic BP, mean arterial pressure, and pulse pressure). Short and long sleepers were defined as cohort-specific 20% and 80% quantiles of age- and sex- adjusted sleep duration residuals. Analyses were performed with and without adjustment for BMI to investigate biological pathways independent of obesity. The primary analysis included 62,970 individuals, of which 40,385 were of European and 15,021 were of African ancestry, from 28 cohort studies. In multi-ancestry analysis, we identified 15 known and 20 potentially novel loci for BP ($p < 5 \times 10^{-8}$), of which 3 loci displayed significant interactions with long sleep duration (*PODXL*, *C20orf96*, *CYYR1*; $p < 5 \times 10^{-8}$). Secondary ancestry-specific analyses identified 9 known and 59 potentially novel loci in individuals of African ancestry and 4 known and 7 potentially novel loci in those of European ancestry. Novel loci overlapped genes enriched for expression in the hypothalamus, including several genes previously associated with cognitive and psychological traits (*CNTN5*, *PTPRN2*, *ASTN2*, *ROBO1*, etc.). These data suggest that long sleep duration, as a direct exposure or marker of other disorders, may modify genetic susceptibility to hypertension. Replication analyses using independent samples are ongoing.

3193T

Genetic association studies of small vessel ischemic stroke stratified by APOE ϵ 2/3/4 genotype. J. Chung^{1,2}, T. Jaworek³, S. Marini^{1,2}, B.D. Mitchell^{3,4}, J. Rosand^{1,2}, C.D. Anderson^{1,2}, International Stroke Genetics Consortium. 1) Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA; 2) Program in Medical and Population Genetics, Broad Institute, Boston, MA, USA; 3) Department of Medicine, Program for Personalized and Genomic Medicine, University of Maryland School of Medicine, Baltimore, MD, USA; 4) Geriatrics Research and Education Clinical Center, Baltimore Veterans Administration Medical Center, Baltimore, MD, USA.

Background Alzheimer's disease (AD) and cerebral small vessel disease (CSVD) share multiple clinical risk factors and pathological features. Apolipoprotein E (APOE) ϵ 4 genotype is a well-known genetic risk factor for AD as well as for CSVD. A recent genome-wide association study (GWAS) in AD identified distinct genetic associations after stratification by APOE genotypes. We performed a similar GWAS for small vessel stroke (SVS), an acute manifestation of CSVD, stratified by APOE genotypes under the hypothesis that a similar novel result could be expected. **Methods** We analyzed samples from 7 genotyping platforms in the NINDS Stroke Genetics Network (NINDS-SiGN) of 2,436 subjects (145 cases; 2,291 controls) with APOE genotypes ϵ 22 or ϵ 23 (ϵ 2+), 15,819 subjects (1,350 cases; 14,469 controls) with ϵ 33, and 5,317 subjects (396 cases; 4,921 controls) with ϵ 34 or ϵ 44 (ϵ 4+). After imputation to the Haplotype Reference Consortium, variants were tested for association in logistic regression against SVS status, adjusting for age, sex, and population structure. GWAS of SVS risk within APOE genotypes ϵ 2/3/4 were conducted in each platform, with results combined by meta-analysis. **Results** We observed genome-wide significant associations at four novel loci including *ISL1* (rs115984524, minor allele frequency [MAF]=2.7%, $P=1.3 \times 10^{-8}$, heterogeneity $I^2=75.7$) and *UNC13C* (rs79250948, MAF=1.4%, $P=3.1 \times 10^{-8}$, $I^2=83.4$) in the APOE ϵ 2+ subgroup, *VN1R95P* (rs402464, MAF=1.6%, $P=1.0 \times 10^{-8}$, $I^2=80.6$) in the ϵ 3, *KCNA4* (rs72883710, MAF=7.2%, $P=1.2 \times 10^{-8}$, $I^2=23.6$) in the ϵ 4+. Cross-referencing these associations in GTEx, rs115984524, rs79250948 and rs72883710 are significant cis-eQTL for *ISL1* ($P=0.02$ in pituitary), *UNC13C* ($P=0.02$ in artery aorta) and *KCNA4* ($P=0.007$ in adipose subcutaneous), respectively. **Conclusion** Our GWAS of SVS stratified by APOE ϵ 2/3/4 alleles reveals several novel loci, as was the case for analogous analyses in AD. Among those, *ISL1*, related to insulin gene regulation and mature-onset diabetes, seems particularly promising given the emerging importance of glucose dysregulation in SVD pathogenesis. High heterogeneity was observed among some findings but may be due to small sample size. We have begun to replicate these associations in independent datasets across SVD traits in the International Stroke Genetics Consortium.

3194F

Hyperuricemia is an early onset metabolic disorder causally associated with cardiovascular disease event in Han Chinese. K.M. Chiang¹, Y.C. Tsay², D.C. Wu², H.C. Yang², Y.T. Huang², C.H. Chen², W.H. Pan¹. 1) Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan; 2) Institute of Statistical Science, Academia Sinica, Taipei, Taiwan.

Background: Serum uric acid (SUA) has been gradually recognized as a potential risk factor for cardiovascular disease (CVD) in recent decades. However, whether the relationship is causal remains a topic of controversy. **Methods:** In this study, we employed two methods to show the importance of SUA in the pathogenesis of CVD development. Firstly, we performed an onset sequence study to examine the onset sequence of hyperuricemia in relation to five cardiometabolic diseases. Secondly, we carried out a Mendelian-randomization (MR) study, a nature's randomized trial, to make causal inference for the SUA-CVD relationship based on the weighted genetic risk score (wGRS) constructed from the SUA-associated single nucleotide polymorphisms (SNPs). The information and SNPs collected from the Cardiovascular Disease Risk Factors Two-Township Study (N=4,330) and Taiwan Biobank (n=10,000) was used, respectively. **Results:** The results of onset sequence study show that both of susceptible males and females tended to have onsets of hyperuricemia, hypo- α -lipoproteinemia in young adulthood, followed by other cardiometabolic diseases. For the MR analysis, the high SUA-wGRS group had a significantly elevated cumulative lifetime risk of CVD, compared to the low group (High vs. Low: odds ratio, OR=1.62, CI: 1.17-2.23, $p=0.003$) with the adjustment for sex (Male vs. Female: OR=2.85, CI: 1.99-4.07, $p<0.001$). Similar results were obtained using wGRS derived from other population's SUA influential SNPs. **Conclusions:** In this study, we showed that hyperuricemia is an earlier onset metabolic disorder compared to hypertension, hypertriglyceridemia and diabetes, indicating high SUA may play an upstream role in the cardiometabolic disease development. Our results from MR study, support a causal role of hyperuricemia on the CVD development.

3195W

Genetic association of arterial stiffness index with incident coronary artery disease and congestive heart failure. S.M. Zekavat^{1,2,3,4,5}, M. Haas², K. Aragam^{2,4,5}, C. Emdin^{2,6}, A.V. Khera^{2,4,5,7}, D. Klarin^{2,4,5}, H. Zhao^{3,8}, P. Natarajan^{2,4,5,7}. 1) Yale School of Medicine, New Haven, CT; 2) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA; 3) Computational Biology & Bioinformatics Program, Yale University, New Haven, CT; 4) Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA; 5) Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA; 6) Harvard Medical School, Boston, MA; 7) Department of Medicine, Harvard Medical School, Boston, MA; 8) Department of Biostatistics, Yale School of Public Health, New Haven, CT.

Arterial stiffness index (ASI) is a noninvasive, rapid measurement extracted from finger infrared analysis. ASI has been independently associated with cardiovascular disease risk in multiple epidemiological studies; however, it is unknown whether these associations represent causal relationships. Here, we used Mendelian randomization to determine whether a genetic predisposition to increased arterial stiffness is associated with increased risk for incident CHF and CAD. We first performed a genome-wide association analysis of arterial stiffness index in 131,686 participants from the UK Biobank. Using these results, we developed a 744-variant polygenic risk score that robustly associated with ASI ($P < 1 \times 10^{-300}$, F-statistic=10,020), and tested its association with incident CHF and CAD through Mendelian randomization. Across nearly 400K participants of the UK Biobank and 200K participants of the CARDIOGRAM-plusC4D consortium, we do not find evidence supporting a causal association with incident CAD ($P > 0.05$). However, we do observe evidence supporting a causal association of ASI with incident CHF (HR=1.23 per SD increase in genetically-mediated ASI, $P = 5.7 \times 10^{-3}$), independent of other cardiometabolic risk factors. Furthermore, we find that the association with incident CHF is present only among individuals without hypertension (HR=1.52 per SD increase in genetically-mediated ASI, $P = 1.3 \times 10^{-4}$) and not among those with hypertension ($P > 0.05$) ($P_{\text{interaction}} = 9.6 \times 10^{-3}$). These results are consistent with a causal association between ASI and CHF but not between ASI and CAD, and support ASI as an orthogonal clinical tool particularly for individuals without hypertension. Furthermore, these data suggest that reducing ASI, particularly among those without hypertension, may reduce risk for CHF.

3196T

Genome-wide Mendelian randomization under pervasive pleiotropy.

Q. Zhao¹, J. Wang¹, J. Bowden², G. Hemani², Y. Chen³, G. Davey Smith², N. Zhang¹, D.S. Small¹. 1) University of Pennsylvania, Philadelphia, PA; 2) University of Bristol, Bristol, UK; 3) University of Michigan, Ann Arbor, MI.

Mendelian randomization (MR) uses genetic variants as instrumental variables (IV) to estimate the causal effect of biomarkers on complex traits. The recent "omnigenic" model suggests that, for most complex traits, there are extremely large numbers of weak genetic signals spread across the genome. This poses two major challenges to MR: most genetic instruments for the target biomarker are expected to be weak, and it is highly likely that they influence the trait via other causal pathways (aka horizontal pleiotropy) and thus violate a core IV assumption—exclusion restriction. We propose a principled statistical method, MR-RAPS (Robust Adjusted Profile Score), that allows the utilization of all instruments across the genome, both weak and strong in magnitude. MR-RAPS requires GWAS summary data and consistently estimates the causal effects under pervasive but balanced pleiotropy. It is also robust to invalid IVs with large direct effects on the outcome as long as they are sparse. Through extensive validation studies, we find MR-RAPS to be much more powerful and robust than existing MR methods, especially in scenarios with many weak genetic IVs. We further propose a comprehensive framework, GRAPPLE (Genome-wide mR Analysis under Pervasive PLEiotropy), for causal discoveries using MR-RAPS. Through an integrated workflow, GRAPPLE allows the identification and adjustment of multiple unbalanced pleiotropy pathways. Through mode analysis of the RAPS likelihood, GRAPPLE also identifies the causal direction. Finally, GRAPPLE allows for sample overlap in the GWAS cohorts. We apply GRAPPLE to public GWAS summary data to screen for causal relationships between blood lipids, BMI, CRP (C-reactive protein), and a panel of disease outcomes. In real data scenarios where the causal relationships are known, GRAPPLE reaffirms the relationships and substantially improves the statistical power and diagnostics. Our new findings include: 1. There are multiple pleiotropic pathways between CRP and CAD (coronary artery disease); 2. HDL cholesterol has a statistically significant protective effect on type 2 diabetes and CAD; The last finding of HDL and CAD is consistent with the inverse association long witnessed in observational epidemiology, but contrasts sharply with the null findings in most existing MR studies of HDL. This suggests that the recent controversy of the HDL hypothesis may be due to insufficient statistical power when only strong SNPs are used in MR analyses.

3197F

NOTCH3 p.R544C mutation is a common and significant risk factor for ischemic stroke in Taiwanese population. Y.C. Liao^{1,2}, C.P. Chung^{1,2}, M.H. Chang³, Y.C. Lee^{1,2}. 1) Taipei Veterans General Hospital, Taipei, Taiwan; 2) Department of Neurology, National Yang-Ming University School of Medicine, Taipei, Taiwan; 3) Department of Neurology, Taichung Veterans General Hospital, Taipei, Taiwan.

Objective Mutations in *NOTCH3* gene cause cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL), which was previously considered a rare familial disease with the prevalence of 2-5 cases per 100,000. A recent study showed that the frequency of cysteine-altering *NOTCH3* mutations in public exomic database is 100-fold higher than that expected. We previously found 70.5% of the CADASIL pedigrees in Taiwan was attributed to the *NOTCH3* p.R544C mutation and approximately 3% of the CADASIL patients carried a homozygous p.R544C mutation, suggesting that this mutation may be prevalent in Taiwanese population. The present study aims at determining the frequency of *NOTCH3* p.R544C mutation in Taiwanese general population and investigating its influence on ischemic stroke. **Methods** The study participants included 6647 anonymous healthy subjects from Taiwan Biobank, 490 stroke-free individuals enrolled from Taipei Veterans General Hospital, and 815 ischemic stroke patients recruited from two medical centers in Taiwan. Genotyping of *NOTCH3* p.R544C mutation was performed using Applied Biosystems TaqMan technology. The genetic effect of *NOTCH3* p.R544C mutation was evaluated by multivariate regression model with adjustment of traditional vascular risk factors. The Trial of Org 10172 in Acute Stroke Treatment (TOAST) classification was used to define stroke subtypes. **Results** Sixty-one of the 6647 community controls was found to carry the *NOTCH3* p.R544C heterozygous mutation, suggesting 0.9% of Taiwanese general populations harboring this genetic risk factor. From the 490 stroke-free, cognitive-intact individuals recruited from hospital, four were found to carry the p.R544C mutation and all of them had diffuse leukoencephalopathy on magnetic resonance images of brain. The *NOTCH3* p.R544C mutation was detected in 16 of the 815 (1.96%) ischemic stroke patients, including one with cardioembolic stroke and 15 with small vessel occlusion. *NOTCH3* p.R544C mutation conferred a 6.89-fold risk for lacunar infarct and a 2.18-fold risk for overall stroke. **Conclusions** *NOTCH3* p.R544C mutation is a common and significant risk factor for ischemic stroke in Taiwanese population. This genetic risk factor is not only detected in familial stroke patients, but also present in one percent of Taiwanese population. Screening the *NOTCH3* p.R544C mutation in asymptomatic persons may help primary stroke prevention via blood pressure control and smoking cessation.

3198W

An extreme phenotype approach identifies novel genes associated with carotid plaque in a multi-ethnic cohort. B. Doliner¹, N. Dueker², S.H. Blanton², L. Wang², C. Dong¹, H. Gardner¹, R. Sacco^{1,2}, T. Rundek^{1,2}. 1) Department of Neurology, University of Miami, Miami, FL; 2) Department of Human Genetics, University of Miami, Miami, FL.

Identifying subclinical cardiovascular disease confers a benefit to patients in the absence of traditional risk factors. Carotid plaque is an accepted marker of subclinical atherosclerosis and predictor of cardiovascular events. Specific genetic variants contributing to plaque have not been well established. We previously identified linkage and association of common variants (CVs) on four chromosomal regions with total carotid plaque area (TCPA). However, identified variants account for only a portion of the total heritability. We aimed to identify additional CVs contributing to TCPA using an extreme phenotype approach in a multi-ethnic cohort. To identify TCPA extremes, the residual TCPA for each participant was determined by regressing TCPA on significant known risk factors. Individuals in the upper and lower 10% (Hispanics; n=180) or 20% (non-Hispanic blacks, n=94 and whites, n=60) threshold of the residual distributions were included in this study. Participants were genotyped using the Affymetrix 6.0 SNP array and these data were imputed to the 1000 Genomes phase 1, version 3 reference panel. Using logistic regression, adjusting for principal components of ancestry, a genome-wide association study was performed separately in each of the three race/ethnic groups and results were meta-analyzed using METAL. Genome-wide gene-based analyses were performed using VEGAS2 and overrepresentation enrichment pathway analysis was implemented in WebGestalt 2.0. Among SNP based meta-analyses, nine SNPs were suggestively associated with extreme TCPA ($p < 0.0001$). The top SNP, rs10779703, was an intronic variant in *RERE* ($p = 2.0 \times 10^{-6}$). In gene-based meta-analyses, 19 genes were suggestively associated with extreme TCPA ($p < 0.001$) and pathway analyses revealed an enrichment for the bitter taste receptors (TAS2Rs) among these genes. This pathway consisted of two G-protein coupled receptors, *TAS2R50* ($p = 4.9 \times 10^{-4}$) and *TAS2R20* ($p = 6.9 \times 10^{-4}$), as well as a downstream component of the signaling cascade, *ITPR3* ($p = 7.9 \times 10^{-4}$), known to modulate intracellular calcium release. TAS2R agonists induce vasodilation in vascular smooth muscle (VSM) and have anti-mitogenic properties in airway smooth muscle by downregulating growth factors and promoting apoptosis in Bnip-3 linked pathway. VSM proliferation is a core component of atherogenesis and, as such, variation in bitter taste receptors may influence atherosclerosis, making taste receptor genes promising candidates for carotid plaque.

3199T

Genetics of cardiovascular disease risk factors in the Tsimané. H.M.

Highland¹, B. Trumble², A.E. Justice³, R. Bedi⁴, R.C. Thompson⁵, L.S. Wann⁶, C.J. Rowan⁷, A.H. Allam⁸, M.I. Miyamoto⁹, D.E. Michalik¹⁰, B. Frohlich¹¹, M.L. Sutherland¹², J.D. Sutherland¹³, G.P. Lombardi¹⁴, C.E. Finch¹⁵, J. Narula¹⁶, G.S. Thomas¹⁷, J. Stieglitz¹⁸, M. Gurven¹⁹, K.E. North¹, H. Kaplan²⁰. 1) Department of Epidemiology, University of North Carolina, Chapel Hill, NC; 2) School of Human Evolution and Social Change, Arizona State University, Tempe, AZ; 3) Biomedical and Translational Informatics Institute, Geisinger Health System, Danville, PA; 4) Department of Bioengineering, University of Washington, Seattle WA; 5) St. Luke's Mid America Heart Inst, Kansas City, MO; 6) Ascension Healthcare, Milwaukee, WI; 7) Renown Inst for Heart and Vascular Health, Reno, NV; 8) Al Azhar Univ, Cairo, Egypt; 9) Providence St. Joseph Health, Mission Viejo, CA; 10) Miller Children's Hosp Long Beach, Long Beach, CA; 11) Smithsonian Institution, Washington, DC; 12) Newport Diagnostic Ctr, Newport Beach, CA; 13) South Coast Radiological Medical Group, Laguna Hills, CA; 14) University Peruana Cayetano Heredia, Lima, Peru; 15) Dornsife College, University of Southern California, Los Angeles, CA; 16) Icahn School of Med at Mount Sinai, New York, NY; 17) MemorialCare, Fountain Valley, CA; 18) Université Toulouse 1 Capitole, France; 19) University of California at Santa Barbara, Santa Barbara, CA; 20) Chapman University, Orange, CA.

To date, majority of participants included in genome wide association studies (GWAS) are of European descent (EUD). American Indian (AI) populations remain underrepresented in GWAS, including lipid levels, inflammatory markers, and subclinical cardiovascular disease (CVD). The Tsimané are a Bolivian AI population of forager-horticulturalists living a traditional lifestyle similar to that experienced throughout human pre-history. This population exhibits low-density lipoprotein cholesterol (LDL) level similar to the target levels of current statin and add-on pharmacological interventions, such as PCSK9 inhibitors. Given their genetic architecture and unique CVD risk profile, this population provides a unique opportunity to study the genetic underpinnings of CVD health. Admixture estimates are 97% Native American, <1% African, and 3% European ancestry on average. All samples were genotyped on the Illumina MEGA array and imputed to 1000 genomes phase 3. We performed GWAS association analysis for serum lipid values (LDL, high-density lipoprotein cholesterol [HDL], triglyceride [TG] and total cholesterol [TC], n=829) and presence of atherosclerotic plaques (carotid or femoral artery, n=303) in 1,214 individuals representing 180 families with 8,602 relative pair types with genotype and imputed data (Illumina MEGA array imputed to both the TOPMed and 1000 Genomes AMR reference panel). We performed linear mixed models, assumed an additive genetic model, adjusted for sex, age, body mass index and ancestry principal components, and accounted for family structure through kinship coefficients. In this subset of data, we identified a novel variant (rs7770909) associated with the presence of plaques in the right carotid artery ($p = 1.7823 \times 10^{-6}$). Further, we replicated many previously identified lipids loci, and have several suggestive associations pending replication. The identification of potential novel CVD-risk loci in the Tsimané underscores the importance of examining diverse populations. We hope these findings, like past investigations in population isolates, will lead to identification of CVD-associated variants that offer protection against CVD.

3200F

Whole genome sequencing identifies multiple rare variants in GT-Pase-activating protein (DLC1) associated with sleep disordered breathing. J. Liang¹, K. He¹, J. Lee², B. Cade², H. Wang², X. Guo³, D. Gottlieb², S. Gharib⁴, J. Rotter⁵, X. Lin⁶, S. Redline², X. Zhu¹ On Behalf of the TOPMed Sleep Working Group.

1) Department of Population and Quantitative Health, Case Western Reserve University, Cleveland, OH; 2) Division of Sleep Medicine, Harvard Medical School, Boston, Massachusetts; 3) Institute for Translational Genomics and Population Sciences and Department of Pediatrics, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, California; 4) Computational Medicine Core, Center for Lung Biology, University of Washington Medicine Sleep Center, Division of Pulmonary, Critical Care and Sleep Medicine, University of Washington, Seattle, Washington; 5) Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute and Department of Pediatrics at Harbor-University of California Los Angeles Medical Center, Torrance, California; 6) Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, Massachusetts.

Average nocturnal oxygen saturation (SaO₂) is a clinically relevant measure of physiological stress associated with sleep disordered breathing that predicts cardiovascular disease and mortality. We report genetic associations of average SaO₂ measured during sleep using the high depth whole genome sequencing data from the NHLBI TOPMed project. To improve power, we focused our analyses on a 21 Mb linkage region identified in European Americans from the Cleveland Family Study (CFS). We prioritized the low frequency and rare variants that contributed to the observed linkage evidence. Non-coding variants were filtered using CADD scores >10, reflecting in silico pathogenicity. We identified 12 candidate genes ($P < 0.05$) associated with average nocturnal SaO₂ in this region. Follow-up analysis of these variants in the 12 candidate genes in 5 independent samples enrolled in the TOPMed project identified 6 coding and 51 noncoding variants in GTPase-activating protein (DLC1) significantly associated with average nocturnal SaO₂ after Bonferroni correction in gene based tests ($P = 0.021$ for coding and $P = 2.11 \times 10^{-6}$ for noncoding variants). The noncoding variants were significantly enriched in regulatory features in the Human Lung Fibroblast cell line ($P = 0.003$) and significantly contribute to the DLC1 expression level in human skin cells-transformed fibroblasts ($P = 2.6 \times 10^{-4}$). The polygenic score built on the effect sizes estimated from the independent samples explained 0.97% of the variation in SaO₂ levels and a significant portion of the observed linkage evidence in CFS ($P < 0.001$). Although DLC1 has been reported to be associated with lung function, our analysis demonstrated that the low and rare variant association of DLC1 with average nocturnal SaO₂ persisted after adjusting for spirometric measures of lung function, suggesting pleiotropic effects. This analysis highlights that although individual effect sizes of rare and low-frequency variant are often small, when aggregated, they may account for substantial variation of polygenic human phenotypes such as SaO₂, and these variants can be uncovered by family cohorts with moderate sample size.

3201W

Investigation of novel variations of *ORAI1* gene and their association with Kawasaki disease. K. Thiha¹, Y. Mashimo¹, H. Suzuki², H. Hamada³, A. Hata¹, T. Hara⁴, T. Tanaka⁵, K. Ito⁶, Y. Onouchi^{1,6}. 1) Department of Public Health, Chiba University Graduate School of Medicine, Chiba, Japan; 2) Department of Pediatrics, Wakayama Medical University, Wakayama, Japan; 3) Department of Pediatrics, Tokyo Women's Medical University, Yachiyo Medical Center, Yachiyo, Chiba, Japan; 4) Fukuoka Children's Hospital and Medical Center for Infectious Diseases, Fukuoka, Japan; 5) Department of Human Genetics and Disease Diversity, Tokyo Medical and Dental University (TMDU), Tokyo, Japan; 6) Laboratory for Cardiovascular Disease, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan.

Kawasaki disease (KD; MIM#611775) is an acute systemic vasculitis syndrome which predominantly affects children younger than 5 years and is the leading cause of acquired heart disease of childhood in developed countries. Genome-wide studies have identified several common susceptibility variants for KD but the involvement of rare variants in KD pathogenesis is largely un-investigated. In this study we conducted a large scale association study for rare variants of *ORAI1* gene in which one common and one rare variants have been reported to confer KD. The study included 3,900 KD and 2,645 controls to determine *ORAI1* protein coding region using direct sequencing. The scoring for deleteriousness of the single variants as well as insertion/deletion variants were predicted by Combined Annotation Dependent Depletion (CADD). The Collapse method was applied for variants with low allele frequencies for which ≥ 0.75 of statistic power to detect association with $P < 0.05$ under assumption of odds ratio of 3.0 was not expected. Association of the individual as well as collapsed multiple rare variants with KD were assessed by Fisher's exact test. Thirty-four variants (5 common ones with minor allele frequencies (MAF) ≥ 0.05 and 29 rare ones with MAF < 0.01) were identified within 903 bases of *ORAI1* coding region. The 27 rare variants were previously not examined and had no association with KD either as a single variant or as a single collapsed one. Interestingly, six rare deleterious missense variants (MAF < 0.001 , C-score ≥ 20) including p.Gly98Asp which has been proved to lead constitutive Ca^{2+} entry to the cell were exclusively seen in KD patients (OR = ∞ , $p = 0.046$). Conversely, frameshift variants ($n = 5$), identified near the N-terminus and assumed to disrupt *ORAI1* function, showed the opposite trend of association (OR = 0.35, $p = 0.24$). These findings suggest involvement of rare variants of *ORAI1* into KD pathogenesis and further support our hypothesis that upregulation of Ca^{2+} /NFAT pathway underlies KD inflammation. Precise information about the functional significance (i.e. neutral, disruptive or gain-of-function) might be helpful in assessing association in the gene-based collapsing method.

3202T

Loss-of-function variant in *ANGPTL4*, *ANGPTL8* and risk of coronary and metabolic disease. P. Helkkula¹, I. Surakka^{1,3}, M.T. Kinnunen¹, A.S. Havulinna^{1,4}, T. Kiiskinen⁵, O.T. Raitakar⁶, T. Lehtimäki⁶, J. Eriksson⁷, M. Männikkö⁸, S. Koskinen⁴, V. Salomaa⁴, H. Laivuori^{9,10,11}, E. Widen¹, M.J. Daly^{1,12,13}, A. Palotie^{1,13,14}, S. Ripatti^{1,2}. 1) Institute for Molecular Medicine Finland, Helsinki, Finland; 2) Department of Public Health, University of Helsinki, Helsinki, Finland; 3) Department of Internal Medicine, University of Michigan, Ann Arbor, MI, United States; 4) National Institute for Health and Welfare, Helsinki, Finland; 5) Department of Clinical Physiology and Nuclear Medicine, University of Turku and Turku University Hospital, Turku, Finland; 6) Department of Clinical Chemistry, Fimlab Laboratories, Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland; 7) Department of General Practice and Primary Health Care, University of Helsinki and Helsinki University Hospital, Helsinki, Finland; 8) Center for Life Course Health Research, Faculty of Medicine, University of Oulu, Oulu, Finland; 9) Medical and Clinical Genetics, University of Helsinki and Helsinki University Hospital, Helsinki, Finland; 10) Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland; 11) Department of Obstetrics and Gynecology, Tampere University Hospital, Tampere, Finland; 12) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; 13) The Stanley Center for Psychiatric Research and Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA; 14) Analytic and Translational Genetics Unit, Departments of Medicine, of Neurology and of Psychiatry, Massachusetts General Hospital, Boston, MA, USA.

Population isolates, such as Finland, show an enrichment of loss-of-function (LoF) variants with potential impact on phenotypes and disease risk. We tested if some of the LoF variants were associated with blood lipids in Finns. Our study consisted of 2,017 autosomal LoF variants with MAF between 1.0×10^{-3} and 0.1 in 1,781 genes. We tested their association with LDL-C, HDL-C, TC and triglycerides in the population-based FINRISK cohort ($N = 25,414$) with genome-wide genotyping and imputation using a Finnish imputation reference panel. We found high-effect lipid-modulating LoF variants in four genes without previously reported studies of human knockouts. Protein-disrupting variants in *ANGPTL4* (rs746226153, MAF = 4.8×10^{-3} , effect = -0.30 mmol/L, $P = 7.9 \times 10^{-10}$) and *ANGPTL8* (rs760351239, MAF = 1.9×10^{-3} , effect = -0.48 mmol/L, $P = 4.2 \times 10^{-8}$) were associated with lower triglyceride levels. Angptl4 and Angptl8 are LPL-inhibiting proteins, with an essential role in triglyceride trafficking during fasting and feeding respectively. The triglyceride-lowering effect of the *ANGPTL4* and *ANGPTL8* LoF mutations was stronger in individuals with longer and shorter fasting times respectively, which is in line with murine models. We also observed LoF variants with substantial HDL-increasing effect in *LIPG* (rs200435657, MAF = 1.7×10^{-3} , effect = 0.10 mmol/L, $P = 2.7 \times 10^{-13}$) and *CETP* (rs751916721, MAF = 1.5×10^{-3} , effect = 0.19 mmol/L, $P = 9.0 \times 10^{-34}$). Pooling individuals from five Finnish population-based cohorts ($N = 38,618$) we observe a protection against coronary artery disease by the *ANGPTL4* (OR = 0.57, $P = 8.2 \times 10^{-3}$) and *ANGPTL8* (OR = 0.24, $P = 6.1 \times 10^{-3}$) LoF alleles. The *ANGPTL8* truncating variant was also associated with lower risk of diabetes (OR = 0.47, $P = 0.042$). In a phenome-wide scan across EHR data, the *ANGPTL4* LoF allele associated with an increased risk of lymphatic disorders: Non-Hodgkin lymphoma (OR = 3.04, $P = 0.018$) and lymphoma (OR = 2.77, $P = 0.028$). The therapeutic safety of *ANGPTL4* inhibition has been questioned due to lymphatic lipid accumulation in animals. A phenome-wide screen of the *ANGPTL8* LoF mutation showed increased risk of multiple metabolic and psychiatric endpoints, including thyroid gland disorders (OR = 3.60, $P = 1.9 \times 10^{-4}$), inflammatory liver disease (OR = 8.40, $P = 0.0119$) and depression (OR = 2.20, $P = 0.018$). Our study shows the potential in combining genome data with rich nationwide health registry data in an isolated population such as Finland to learn about potential drug targets and their possible safety.

3203F**Bayesian whole genome association analysis of gene-age interactions.**

M. Kerin^{1,2}, *K. Sharp*¹, *J. Marchini*^{1,2}. 1) University of Oxford, Oxford, Oxfordshire, United Kingdom; 2) Wellcome Trust Center for Human Genetics, University of Oxford.

Age is often included in genetic association studies as an important explanatory factor, but few studies have investigated interactions between age and genetic variants, and whether this can explain a fraction of the heritability in a given trait. We have developed a Bayesian whole genome regression approach that jointly models the main effects of each SNP as well as multiplicative SNP-age interaction effects. Our model has separate sparse spike and slab priors on the effect sizes of main effects and interaction effects. We use Variational Bayes to fit the model and use a mixture of exact and gradient-based updates to optimize over the hyper-parameters. This avoids having to search over a grid of hyper-parameter values. The method estimates the proportion of variance explained (PVE) by main effects and by SNP-age interactions, as well as identifying loci with non-zero main and interaction effects. The method is implemented in C++ using multi-threading to increase computational efficiency. The computational complexity of the method is linear in sample size and number of SNPs. On the simulated data with 10,000 individuals at 100,000 SNP loci the method takes 30 minutes to run to convergence using 5 cores. When data is simulated PVE by main effects and by SNP-age interactions of 0.4 and 0.1, our method produces estimates of 0.398 (se 0.003) and 0.099 (se 0.001) respectively. We applied the method to a genetic association study of Pulse Pressure in 342,673 participants at 721,836 genotyped SNPs from the UK Biobank study. Running on a single core, our analysis takes a median of 83 hours and uses 240GB of RAM. The method estimates that ~1% of SNPs have a non-zero main effect and that the PVE by main effects is 11.69%. For SNP-age interactions our method infers a PVE of only 0.12%. However, our method infers 8 loci with a posterior probability of inclusion (PIP) above 0.9. Using a simpler GWAS approach that analyzes each SNP independently, we find that all of these 8 SNPs show genome-wide significant main effects, but 3 of the 8 SNPs (rs4717862, rs4908857 and rs35688424) do not show significant SNP-Age interaction effects. We have observed that two of these SNPs (rs4717862 and rs35688424) are near genes (*ELN* and *MRC2*) that play a role in extracellular matrix remodelling.

3204W**A statistical method for testing pleiotropy using GWAS summary statistics.** *X. Li, X. Zhu.* Department of Population and Quantitative Health Sciences, School of Medicine, Case Western Reserve University, Cleveland, OH.

Genome-wide association studies (GWASs) have identified many genetic loci which appear to harbor variants that are associated with multiple traits, which terms a cross-phenotype (CP) association. CP association can be caused by biological, mediated or spurious pleiotropy. In this study we propose a two stage method using summary statistics from GWAS to differentiate biological, mediated and spurious pleiotropy. At stage 1 we separate spurious pleiotropy from the others. At stage 2 we further separate mediated from biological pleiotropy. Our simulations suggest that the proposed method can differentiate different pleiotropies. We applied our method to the summary statistics of height and BMI from GIANT consortium (Yengo et al. 2018 bioRxiv). Among the reported 3,290 and 716 independent common variants for height and BMI, respectively, 302 variants demonstrate either biological pleiotropy or colocalization effects after adjusting for multiple tests ($P < 1.52E-5$), suggesting shared genetic pathways. We also applied the method to the 901 reported systolic, diastolic and pulse blood pressure loci (Evangelou et al. 2017 bioRxiv) using the summary statistics from the International Consortium of Blood Pressure. We observed eight loci with biological pleiotropy/colocalization evidence, including *PKHD1*, *VMP1*, *ULK4*, *HSD52-MIR4711*, *CDKN2B*, *FAM208B*, *CCDC30/CLDN19* and *PDE3A/CACNA1C* ($P < 5.55E-5$). Our study demonstrates that summary statistics from GWAS can further help to understand the inter-phenotype relationship contributed by genetic factors.

3205T

Joint modeling of rare inherited and *de novo* variants identifies congenital heart disease risk genes. M. Li¹, L. Jin², Q. Lu³, S. Jin⁴, X. Zeng⁴, W. Dong⁴, M. Brueckner⁴, R. Lifton^{4,5}, H. Zhao^{1,6}. 1) Department of Biostatistics, Yale School of Public Health; 2) Department of Molecular, Cellular and Developmental Biology, Yale University; 3) Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison; 4) Department of Genetics, Yale University; 5) Laboratory of Human Genetics and Genomics, Rockefeller University; 6) Program of Computational Biology and Bioinformatics, Yale University.

Recent studies have demonstrated the important role of *de novo* mutations (DNMs) in the etiology of early-onset diseases. The study of DNMs in congenital heart disease (CHD) parent-offspring trios has led to the identification of multiple disease genes. However, gene identification based on DNMs remains challenging due to low mutation rate. With few recurrence of DNMs in the same gene, such studies may be underpowered. In this project, we describe a hierarchical Bayesian framework for gene-level association test which jointly analyzes *de novo* and rare inherited variants. By explicitly modeling the joint probability of multiple types of variants in trios, our method has higher statistical power than analyses based on DNMs only. We incorporated gene-level annotations (e.g. gene length and sequence context) and data from a reference panel (e.g. gnomAD) to accurately characterize the expected frequencies of both *de novo* and transmitted rare variants, which further improved the statistical power of our method. We applied our model to whole-exome sequencing data from 2,645 CHD parent-offspring trios (Jin et al. 2017), and confirmed a number of previous findings. We replicated multiple known CHD genes that were previously identified using DNM enrichment analysis (e.g. *PTPN11*, *KMT2D*, and *CHD7*), as well as a few risk genes that were enriched for rare, transmitted damaging variants (e.g. *GDF1* and *MYH6*). More importantly, our model implicated novel risk genes for CHD. One example is *SULF1*, a gene that is intolerant to loss-of-function variants ($pLI=0.88$), specifically abundant in mouse developing heart, and known to associate with mesomelia-synostosis syndrome. These results suggest that, through joint modeling of rare inherited and *de novo* variants, our method can effectively improve disease gene discovery. Our analysis also identified novel risk genes and provided new insight into the genetic basis of CHD. Our method may benefit future sequencing-based studies in disease trios and accelerate findings of risk genes.

3206F

Multi-phenotype genome-wide association study of protein levels in individuals with pulmonary arterial hypertension. E. Pileckyte¹, M. Kaakinen^{1,2}, C.J. Rhodes², M. Wilkins², I. Prokopenko², NIHR Bioresource for Rare Diseases. 1) Section of Genomics of Common Disease, Department of Medicine, Imperial College London, London, United Kingdom; 2) The Centre for Pharmacology and Therapeutics, Department of Medicine, Imperial College London, London, United Kingdom.

Genome-wide association studies (GWAS) have identified numerous genetic risk factors associated with complex human traits and diseases but large proportion of heritability remains unexplained. To increase the locus discovery power, intermediate phenotypes, e.g. protein levels, and multi-phenotype analyses can be used. We aimed to identify genetic risk variants underlying susceptibility to pulmonary arterial hypertension (PAH), a rare but severe complex disease. We performed standard GWAS and multi-phenotype (MP) GWAS of SOMAscan platform 1,124 blood plasma protein levels in 128 European idiopathic, heritable, or drug-/toxin-induced PAH patients to identify one-variant-one-phenotype and one-variant-multiple-phenotypes associations, respectively. We defined proteins related to cardiovascular, metabolic, inflammatory, pulmonary, immune processes, and stress response, based on SOMAscan protein platform classification. Prior to analysis, each protein level was natural logarithm transformed and adjusted for age, sex, and four principal components to control for population stratification. The resulting residuals were further inverse-normal transformed to assure normality. In standard GWAS, we discovered 27 protein quantitative loci (pQTLs) with genome-wide significance corrected for multiple testing ($P\text{-value}<4.45\times 10^{-11}$), among which four were novel *trans*-pQTLs: at *ELK2AP* for DR3, near *TMEM215* for Properdin, near *MIR4435-1* for C1r, and near *RETNLB* for HGFA. We then performed MP-GWAS using SCOPA software that implements "reverse regression" approach. We analysed a cluster of 26 proteins with pulmonary, stress response, and immune functions, and identified one pQTL at *PDGFRB* with genome-wide significance ($P\text{-value}<5\times 10^{-6}$). The association is driven by proteins Calcineurin alpha isozyme, Cathepsin H, epidermal growth factor receptor (ERBB1), Nectin-like protein 2, and platelet-derived growth factor receptor beta polypeptide (PDGF Rb). The pQTL is novel for the first four proteins and is in *trans*. Calcineurin alpha isozyme, Cathepsin H, and ERBB1 are involved in blood vessel and lung development, and Calcineurin alpha isozyme mutations are associated with cardiac hypertrophy, an abnormal heart muscle enlargement, that can be caused by hypertension. Nectin-like protein 2 and ERBB1 mutations are involved in lung cancer. These large-scale proteomics GWAS in PAH patients identified potentially PAH-specific pQTLs not reported previously in studies of healthy individuals.

3207W

A multi-phenotype analysis of coagulation factor plasma levels identifies genetic variants influencing levels of fibrinogen and Factor XIII. M. Lerro¹, P. Wells², P. Morange³, D. Tregouet^{4,5}, F. Gagnon¹. 1) University of Toronto, Toronto, Ontario, Canada; 2) University of Ottawa, Ottawa, Canada; 3) INSERM, Marseille, France; 4) Sorbonne Universites, Paris, France; 5) ICAN, Paris, France.

Plasma levels of coagulation-related proteins affect risk of venous thromboembolism (VTE). Genetic variants that explain their levels have been identified in analyses using a univariate outcome. We assessed association of genome-wide SNPs with plasma levels of 13 coagulation-related proteins using groups of variables as multivariate outcomes in a sample (n=255) of five French-Canadian pedigrees ascertained on VTE. Testing of associations with multiple outcomes can be more powerful when using correlated traits and it allows fewer statistical tests. Proteins were clustered into correlated groups based on plasma levels and association analyses were performed for each group. Four clusters were identified using a PCA-based approach: 1. vWF and FVIII; 2. FIX, FX, and prothrombin; 3. plasminogen activator inhibitor, FXI, FXII, and tissue factor pathway inhibitor; 4. fibrinogen and FXIII. 489,241 SNPs were tested for association while controlling for age, sex, and family structure. The cluster of FVIII and vWF showed a significant association on chromosome nine at the ABO locus (p = 5E-8), a known predictor, and a suggestive association on chromosome 15 in SRP14-AS1 (p = 6E-7). SRP14-AS1 relates to the exportation of proteins from cells. Fibrinogen and FXIII showed an association on chromosome 6 at BTBD9 (p = 9E-8) and on chromosome 20 within MACROD2 (p=5E-9). The SNP identified in BTBD9 has been previously associated with ischemic stroke. MACROD2 is associated with the presence of anti-phospholipid antibodies. Antiphospholipid syndrome is a condition associated with VTE, and was an exclusion criterion of the FVL study. Additionally, MACROD2 variants have been seen to be associated with plasma levels of MMP7, a protein that is elevated in patients following a VTE event. The genetic associations showed strong signals in single-outcome analyses, but were not genome-wide significant, demonstrating improvements made by this approach. We are currently replicating in a second French sample. We also tested for differential expression of the genes MACROD2 and BTBD9 in blood of a separate VTE case-control study to understand potential mechanisms. BTBD9 was more highly expressed in cases than controls after adjustment for blood cell type proportions (p = 3E-3), and MACROD2 showed under-expression in cases, but this was not significant (p = 9E-2). We plan to further examine the effects of the identified variants by testing their association with nearby DNA methylation.

3208T

Investigating the role of $\beta 4$ GALT-I and $\beta 4$ GALT-I N352S in lipid metabolism and atherosclerosis. G. Della Gatta¹, C. Van Hout¹, M. Montasser², B. Shen³, N. Li³, L. Miloscio¹, G. Tzoneva¹, S. Taylor², E. Streeter², G. Gong², S. Hartford³, D. White³, E. Pefanis³, T. Persaud¹, V. Gusarova³, J. Haller², C. Schoenherr^{1,3}, B. Zambrowicz³, A. Economides^{1,3}, A.R. Shuldiner^{1,2}. 1) Regeneron Genetics Center, Regeneron Pharmaceuticals, Tarrytown, NY 10591, USA; 2) School of Medicine, Division of Endocrinology, Diabetes and Nutrition, and Program for Personalized and Genomic Medicine University of Maryland, Baltimore, MD, 21201, USA; 3) Regeneron Pharmaceuticals, Tarrytown, NY 10591, USA.

The distinctive genomic architecture of founder populations, such as the Old Order Amish (OOA) of Lancaster PA, provides unique opportunities to discover novel genotype-phenotype associations relevant to human health and disease. Exome sequencing (n=5,297) in the OOA identified a variant, c.1055A>G in the $\beta 4$ -galactosyltransferase-I gene corresponding to an asparagine to serine substitution at position 352 of $\beta 4$ GALT-I protein (N352S). While very rare in the general population, the N352S variant drifted to a frequency of 6% in the Amish. $\beta 4$ GALT-I N352S carriers showed: (1) lower LDL-cholesterol (17.3 and 31.2 mg/dL LDL in heterozygotes and homozygotes, respectively (p=1.3x10⁻⁶); (2) lower fibrinogen (p=3.2x10⁻⁵); (3) and decreased coronary artery calcification (p=0.04). $\beta 4$ GALT-I belongs to the beta-1,4-galactosyltransferase family of type II membrane-bound glycoproteins that mostly reside in the trans-Golgi apparatus. $\beta 4$ GALT-I contributes to the synthesis of N-linked glycan chains, transferring galactose from UDP-galactose to N-acetylglucosamine (GlcNAc). The N352S variant resides very close to the active site of the enzyme. To understand the mechanisms underlying the cardioprotective effects of N352S on serum lipids and fibrinogen, we measured by LC-MS N-linked glycan profiles in serum of N352N and N352S homozygotes (n=12 each). Compared to N352N homozygotes, N352S homozygotes had significantly decreased galactosylation of total serum glycoproteins (6.7%), fibrinogen (15.1%) and IgG (25.7%, p<0.0001 for all). Moreover, ELISA assays on the same samples, showed significant decreases of LDL-c, APOB and PCSK9 protein levels (32%, p= 0.0003; 32%, p=0.001; 14%, p=0.07, respectively) as well as an increase of APOA1 levels (26%, p= 0.002) in N352S homozygotes compared to the N352N homozygotes. Finally, preliminary *in vivo* analyses of $\beta 4$ galt-I gene-specific liver KO using Cas9 ready mice and gRNA delivery by AAV8, showed that liver-specific $\beta 4$ galt-I KO leads to lower serum LDL-c (55%, p<0.0001, at 8 weeks from AAV8 injection) compared to Cas9 ready liver knock out control mice. Overall, our data suggest a previously uncharacterized role of $\beta 4$ GALT-I in cholesterol metabolism. Further studies of $\beta 4$ GALT-1 could lead to the discovery of new molecular pathways and therapeutic targets for lipid disorders and cardiovascular disease.

3209F

Inter-species differences in response to hypoxia and oxidative stress in iPSC-derived cardiomyocytes from humans and chimpanzees. *M.C. Ward¹, Y. Gilad^{1,2}*. 1) Department of Medicine, University of Chicago, IL; 2) Department of Human Genetics, University of Chicago, IL.

The basis for complex polygenic diseases such as cardiovascular disease is unclear; however despite anatomical similarities, there appear to be differences in susceptibility between primates. For example, humans are prone to ischaemia-induced myocardial infarction unlike chimpanzees, which tend to suffer from fibrotic disease. At a molecular level, it has been suggested that differences in oxidative stress contribute to inter-species susceptibility to disease. However, it has been difficult to test whether these are true intrinsic inter-species differences due to uncontrolled effects of environmental factors and small sample sizes. The ability to differentiate cardiomyocytes from induced pluripotent stem cells (iPSCs), in both species, now allows for direct inter-species comparisons. In order to understand human-specific regulatory adaptations in the heart, and to gain insight into the evolution of disease susceptibility and resistance, we developed a model of hypoxia and oxidative stress in human and chimpanzee cardiomyocytes. We differentiated eight human and seven chimpanzee iPSC lines, together with six technical replicates, into cardiac troponin-positive iPSC-derived cardiomyocytes under normoxic conditions. We subjected these cells to 6 hours of hypoxia, followed by 6 or 24 hours of re-oxygenation, and collected genome-wide gene expression data, as well as measurements of cellular stress at each time-point. The overall cellular and transcriptional response to hypoxic stress is highly correlated across species with 1,920 genes responding similarly between humans and chimpanzees. 636 genes show a species-specific trajectory during the course of the experiment, while 147 genes show a species-specific response only in the hypoxic state. For example, *RASD1* is up-regulated specifically in human following hypoxia, and is a coronary artery disease GWAS hit of unknown function. Our results provide the first direct molecular measurements following cellular stress in human and chimpanzee cardiomyocytes, and suggest a degree of species-specificity in the transcriptional response in these two closely related species.

3210W

Integrative Bayesian analysis of high-throughput functionalization assays identifies variants controlling the expression of CD36 and other platelet genes. *A. Ghazi¹, N. Madan², L. Edelstein², E. Chen³, C. Shaw²*. 1) Quantitative and Computational Biosciences, Baylor College of Medicine, Houston, TX; 2) Cardeza Foundation for Hematologic Research, Thomas Jefferson University, Philadelphia, PA; 3) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

Most human DNA is non-coding, and recent studies have shown the importance of non-coding variation for gene regulation in both pathological and physiological functions. There is a considerable level of gene expression and phenotypic variation in human platelets that associates with non-coding genetic variants. However, inspecting the function and transcriptional impact of individual variants on regulatory activity with traditional assays is time-consuming and costly. The Massively Parallel Reporter Assay (MPRA) addresses this problem by examining the transcriptional activity of thousands of variants in parallel using next-generation sequencing (NGS). We have used MPRA to study the impact of non-coding regulatory variation present in human megakaryocytes, which give rise to human platelets. The use of MPRA study permits for the first time the examination of platelet eQTLs for functional analysis. However this assay in turn poses large statistical problems: a heavy multiple testing burden, high levels of noise, and experimental data structure that defies the assumptions underlying traditional statistical methods. In order to gain the statistical efficiency needed to address these problems, this project introduces a fully Bayesian negative binomial model of MPRA data that allows for the integration of diverse prior information. We've applied this model to existing MPRA datasets, showing that the new method identifies large numbers of putatively functional variants and provides clear inference on model parameters. We also investigate the ability of a diverse set of external genomic annotations to provide prior information depending on MPRA assay cell types and conditions. Finally, we also show that this variety of Bayesian negative binomial models are readily extensible to other high-throughput NGS-based assays such as CRISPR screens. To pilot our approach, we performed our own MPRA experiment on eighty-one variants in CD36, an important platelet lipoprotein receptor. Using our statistical method, the MPRA results and additional validation experiments lead us to identify a single SNP, rs2366739, that causes a large negative reduction in CD36 expression. An ongoing platelet-focused, large scale MPRA is testing the impact of thousands of variants on many platelet-related genes.

3211T

DIVaH: A variant calling algorithm to identify CRISPR/Cas9 induced mutations in zebrafish. E. Mazzaferro^{1,2}, D. Manyara Wairimu^{1,2}, C. Song^{1,2}, M. den Hoed^{1,2}. 1) Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden; 2) Science for Life Laboratory, Uppsala University, Uppsala, Sweden.

The rapid development and progress of genome editing and sequencing techniques facilitates modeling of diseases and studying of gene function *in vivo*. However, while being an essential step to identify causal genes, haplotype-specific calling of CRISPR/Cas9-induced mutations remains challenging. We developed an algorithm to accomplish this in a computationally efficient manner. The variant calling algorithm DIVaH (Danio rerio Identification of Variant by Haplotype) can identify CRISPR/Cas9-induced mutations from paired-end sequencing of 150 to 400 bp amplicons across multiple targeted sites in a large number of samples haplotype specifically. The algorithm extracts sequencing reads from each BAM file, identifies unique strings, excludes reads that occur fewer than five times or that do not match the reference sequence, and identifies the two most observed sequences at each targeted site. The two strings are then aligned against the reference sequence to identify the mutations present in each of them. i.e. insertions, deletions, or mismatches of one or more nucleotides. Fish are considered homozygous for mutations on the most observed allele if the condition is TRUE, and to be (compound) heterozygous for mutations on the two most prominent alleles otherwise. To test the algorithm, it was applied alongside two validated variant calling algorithms, i.e. VarScan and Freebayes, to paired-end sequencing results from 384 zebrafish larvae with 0 to 2 mutated alleles in 0 to 7 CRISPR/Cas9-targeted proof-of-concept genes for familial hypercholesterolemia (i.e. 6513 BAM files). The script could process each BAM file in an average of 5 secs, and the results show that DIVaH identified 2x more indels than VarScan and Freebayes, with a length of up to 170 bp. The subsequent association analysis showed a significant association of LDL cholesterol levels with the number of mutated alleles in *apoea*, and of vascular lipid deposition with the number of mutated alleles in *ldlra*, amongst others. Our results show that DIVaH reliably and efficiently calls CRISPR-Cas9-induced mutations and haplotypes in large numbers of BAM files in batch. This facilitates the translation of results from genetic association studies in humans using high-throughput zebrafish model systems.

3212F

Novel molecular signatures for heart failure: A comprehensive study of mRNA, lncRNA, and miRNA expression. Y. Wang¹, X. Hua², Q. Xiong¹, P. Jia¹, Z. Zhao¹, J. Song². 1) Center of Precision Health, School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX; 2) Department of Cardiac Surgery, State Key Laboratory of Cardiovascular Disease Fuwai Hospital, National Center for Cardiovascular Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, China.

Heart failure (HF), characterized by structural and functional impairment of the heart, has been a major cause of hospitalization, morbidity and mortality. Although numerous advances have been made in understanding its disease etiology and molecular mechanisms, a systematic investigation of the genetic changes at the transcriptional level has not been much done yet. Such a study may help us find molecular markers or features underlying the cellular condition during heart failure, and hopefully, discovering the biomarkers for the development of therapeutic strategy for HF. In addition to protein-coding genes, recent studies have revealed non-coding RNAs, such as microRNAs (miRNAs) and long noncoding RNAs (lncRNAs), may play important regulatory roles in the pathogenesis of HF. So far, the specific noncoding RNAs and their roles in HF still remain as preliminary, but promising. In this study, we generated whole transcriptome data from both mRNA and small RNA of 21 HF samples and 9 matched control samples using Illumina HiSeq 2500 platform. Using our in-house transcriptomic data analysis pipeline, we identified 126 protein-coding genes, 42 miRNAs, and 16 lncRNAs that are differentially expressed in the HF samples versus the controls. Our further analysis of these differentially expressed mRNAs, miRNAs and lncRNAs indicated that they were related to the HF-associated pathways and biological processes, such as regulation of Insulin-like Growth Factor (IGF) transport and regulation of blood pressure. These results supported these gene signatures to be involved in the HF pathogenesis. We further investigated their co-regulation by developing miRNA-mRNA and lncRNA-mRNA regulatory network approaches. Several differentially expressed miRNAs (e.g., *miR-190-5p*, *miR-129-5p* and *miR-1-3p*), lncRNAs (e.g., *LOC101926975*, *PDZRN3-AS1* and *BANCR*) and genes (e.g., *JAK2*, *PROS1* and *BDNF*) were found to be critical in the co-regulatory network, indicating their potential co-functional roles in HF condition. Finally, some of the results were validated by an independent transcriptional dataset (7 HF and 6 matched control samples) by using the same approach (RNA-seq and small RNA-seq) and a proteomic dataset (4 matched HF and one mixed controls of 5 matched samples) based on TMT quantitation. In summary, we identified a number of novel, differentially expressed mRNAs, miRNAs and lncRNAs, and their co-regulation results provided insights into the molecular mechanisms of HF.

3213W

Pre-annotation of genomic variants related to cardiovascular phenotypes for clinical reporting with the Neptune Pipeline. *E. Venner, Y. Jiang, D. Murdock, S.E. Kalla, M. Murugan, T.J. Wu, C. Kovar, K. Walker, J. Hu, T. Chiang, W. Salerno, G. Metcalf, D. Muzney, R. Gibbs.* Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX.

Pairing electronic medical records (EMR) with genomic data will drive precision medicine, but clinical reporting of genomic data is a costly and manual process. As a pilot project that enables precision medicine in an active clinical setting, we are validating a new clinical assay of 158 genes known to be associated with cardiovascular phenotypes, including coronary artery disease, arrhythmia, cardiomyopathy and aortic aneurysm. We will be launching this test with our clinical partners in the third quarter of 2018. To address this challenge, we have developed Neptune, an analytical platform to sign-out and deliver clinical reports. Following DNA sequencing and processing with the Human Genome Sequencing Center's Mercury Pipeline (PMID: 24475911), clinical DNA samples are analyzed via Neptune's custom annotation software that identifies a small subset of variants for manual review and addition to a "VIP" database of clinically-relevant variation. When all variants that require review in a sample have been curated, clinical reports are automatically generated, integrating SNVs, Indels, CNVs and CPIC level A pharmacogenomic variants. A structured version of the report is available in xml and json formats, with an HL7-FHIR standards compliant version in development. To facilitate clinical reporting for this new project, we have begun to augment our existing VIP database with annotations of variants and key publications specific to these cardiovascular genes. We have begun identifying publications with functional and case/control data that are relevant to this project. These publications are being pre-assessed in order to speed up reporting during the project. In addition, we are curating over 8,000 variants, identified in the ARIC cohort [PMID: 2646917] that lack strong ClinVar curations but which are likely to show up in our cohorts. By managing variant classifications and interpretations in a VIP database, we decouple variant curation from reporting, which in turn simplifies the logistics of the variant interpretation process and facilitates variant classification by external partners.

3214T

Validation of CNV detection in targeted panels and results from reanalysis of a historical cohort of cardiac patients. *S.P. Sadedin¹, C. Ieng², A. Oshtack¹.* 1) Murdoch Childrens Research Institute, Parkville, Victoria, Australia; 2) Victorian Clinical Genetics Services, Parkville, Victoria, Australia.

Copy number variants (CNVs) are a well known cause of rare genetic disease, and are routinely screened for in clinical laboratories. Historically, this has been performed either by microarray (to detect larger imbalances), or single-gene methods when a candidate gene can be identified. However, sensitive detection of intragenic and whole gene deletions and duplications across a wide set of genes, has remained challenging. Targeted sequencing gene panels used for detection of SNVs and indels can also detect CNVs. However the data generated is subject to distortions and biases that make this application significantly more challenging than detection of SNVs and indels. These challenges are compounded by a paucity of positive controls and reference samples for use in validation. Here we present our experience from validating CNV detection for clinical use in a targeted gene panel of 130 genes related to cardiac disease. Our method includes use of Ximmer, a freely available CNV simulation and analysis framework that supports rapid evaluation and tuning of CNV detection performance. Using Ximmer, we perform a comprehensive comparison of methods for CNV calling and identify factors affecting sensitivity and precision. By combining Ximmer results with sequencing of a set of positive controls, we are able to confidently predict performance over each exon. Using the accredited analysis pipeline, we re-analysed samples from a historical cohort of 545 patients, informing on the increase in diagnostic yield from application of CNV calling in cardiac testing.

3215F

A comprehensive and accurate knowledgebase system for variant clinical interpretation using ACMG-AMP standard guideline in inherited cardiac diseases and hereditary cancers. C. Chen, Q. Song, F. Xu, P. Yu, G. Hu. Admera Health LLC, South Plainfield, NJ.

With the rapid development and adoption of next-generation sequencing (NGS) technologies, the complexity of variant interpretation has increased significantly and new challenges have emerged in the clinical interpretation of Mendelian and complex diseases. While the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) recommended standards and guidelines for the interpretation of sequence variations in 2015, it did not offer specific approaches or algorithms for implementing these guidelines in detail, making discrepancies of variant interpretation widely exist between intra- and inter-laboratory settings. Here we present a crowdsourced knowledgebase system to provide accurate clinical interpretation of variants related to a number of human inherited diseases, including inherited cardiac diseases and hereditary cancers. Strictly following the 28 criteria provided by ACMG-AMP standards and guidelines, we have built a knowledgebase which curated comprehensive information from Human Gene Mutation Database (HGMD), ClinVar, 1000 Genome Project, dbSNP, and other population, disease-specific, and sequence related databases, combined with a number of *in silico* predictive algorithms (such as SIFT and PolyPhen-2). The comprehensive knowledgebase is able to provide reliable and consistent genetic variant interpretation with the five-tier classification system (Pathogenic/Likely pathogenic/Benign/Likely Benign/Uncertain significance) and help human interpreters understand the clinical significance of genetic variants intuitively. Our implementation allows any potential new variants to be included and provides the most updated information to the scientific community. The knowledgebase can be easily used by researchers and clinicians, and will significantly enhance our understanding of the functional consequences of genetic variants in human diseases.

3216W

How well can we classify coronary artery disease using all genetic data and choosing the best classification algorithm. D. Gola^{1,2}, T. Andlauer³, N. Mirza-Schreiber³, L. Zeng⁴, G. Delgado⁵, M. Kleber⁶, I. Gerge⁶, W. März⁵, S. Kanoni⁶, P. Deloukas^{6,7,8}, H. Schunkert^{4,12}, N. Samani⁹, J. Erdmann^{2,10,11}, B. Müller-Myhsok³, I.R. König^{1,2,13}. 1) Institut für Medizinische Biometrie und Statistik, Universität zu Lübeck, Lübeck, Schleswig-Holstein, Germany; 2) German Centre for Cardiovascular Research (DZHK), partner site Hamburg/Kiel/Lübeck, Lübeck, Germany; 3) Department of Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Munich, Germany; 4) Deutsches Herzzentrum München, Technische Universität München, München, Germany; 5) Medical Clinic V, Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany; 6) Department of Human Genetics, Wellcome Trust Sanger Institute, Hinxton, United Kingdom; 7) William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom; 8) Princess Al-Jawhara Al-Brahim Centre of Excellence in Research of Hereditary Disorders (PACER-HD), King Abdulaziz University, Jeddah, Saudi Arabia; 9) Department of Cardiovascular Sciences University of Leicester and NIHR Leicester Cardiovascular Biomedical Research Unit, Glenfield Hospital, Leicester, United Kingdom; 10) Institute for Cardiogenetics, Universität zu Lübeck, Lübeck, Germany; 11) University Heart Center Lübeck, Lübeck, Germany; 12) DZHK, partner site Munich Heart Alliance, Munich, Germany; 13) Airway Research Center North (ARCN), Member of the German Center for Lung Research (DZL).

Coronary artery disease (CAD) is the leading global cause of mortality and has substantial heritability with a polygenic architecture. Recent approaches of risk prediction were restricted in that they did not utilize the complete genetic information available, but rather focused on genetic loci found to be associated to CAD. Also, they were based on simple genomic risk scores (RS) not taking possible interdependencies between variants into account. We therefore benchmarked RS, logistic (penalized) regression, naïve Bayes (NB), random forests (RF), support vector machines (SVM) and gradient boosting (GB) on a dataset of 7736 CAD cases and 6774 controls from Germany to identify the algorithms for most accurate classification of CAD patients. After training, the final models were validated on an independent dataset of 527 CAD cases and 473 controls out of the same population and two further independent datasets from UK (1874 CAD cases, 1874 controls) and Germany, France and UK (365 CAD cases, 401 controls). We found RS using 50 633 genetic markers to be the most suitable algorithm for CAD classification, yielding an area under the receiver operating curve (AUC) of 0.92 (95% CI [0.90, 0.95]) in the validation data from the same population. NB and SVM performed better than RF and GB, with AUC of about 0.81 and 0.75, respectively. Applied to datasets from slightly different populations, however, all classification models had markedly reduced AUC of 0.5 – 0.6. We conclude that using all available genetic information can boost classification performance, although intensive validation of prediction models is crucial to assess their usability in populations different from those used to build the models.

3217T

Mosaic mutations contribute to congenital heart disease. A. Hsieh¹, S. Morton², H. Qi³, K.B. Manheimer⁴, D. Bernstein⁵, R.W. Kim⁶, J.W. Newburger⁷, G.A. Porter Jr.⁷, D. Srivastava⁸, M. Tristani-Firouzi⁹, M. Brueckner¹⁰, R.P. Lifton¹⁰, C.E. Seidman², E. Goldmuntz¹², B.D. Gelb⁴, J.G. Seidman^{2,11}, W.K. Chung¹, Y. Shen^{1,13}. 1) Columbia University Medical Center, New York, NY; 2) Boston Children's Hospital, Boston, MA; 3) Columbia University, New York, NY; 4) Icahn School of Medicine at Mount Sinai, New York, NY; 5) Stanford University School of Medicine, Palo Alto, CA; 6) Children's Hospital Los Angeles, Los Angeles, CA; 7) University of Rochester Medical Center, Rochester, NY; 8) University of California San Francisco, San Francisco, CA; 9) University of Utah School of Medicine, Salt Lake City, UT; 10) Yale University School of Medicine, New Haven, CT; 11) Harvard Medical School, Boston, MA; 12) Children's Hospital of Philadelphia, Philadelphia, PA; 13) JP Sulzberger Genome Center, New York, NY.

Mosaicism is due to post-zygotic *de novo* mutation, resulting in two or more cell populations with distinct genotypes in a single individual. These post-zygotic events have been implicated in developmental disorders such as structural brain malformations, epilepsy, and autism, but their role in congenital heart disease (CHD) and other congenital anomalies is still unclear. Additionally, current estimates of the fraction of *de novo* events occurring post-zygotically and the mosaic mutation rate per individual are unstable due to differences in sequencing depth or variant calling methods. To systematically estimate the contribution of mosaicism to CHD, we developed a new computational method to characterize mosaic single nucleotide variants from exome sequencing data. We use an expectation-maximization algorithm to jointly estimate the fraction of mosaics among apparent *de novo* mutations and the false discovery rate of candidate mosaics. We applied our method to data from 2645 CHD proband-parents trios from the Pediatric Cardiac Genomics Consortium and Pediatric Heart Network programs. We detected a total of 288 high-confidence protein-coding mosaic mutations from these blood samples, an average of 0.11 per individual. Independent sequencing confirmed 90% of selected candidates. Among the mosaics in genes highly expressed in developing heart, the variants predicted to be damaging tended to have a higher variant allele fraction than likely-benign variants ($P=0.016$), suggesting that damaging mosaics contribute to CHD and that their contribution is positively correlated with the percentage of cells carrying the mutations. In particular, we observed deleterious mosaics in genes involved in known biological processes relevant to CHD or developmental disorders, such as *ISL1*, *SETD2*, *NOVA2*, *SMAD9*, *LZTR1*, and *FZD5*. However, comparing the proportions of damaging and non-damaging mosaic and germline *de novo* variants in high heart expression genes, we find that mosaic variants detected in blood as a class are less likely to be pathogenic than germline variants ($P=0.005$), probably due to having a smaller chance of being present in the heart during development. Finally, we calibrated detection power given sequencing depth and estimated that the true frequency of coding mosaics with level of mosaicism greater than 20% in the blood is 0.17 per patient, representing an understudied but substantial fraction of the genetic risk to congenital anomalies and other diseases.

3218F

An artificial intelligence-based next-generation integrated genetic-epigenetic prediction of 5-year risk for symptomatic coronary heart disease. M.V. Dogan^{1,2}, R.L. Simons³, S.R.H. Beach⁴, A. Lendasse⁵, R.A. Philibert^{1,2,6,7}. 1) Biomedical Engineering, University of Iowa, Iowa City, IA; 2) Cardio Diagnostics LLC, Coralville, IA; 3) Sociology, University of Georgia, Athens, GA; 4) Psychology, University of Georgia, Athens, GA; 5) Mechanical and Industrial Engineering, University of Iowa, Iowa City, IA; 6) Psychiatry, University of Iowa, Iowa City, IA; 7) Behavioral Diagnostics, Coralville, IA.

Coronary heart disease (CHD) is responsible for >300,000 deaths annually. CHD associated morbidity and mortality can be reduced if those at risk can be identified well in advance, to put in place prevention interventions prior to the occurrence of an acute coronary event. Currently, there are several approaches used to infer risk for CHD. However, their limitations call for additional efforts to more sensitively predict risk for developing symptomatic CHD. To that end, we recently developed a Precision Medicine tool that in addition to predicting 5-year risk, can also be used to guide personalized prevention interventions and monitor changes in risk over time. It was developed using an application of artificial intelligence, machine learning, in the Framingham Heart Study to mine integrated genetic-epigenetic biosignatures that capture the complex interplay between our genome and the environment in conferring risk for CHD. The training and test sets consisted of 1180 (19/695 females and 23/485 males developed symptomatic CHD within 5-years) and 524 (8/293 females and 12/231 males developed symptomatic CHD within 5-years) individuals, respectively, of Northern European ancestry. Non-linear data mining and modeling techniques were employed on the training set to identify prediction signatures from genome-wide SNP and DNA methylation (DNAm) data. As a result, 14 DNAm and 10 SNPs were identified and trained on the training set (training bootstrap average sensitivity and specificity of 0.61 and 0.71, respectively), which then, predicted the 5-year risk in the test set with a sensitivity and specificity of 0.60 and 0.69, respectively. Performance was compared to that of the Framingham Risk Score (FRS) and the ASCVD Risk Estimator. The sensitivity and specificity in the test set were 0.05 and 0.99, respectively, for FRS, and 0.38 and 0.85, respectively, for ASCVD. Even with a small number of incident cases, our tool performed with superior sensitivity, which is vital because, while a false positive would result in additional testing, a false negative can be detrimental. Unlike genetics, because DNAm is dynamic, it can be mapped to actionable risk factors to inform personalized interventions. Post-interventions, changes in DNAm can be used to re-assess changes in risk and hence, continuously monitor heart health. Additional efforts are ongoing to optimize, validate and extend our tool to further improve performance in larger, more ethnically diverse cohorts.

3219W

PseudoNet: reconstructing pseudo-time in single-cell RNA-seq data using neural networks. J. Lakkis¹, G. Hu^{1,4}, C. Xue², H. Pan², S.B. Trignano², H. Zhang², M.P. Reilly², N.R. Zhang³, M. Li¹. 1) Department of Biostatistics Epidemiology, and Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; 2) Division of Cardiology, Department of Medicine, Columbia University Medical Center, New York, NY; 3) Department of Statistics, University of Pennsylvania Wharton School of Business, Philadelphia, PA; 4) School of Mathematical Sciences and LPMC, Nankai University, Tianjin, PR China.

Single-cell gene expression profiling can be used to quantify transcriptional dynamics in temporal processes, such as cell differentiation, using computational methods to label each cell with a "pseudo-time." Pseudo-time analysis models the gradual transition of a cell rather than assigning it to a discrete cluster. Popular pseudotime reconstruction methods such as Monocle2 do not scale well as the number of cells increases. As a result, these methods cannot jointly analyze data across multiple subjects when each subject has a few thousand single cells. As single-cell RNA-seq becomes widely used in biomedical research, many research groups are now sequencing a large number of cells across many subjects. It is important to identify genes whose expression changes over pseudo-time across subjects in different disease conditions. However, when analyzing data individually, it becomes challenging to compare gene expression across subjects as the estimated pseudotimes are not on the same scale. To overcome these limitations, we present PseudoNet, a new, supervised machine learning approach for pseudotime reconstruction to model a non-branching temporal cell process. A feedforward neural network is trained to predict the probability that a cell is near a pre-specified end of the temporal process based on its gene expression, and the predicted probability is used as a proxy for pseudo-time. Training is performed on cells that are known or are certain to be near either end of the temporal process. PseudoNet efficiently borrows information across subjects and returns a pseudotime standardized to the [0, 1] scale, which allows the identification of genes that are differentially expressed over pseudo-time across subjects in different conditions. To evaluate the performance of PseudoNet, we analyze benchmark datasets for cell cycle and cell differentiation, where true cell states are known. We show that our method requires less computation time and classifies cells with greater accuracy and higher resolution than other popular pseudo-time reconstruction methods. We also apply this method to a single-cell RNA-seq dataset consisting of monocyte gene expression data for 15 patients, totaling ~60,000 single cells. We show that PseudoNet identifies genes differentially expressed between healthy people and those with cardiometabolic disease risk factors (e.g., smoking, diabetes, high blood lipids). An R package of PseudoNet is made available.

3220T

Biomarker predictors of acute ischemic stroke: Results of metabolome in ischemic stroke study (MISS). C. Bejar¹, B. Apple¹, L. Reddivari², E. Sidorov¹, J.K.P. Vanamala², D.K. Sanghera¹. 1) University of Oklahoma Health Science Center, Oklahoma City, OK; 2) Penn State University, University Park, PA, USA.

Acute ischemic stroke (AIS) is the fourth leading cause of death in the United States. Despite advances in research, early detection and treatment strategies still suffer from significant limitations. There has been an increased interest to identify novel biomarkers, which may correlate with disease etiologies using advanced metabolite profiling. However, most such studies have been performed in animals; very few studies have examined AIS in humans. This investigation is part of our ongoing Metabolome in Ischemic Stroke Study (MISS) in which we have performed a global metabolome profiling using serum samples of 119 subjects. These include 44 MISS patients admitted at the OU Medical Center's emergency and were diagnosed with the AIS; on these blood and urine samples available at both acute- (within 72 hours of AIS) and chronic phase (2-3 month follow-up); 31 chronic patients who had stroke in the past, and their specimens were collected only at the chronic phase. We also investigated 44 age- and gender- matched healthy controls for comparison. The global metabolomic profiles were generated using AB Sciex TripleTOF 5600 for untargeted analysis and Thermo Scientific Q Exactive Plus for targeted analysis. We identified and prioritize top 1,919 metabolites that correlated with AIS in serum. Putative identities were assigned to ~70 metabolites, and both known and novel metabolites were confirmed using appropriate standards. Several metabolites exhibited significantly altered patterns between patients and controls in serum and urine. Myo-inositol, 2,3 Hydroxybenzoic Acid, Pyridoxic Acid, Allantoin, Hydroxyphenyllactate, Ketoisovalerate, and few essential amino acids were significantly perturbed ($p < 0.0001$) among the AIS patients at both acute and chronic phase, while the magnitude of variation was significantly pronounced at the acute phase (which ranged from 20% to 50%) compared with the chronic phase. However, the other essential amino acids including BCAAs, Histidine, Methionine, Proline, and Serine were only reduced during the acute phase. Our results suggest that the dysregulation of specific metabolic pathways may have long-term or irreversible effects after the cerebral ischemic episode. These findings once replicated in the expanded MISS cohort would provide novel insights on underlying molecular pathways altered in AIS and may facilitate the development of diagnostic and therapeutic approaches for early detection and prevention. .

3221F

Metabolic profiling and sequencing at biobank scale for prioritizing drug targets. S. Ruosaari. Nightingale Health, Helsinki, Finland.

Advances in metabolomics now allow comprehensive biomarker profiling of entire biobanks and clinical trials. This provides a plethora of scientific opportunities such as discovery of novel biomarkers for onset of cardiovascular diseases and tracking their progression, as well as etiological insights into established cardiometabolic risk factors. Here we showcase scientific applications in combining metabolomic data with large-scale genomics and electronic health care records in entire biobank cohorts. Nightingale Health Ltd has developed a metabolomics platform for large-scale 'omics and screening projects. Previous studies show that the platform significantly improves risk stratification in primary and secondary CVD prevention (Circulation 2015;131:774). Also, it has been used to examine metabolomic consequences of PCSK9 inhibition in ~72,000 individuals from European cohort studies using a loss-of-function variant in PCSK9 R46L to mimic the therapeutic effects of PCSK9 inhibitors (unpublished data). To clarify the clinical utility of the findings, the metabolomics platform is currently being applied to profile close to 1,000,000 biobanked blood samples from diverse ethnicities with extensive electronic health care records, and the resulting biomarker data will be made available for the medical research community. Simultaneously, the metabolomic profiling is for the first time being piloted in healthcare settings for guiding clinical decision making for individual patient, exemplifying how metabolomic profiling of large biobank studies pave way for precision medicine approaches with strong evidence base. The benefits of metabolomic profiling in a precision medicine setting are clear. They help improve risk stratification and provide more clear clinical interpretation of the biomarker results. We also illustrate the benefits of tracking the metabolomic effects in response to various lifestyle and pharmacological interventions. These results demonstrate the translational utility of NMR metabolomics studies in large biobanks providing enhanced cardiovascular risk prediction and novel means to track effectiveness of interventions.

3222W

Genetic background drives molecular variability in a large-scale collection of iPSC-derived cardiomyocytes. M.K.R. Donovan^{1,2}, A. D'Antonio-Chronowska³, P. Benaglio³, E.N. Smith³, M. D'Antonio⁴, K.A. Frazer^{3,4}.

1) Bioinformatics and Systems Biology Graduate Program, University of California, San Diego, La Jolla, CA 92093, USA; 2) Department of Biomedical Informatics, University of California, San Diego, La Jolla, CA 92093, USA; 3) Department of Pediatrics, University of California, San Diego, La Jolla, CA 92093, USA; 4) Institute for Genomic Medicine, University of California, San Diego, La Jolla, CA 92093, USA.

Induced pluripotent stem cell (iPSC)-derived cardiomyocytes (iPSC-CMs) have emerged as a promising model system to identify and characterize regulatory variation underlying cardiac traits and disease. Using 144 iPSCs from whole-genome sequenced individuals, we have derived the largest collection of iPSC-CMs to date. We characterized heterogeneity, including: 1) varying differentiation efficiencies of iPSCs, and 2) iPSC-CM cellular compositions, to examine its impact on using large collections of iPSC-CMs for studying regulatory variation. We investigated if iPSC molecular and regulatory characteristics were predictive of differentiation outcome. Sex was significantly associated ($p = 0.0025$ Fisher's exact test), with female iPSCs differentiating successfully more often than male iPSCs. Analyzing RNA-seq for all 144 iPSC lines, we also observed increased erosion of the inactive X chromosome in successfully differentiated female iPSCs. Gene expression (bulk RNA-seq) analysis in iPSCs revealed that *TLE3*, a master regulator of the Wnt-signaling pathway, was under-expressed in successfully differentiated iPSCs, consistent with proper Wnt signaling being critical to the success of iPSC-CM differentiation. We performed single cell RNA-seq on 8 iPSC-CMs (33,914 cells) with a range of differentiation efficiencies (%cTnT), and observed that two distinct cell populations were present. The majority of cells belonged to population 1, which was strongly enriched for cardiac development-associated genes including *TNNI2* (cTnT), *NKX2-5*, and *MYL7* (MLC2a), while population 2 was enriched for functions associated with the extracellular matrix. The percentage of population 1 cells in each iPSC-CM sample was directly reflective of its differentiation efficiency. To investigate the extent to which cellular heterogeneity impacts the association between genetic variation and molecular phenotypes in iPSC-CMs, we analyzed RNA-seq, ATAC-seq, and ChIP-seq for H3K27ac from 24 iPSC-CMs derived from 5 MZ twin pairs. We estimated the relative distributions of the two cellular populations in all 24 iPSC-CM lines with Ciber-sort, and showed that while in the unadjusted samples genetic background accounted for most molecular variability, using cell population as a covariate resulted in an even higher association. Overall, our study shows that molecular phenotypes in iPSC-CMs are largely associated with genetic background, which supports the use of large collections of iPSC-CMs for genetic studies.

3223T

Disease Navigator: Utilizing model organisms for functionalizing human genome variation. *M. Tomczuk¹, J. De Pons², C. Smith¹, J. Richardson¹, J. Blake¹, C. Bult¹, M. Shimoyama².* 1) Mouse Genome Informatics, The Jackson Laboratory, Bar Harbor, ME; 2) Rat Genome Database, Medical College of Wisconsin, Milwaukee, WI.

Integration of and accessibility to multiple datasets can accelerate new discoveries and can aid in understanding human health, and in disease prevention, diagnosis and generation of therapies. The Disease Navigator, a project that is part of the NIH Data Commons pilot initiative, is a new portal that links human genomic, expression and phenotypic disease data with relevant model organism information from the Alliance of Genome Resources to help elucidate the mechanisms of human disease. The Alliance of Genome Resources (www.alliance.org) is a consortium of six model organism databases (Mouse Genome Database—MGD, Rat Genome Database—RGD, Saccharomyces Genome Database, WormBase, FlyBase, and Zebrafish Information Network), and the Gene Ontology Consortium – all working together to standardize and incorporate data from human and these organisms into a single integrated resource. Disease Navigator currently focuses on cardiovascular disease, particularly cardiomyopathy, and incorporates human data from the Genotype-Tissue Expression (GTEx) project and Trans-Omics for Precision Medicine (TOPMed), and animal model data from MGD and RGD. Disease terms are sourced from ontologies/vocabularies (Disease Ontology, Mammalian Phenotype Ontology, Human Phenotype Ontology, Vertebrate Trait Ontology, CheBI and Uberon). The portal is gene-centric and contains disease-associated genes from human, mouse, and rat. Additional material includes quantitative phenotype data, rat and mouse strains and models, genomic data and eQTL from TopMed, human gene expression data from GTEx, and human gene variation data from ClinVAR. We will show how the Disease Navigator integrates datasets and can be used to take advantage of model organism data to understand the functional significance of human genome variation. A search by a human gene returns links to gene reports for the human and mouse and rat orthologs, eQTL data, expression data, and associated disease models. A search by disease term such as cardiomyopathy, returns genes, variants, genotypes, QTLs, heritable phenotypes, strains, disease models and expression data. Lists of genes associated to a disease term can be further analyzed for functional commonalities or identify overlap with other genes. Future efforts will add more diseases, genomic and phenotypic data from other model organisms and multiple human projects, as well as tools to aid in analysis of the myriad of available data. Supported by NIH grant U41 HG001315-24S1.

3224F

Comprehensive analysis of chromatin organization delineates regulatory programs of human cardiomyocyte differentiation. *Y. Zhang¹, S. Preissl^{1,2}, T. Li³, H. Yang³, J. Grinstein³, E. Destici³, A. Lee¹, S. Chee¹, Z. Ye¹, K. Ma¹, N. Tedeschi³, S. Kuan¹, J. Vevers⁴, S. Evans^{4,5}, N. Chi³, B. Ren^{1,2,6}.* 1) Ludwig Institute for Cancer Research, San Diego, CA; 2) Department of Cellular and Molecular Medicine, UCSD; 3) Department of Medicine, Division of Cardiology, UCSD; 4) Department of Medicine, UCSD; 5) Department of Pharmacology and Skaggs School of Pharmacy and Pharmaceutical Sciences, UCSD; 6) Institute of Genomic Medicine, and Moores Cancer Center, UCSD.

Chromosome conformation capture technologies have revealed important impacts of chromatin folding on gene regulation. However, little is known about how the 3D chromatin structure is established and reorganized during human cardiomyocyte development and dysregulated in heart diseases. Here, we performed high-resolution in-situ Hi-C, histone ChIP-seq and ATAC-seq to interrogate the chromatin architecture during six key developmental stages of human cardiomyocyte differentiation. We find substantial dynamic organization of the chromatin structure at all scales including compartments, topologically associated domains (TADs) and chromatin loops. Interestingly, chromatin loop interactions correlate with stage-specific gene expression program and form interaction hubs on the promoters of cardiac transcription factors. Integration of chromatin loops with noncoding variants identified in genome-wide association studies (GWAS) predicts multiple candidate genes for heart diseases, including several ion channel genes involved in cardiac conduction and genes involved in early cardiac differentiation. Our work provides a rich resource not only for understanding the chromatin structure changes during cardiac development, but also resolving the impact of non-coding variants in heart diseases.

3225W

Mapping of HiChIP chromatin looping QTLs in human coronary artery smooth muscle cells and their relationship to other regulatory and disease variation. *M. Pjanic*^{1,2}, *M. Dacre*³, *Q. Zhao*^{1,2}, *T. Nguyen*^{1,2}, *B. Liu*⁴, *R. Wirka*^{1,2}, *T. Wang*⁴, *C.L. Miller*⁵, *H. Fraser*⁵, *T. Quertermous*^{1,2}. 1) Department of Medicine, Division of Cardiovascular Medicine, School of Medicine, Stanford, CA; 2) Cardiovascular Institute, Stanford University, Stanford, CA; 3) Department of Biology, Stanford, CA; 4) Department of Genetics, Stanford, CA; 5) Department of Public Health Sciences, School of Medicine, University of Virginia, Charlottesville, VA.

Coronary artery disease (CAD) has become a major chronic condition in modern human societies and is the global leading cause of death in both men and women. Given that approximately half of the risk for CAD is genetic in nature, in this work we aimed to annotate the human genome by mapping regulatory variation to identify disease mechanisms. Toward that end, we have conducted the first large-scale pooling approach for HiChIP to map chromatin looping QTLs (cQTLs) in disease-related human coronary artery smooth muscle cells (HCASMC). We pooled chromatin extracts from 71 individuals and performed the HiChIP protocol with H3K27ac immunoprecipitation. We processed data as described by the Fraser Lab to identify statistically significant SNPs with different pre- vs. post-frequency as related to chromosomal looping. From a total number of 3,636,814 analyzed SNPs, we obtained 24,260 high confidence cQTLs with $p < 0.001$. By applying binomial statistics to test the significance of overlaps between GWAS Catalog variants and cQTL variants in a window of 100 bp, we report that HCASMC cQTLs are enriched in disease-associated SNPs from CARDIoGRAMplusC4D Consortium GWAS studies that comprise a collection of risk SNPs for CAD and myocardial infarction ($p < 1e-4$). In addition, when broader regions of 1 kb surrounding cQTLs were tested, a significant GWAS term was "coronary artery calcification", a risk factor for adverse outcomes in patients with CAD. When transcription factor (TF) position weight matrix (PWM) enrichment was examined in the narrow region of ± 25 bp surrounding the cQTLs, we report enrichment for PWMs for smooth muscle related TFs SRF and TEAD1, as well as CAD associated TFs TCF21 and SMAD3. Our results indicate that by applying the pooled approach and characterizing cQTL variants, we can obtain novel information regarding the regulation of gene transcription, and identify chromatin looping as a genetic mechanism that may be regulated by allelic variation to modulate the risk of CAD.

3226T

A promoter interaction map for cardiovascular disease genetics. *L.E.*

*Montefiori*¹, *D.R. Sobreira*¹, *N.J. Sakabe*¹, *I. Aneas*¹, *A.C. Joslin*¹, *G.T. Hansen*¹, *G. Bozek*¹, *I.P. Moskowitz*², *E.M. McNally*³, *M.A. Nobrega*¹. 1) Human Genetics, University of Chicago, Chicago, IL; 2) Department of Pediatrics and Pathology, The University of Chicago, Chicago IL; 3) Center for Genetic Medicine, Northwestern University Feinberg School of Medicine, Chicago IL.

Over 500 genetic loci have been associated with risk of cardiovascular diseases (CVDs), however most loci are located in gene-distal non-coding regions and their target genes are not known. We generated high-resolution promoter capture Hi-C (PCHI-C) maps in human induced pluripotent stem cells (iPSCs) and iPSC-derived cardiomyocytes (CMs) to provide a resource for identifying and prioritizing the functional targets of CVD associations. We validated these maps by demonstrating that promoters preferentially contact distal sequences enriched for tissue-specific transcription factor motifs and are enriched for chromatin marks that correlate with dynamic changes in gene expression. Furthermore, the majority (80%) of PCHI-C interactions occurred within topologically associating domains (TADs), indicating that this approach recapitulates known features of genome organization. Using the CM PCHI-C map, we linked 1,999 CVD-associated SNPs to 347 target genes; importantly, these genes were enriched for causing cardiovascular phenotypes when deleted in the mouse. Remarkably, more than 90% of SNP-target gene interactions did not involve the nearest gene, while 40% of SNPs interacted with at least two genes. Furthermore, CVD-associated SNPs that corresponded to expression quantitative trait loci (eQTLs) identified in left ventricle tissue were significantly more likely to loop to their associated gene than expected by chance. Our findings suggest that PCHI-C provides an important layer of information to help resolve the functional targets of cardiovascular disease risk loci.

3227F

Methylation at nucleotide C62 in spliceosomal RNA U6 alters mRNA splicing which is important for embryonic development. A. Ogren¹, N. Kibiryeve², J. Marshall², J. O'Brien Jr.², D.C. Bittel^{1,2}. 1) Kansas City University of Medicine and Bioscience, Kansas City, MO 64106; 2) Children's Mercy Hospital, Kansas City, MO 64108.

Understanding the regulation of development can help elucidate the pathogenesis behind many developmental defects found in humans and other vertebrates. Evidence has shown that alternative splicing of mRNA plays a role in developmental regulation, but we know little about the underlying mechanisms. Notably, a subset of small noncoding RNAs known as scaRNAs (small cajal body associated RNAs) contribute to spliceosome maturation and function through covalently modifying spliceosomal RNAs by either methylating or pseudouridylating specific nucleotides, but the developmental significance of these modifications are not well understood. Our focus is on one such scaRNA, known as snord94 or U94, that methylates one specific cytosine (C62) on spliceosomal RNA U6, thus potentially altering spliceosome function during embryogenesis. We previously showed that mRNA splicing is significantly different in myocardium from infants with congenital heart defects (CHD) compared to controls. Furthermore, we showed that modifying expression of scaRNAs changes splicing in human cells, and zebrafish and quail cells and embryos. Here we present evidence that methylation at C62 in U6 is associated with altered splicing and CHD. The potential importance of scaRNAs as a developmentally important regulatory mechanism controlling alternative splicing of mRNA is unappreciated and needs more research.

3228W

Enhancer-associated long non-coding RNA regulates CTGF locus in calcific aortic valve stenosis. A. Chignon, M. Rosa, MC. Boulanger, G. Mkanneze, D. Argaud, R. Devillers, S. Thériault, Y. Bossé, P. Mathieu. Institut universitaire de cardiologie et de pneumologie de Québec, Université Laval, Quebec City, QC, G1V 4G5, Canada.

Introduction: Calcific aortic valve stenosis (CAVS) is a chronic and prevalent heart valve disorder characterized by ectopic mineralization and a fibrogenic response. Growing evidence suggests that long non-coding RNAs (lncRNAs) are transcribed at regulatory loci such as enhancers. In topologically associated domains (TAD), lncRNA may be located in hub enhancers that orchestrate the transcription process. Whether an enhancer-associated lncRNA participates to CAVS and controls gene expression of dysregulated loci is presently unknown. **Methods:** The expression of lncRNAs was determined by RNA sequencing in 19 calcific aortic valves vs 8 normal valves, and dysregulated noncoding transcripts were prioritized by functional annotation of epigenetic marks. Functional characterization was carried out by chromatin immunoprecipitation (ChIP) and gene expression analyses. **Results:** Transcriptomic analysis by RNA sequencing combined with the analysis of H3K27ac epigenetic mark identified that one lncRNA located in an enhancer is dysregulated in CAVS. The overexpressed lncRNA is localized in a TAD comprising the *CTGF* locus, a gene involved in the control of fibrosis. Annotation of the locus in aorta indicates the presence of a cluster of enhancer with CTCF binding sites suggesting the presence of a hub enhancer. ChIP data indicate that H3K27ac at the lncRNA locus is increased by several-fold in the presence of CAVS. The lncRNA is localized in both the nucleus and the cytosol. Whole-genome chromatin conformation capture (Hi-C) data analysis indicates that the lncRNA locus interacts with *CTGF*, a pro-fibrotic gene that shares the same TAD. Treatment of human valve interstitial cells (VICs) with TGF- β , a factor known to up-regulate CTGF, increased the expression of the lncRNA by 2.95 fold. An oligonucleotide antisense-mediated knockdown of the lncRNA lowered its expression and largely abrogated TGF- β -mediated expression of CTGF. In addition, antisense-mediated knockdown of the lncRNA reduced the expression of COL1A1 and the secretion of pro-collagen-1 α . **Conclusion:** An epigenetically activated enhancer-associated lncRNA in CAVS controls the *CTGF* locus and promotes fibrogenesis.

3229T

Gene expression profiling of coronary vasculopathy and rejection episodes after cardiac transplantation. O. McDaniel, M. Garrett, L. McDaniel, C. Moore. Univ Mississippi Med Ctr, Jackson, MS.

Identification of biological markers in the context of coronary vasculopathy (CV) and or allograft rejection episode (RE) after cardiac transplantation is useful for early detection of the development of CV and/or RE. Rejection after cardiac transplantation is a complex immunologic response involving multiple signaling processes, associated with inflammatory pathways. The expression of the innate and adaptive immunity signaling proteins such as cytokines and Toll-like receptors are increased in peripheral blood mononuclear cells (PBMCs) and cardiac tissues during cardiac dysfunction. The goal was to screen by microarray analysis cluster of genes associated with these events and further verify association levels with CV and RE. CDNA microarray was developed from PBMCs of patients with RE and or CV using a combined Human 19K and 8K gene probes. The resulting database was normalized and underwent an unsupervised cluster analysis comparing samples associated with RE/CV vs. no rejection, no CV. Gene ontology-response to stimuli demonstrated in the following distribution: H19K- Grade 0 vs. grade 3 Rejection Gene Ontology Regulation of immune response (IR) 26.58% Innate immune response 26.58% Positive regulation of IR 21.5% Adaptive IR 10.36% Humoral IR 7.59% Therefore, extracting only the gene subset of interest we studied expression levels of innate immune response and cytokine genes in association with RE and CV in a new group of patients who undergone cardiac transplantation. The TLR2, and TLR4 which are expressed in cardiac myocytes were elevated in PBMCs, plasma and biopsy specimen in response to rejection and CV. The mean $\pm$ SEM level of TLR2 mRNA was increased only 1.8-fold in PBMCs ($p < 0.05$) but 4.5-fold in biopsy samples from patients with grade 3A rejection or grade III CV vs. grade 1A/null. Protein concentration of IL-18 and Interferon-gamma (IFN- γ) were in direct correlation with RE and CV. Nonetheless, in the presence of increased concentration of IL-10 and decreased IFN- γ in spite of increased IL-18, the grafts were rejection free. In conclusion, peripheral concentration of cytokines correlate with the local production seen in biopsy specimens. Thus, expression profiles of the innate immune signaling pathway may detect signals to predict early rejection episodes.

3230F

Using endothelial cell molecular QTLs to dissect coronary artery disease risk. L.K. Stolze¹, M.B. Whalen¹, N.K. Hadeli¹, A. Eshghi¹, A. Conklin¹, R. Victor², C.E. Romanoski¹. 1) Department of Cellular and Molecular Medicine, University of Arizona, Tucson, AZ; 2) Department of Molecular and Cellular Biology, University of Arizona, Tucson, AZ.

The leading cause of mortality worldwide is Coronary Artery Disease (CAD), which is caused by genetic and environmental risk factors. Estimates show CAD to be approximately 50% heritable, and accordingly, Genome-Wide Association Studies in over hundreds of thousands of individuals have revealed more than 160 risk haplotypes enriched in CAD patients versus controls. Still, the majority of associated loci reside outside the coding genome, suggesting regulatory function by way of unidentified mechanisms. Further, given the specificity of gene regulation, non-coding variants that perturb target gene expression likely function in specific cell types and disease-promoting activation states. The overarching goal of this project is to overcome challenges in identifying functional non-coding genetic variants linked to CAD risk by exploiting knowledge of regulatory networks in vascular endothelial cells that play a critical role in the onset of CAD pathogenesis. Our strategy is to build and analyze a comprehensive dataset detailing epigenetic molecular traits genome-wide in a population of over 50 primary arterial endothelial cell cultures isolated from aortic explants of individuals from both sexes and three major ancestral populations. This unique dataset includes genotypes, chromatin accessibility, histone modification, transcription factor binding, and transcriptomics genome-wide to identify regulatory elements whose functions are impaired by non-coding genetic variation. We collected molecular phenotypes from endothelial cells with and without exposure to the pro-inflammatory cytokine interleukin 1 beta. This design enables our discovery of genetically-driven individual responsiveness to heightened systemic inflammation that is a risk factor for CAD. Associations detected in these studies will provide important insights into common genetic variants that perturb endothelial-specific gene expression profiles in quiescent and pro-inflammatory states. Non-coding genetic variants that associate with epigenetic phenotypes provide evidence of the mechanisms by which CAD associated loci effect disease risk. Further, associations specific to the inflammatory setting demonstrate gene-by-environment interactions whereby risk alleles only drive risk in this disease-promoting environment. Taken together, our findings have implications not only for the functional genetics of CAD, but also inform studies of cell type-specific gene expression in endothelial cells.

3231W

Rejuvenation of senescent endothelial progenitor cells by extracellular vesicles derived from mesenchymal stromal cells. L. Wang^{1,2}, J.Q. Wei^{3,4}, J.M. Hare^{3,4}, C. Dong^{3,4}. 1) John P. Hussman Institute for Human Genomics; 2) Dr. John T. Macdonald Foundation Department of Human Genetics; 3) Interdisciplinary Stem Cell Institute; 4) Department of Medicine, University of Miami Miller School of Medicine, Miami, FL 33136.

Endothelial progenitor cells (EPCs) play a crucial role in vascular repair and angiogenesis. EPC senescence resulting from aging or exposure to other cardiovascular (CV) risk factors not only augments atherosclerosis, the underlying cause of ischemic CV diseases, but also diminishes angiogenesis following ischemic events. Thus, rejuvenating endogenous senescent EPCs in individuals at risk for CV diseases is of substantial clinical importance. Clinical trials on mesenchymal stromal cell (MSC) transplantation showed that young allogeneic MSCs but not aged autologous MSCs improved circulating EPC functionality in patients with dilated cardiomyopathy (DCM)—a disease characterized with microvascular lesions. We hypothesize that MSCs secrete bioactive factors to regulate EPC functions. To test this hypothesis, we used a transwell system to co-culture MSCs and EPCs without direct cell-cell contact. We found that young but not aged MSCs rejuvenated aged EPCs, as evidenced by decreased expression of cellular senescence markers (senescence-associated β -galactosidase, p16^{INK4A} and p19^{ARF}), and increased proliferation and differentiation capacity. Extracellular vesicles (EVs) are secreted by most types of cells and contain bioactive molecules to mediate intercellular communication. To further test if rejuvenation of EPCs by MSCs is mediated by EVs, we treated senescent EPCs with EVs purified from culture media of MSCs and demonstrated that EVs derived from young but not aged MSCs were capable of rejuvenating aged EPCs. Once EVs are taken up by the recipient cells either by receptor-mediated endocytosis or fusion of membranes, cargos packaged inside EVs, such as miRNA, are released to regulate gene expression in the recipient cells. We and others have shown that miR-126 is essential to maintain endothelial cell functionality and angiogenesis. The expression of miR-126 decreases during aging in both EPCs and MSCs. In an attempt to generate competent EVs for EPC rejuvenation, we overexpressed miR-126 in old MSCs to make them secrete "tailored" EVs (TEVs) loaded with miR-126. These TEVs from aged MSCs effectively rejuvenated aged EPCs *in vitro* and promoted angiogenesis *in vivo* in a mouse hind limb ischemia model (n=3, P<0.05). Our data support that EVs mediate the effects of MSCs on EPCs. In addition, our study suggests that EVs is a promising platform for developing an acellular therapeutic approach to rejuvenate senescent EPCs to treat ischemic CV disease.

3232T

Coronary artery disease genes SMAD3 and TCF21 promote opposing interactive genetic programs that regulate smooth muscle cell differentiation and disease risk. Q. Zhao¹, D. Iyer¹, R. Wirka¹, A. Naravane¹, T. Nguyen¹, B. Liu², M. Nagao¹, P. Cheng¹, C.L. Miller³, J.B. Kim¹, M. Pjanic¹, T. Quertermous¹. 1) Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA; 2) Department of Genetics, Stanford University School of Medicine, Stanford, CA; 3) Center for Public Health Genomics, Department of Public Health Sciences, Biochemistry and Genetics, and Biomedical Engineering, University of Virginia, Charlottesville, VA.

Although numerous genetic loci have been associated with coronary artery disease (CAD) with genome wide association studies, efforts are needed to identify the causal genes in these loci and link them into fundamental signaling pathways. Recent studies have investigated the disease mechanism of CAD associated gene SMAD3, a central transcription factor (TF) in the TGF β pathway, investigating its role in smooth muscle biology. In vitro studies in human coronary artery smooth muscle cells (HCASMC) revealed that SMAD3 modulates cellular phenotype, promoting expression of differentiation marker genes while inhibiting proliferation. RNA sequencing and chromatin immunoprecipitation sequencing studies in HCASMC identified downstream genes that reside in pathways which mediate vascular development and atherosclerosis processes in this cell type. HCASMC phenotype, and gene expression patterns promoted by SMAD3 were noted to have opposing direction of effect compared to another CAD associated TF, TCF21. At sites of SMAD3 and TCF21 colocalization on DNA, SMAD3 binding was inversely correlated with TCF21 binding, due in part to TCF21 locally blocking chromatin accessibility at the SMAD3 binding site. Further, TCF21 was able to directly inhibit SMAD3 activation of gene expression in transfection reporter gene studies. In contrast to TCF21 which is protective toward CAD, SMAD3 expression in HCASMC was shown to be directly correlated with disease risk. We propose that SMAD3 inhibits dedifferentiation that is required for HCASMC to expand and stabilize disease plaque, counteracting the protective dedifferentiating activity of TCF21 and promoting disease risk.

3233F

Genetic and epigenetic fine mapping of complex trait associated loci in the human liver.

M. Caliskan¹, E. Manduchi², H.S. Rao¹, J. Segert¹, M. Holsbach Beltrame¹, M. Trizzino¹, Y. Park¹, A.M. Berrido¹, A. Ches², M.E. Johnson², K.M. Hodge², M.E. Leonard², A.D. Wells², N.J. Hand¹, R.C. Bauer², K.M. Olthoff¹, A. Shaked¹, D.J. Rader^{1,5}, B.E. Engelhardt¹, S.F.A. Grant^{1,7}, C.D. Brown¹. 1) University of Pennsylvania, Philadelphia, PA; 2) Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA; 3) Department of Medicine, Columbia University, New York, NY; 4) Division of Transplant Surgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 5) Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 6) Department of Computer Science, Princeton University, Princeton, NJ; 7) Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

We performed ChIP-Seq experiments for H3K4me3 and H3K27ac histone modifications, annotated putative noncoding functional regions, and identified 972 such regions with genotype-associated activity in the human liver (i.e., hQTL-Peaks). We integrated these findings with genome-wide gene expression data collected in human liver tissue and high-resolution promoter-focused Capture-C data obtained from the human liver-derived HepG2 cell line. We found 2,625 genes with genotype-associated expression levels (i.e., eQTL-Genes) and identified 29,328 significant promoter-H3K4me3 peak and 40,839 promoter-H3K27ac peak interactions. We showed that genes interacting with hQTL-Peaks are significantly more likely to have eQTLs relative to the genes that do not interact with hQTL-Peaks. 116 Gene-Peak pairs in the human liver showed evidence of shared genetic causality. Interestingly, such hQTL-Peaks were often not assigned to their nearest gene; in 71% of the co-regulated gene-peak pairs, there was at least one other gene that is closer to the hQTL-Peak than the eQTL-Gene that it was assigned to. Next, we integrated our findings with GWAS summary statistics for nine phenotypes, including coronary artery disease, blood lipid traits, and blood pressure traits. We found a total of 83 GWAS-Gene and 22 GWAS-Peak pairs with evidence of shared genetic causality. In 44% of the GWAS loci, our findings supported the previously reported gene as the likely causal gene. In 23% of the loci, our findings helped refine the causal gene among several that were suggested to be causal in the literature, and in the remaining loci (33%), our findings pinpointed to novel putatively causal genes in GWAS loci. In loci where we identified putatively causal regulatory region in the absence of a candidate causal gene, we took advantage of our chromatin conformation capture interaction data and identified gene promoters that make contact with the candidate regulatory regions at GWAS loci. Overall, our findings highlight the benefits of integrating multiple cellular traits for the identification and characterization of disease-causing genes and regulatory regions and contribute to basic understanding of genetic and epigenetic regulation of gene expression in the human liver tissue.

3234W

Epigenome-wide methylation survival analyses identify seven loci associated with time to recurrent stroke and composite vascular endpoints in participants from the Vitamin Intervention for Stroke Prevention (VISP) clinical trial.

N.M. Davis Armstrong¹, W.M. Chen^{2,3}, F.C. Hsu⁴, M.S. Brewer¹, S.R. Williams⁵, M.M. Sale^{2,3}, B.B. Worrall^{1,5}, K.L. Keene^{1,6}. 1) Department of Biology, East Carolina University, Greenville, NC; 2) Center for Public Health Genomics, University of Virginia, Charlottesville, VA; 3) Department of Public Health Sciences, University of Virginia, Charlottesville, VA; 4) Department of Biostatistical Sciences, Division of Public Health, Wake Forest School of Medicine, Winston-Salem, NC; 5) Department of Neurology, University of Virginia, Charlottesville, VA; 6) Center for Health Disparities, East Carolina University, Greenville, NC.

Aberrant DNA methylation has been implicated in both cardiovascular disease and stroke. To test for associations of DNA methylation and time to recurrent stroke and composite vascular endpoints, we performed epigenome-wide survival analyses using Cox proportional hazards regression models. Methylation β -values were generated from DNA extracted via whole blood of 180 participants from the Vitamin Intervention for Stroke Prevention (VISP) clinical trial, using Illumina HumanMethylation450K arrays. VISP was a multi-center, double-blind, randomized, controlled trial designed to determine whether daily supplementation of B vitamins (folic acid, B6, and B12) reduced recurrent stroke, myocardial infarction (MI), or death. After filtering, β -values for 473,864 autosomal CpG loci were analyzed, adjusted for covariates, and stratified by race: 76 African Americans (AA) and 104 European Americans (EA). The outcomes of interest included (1) VISP recurrent stroke, defined as a stroke during a two-year follow-up and (2) composite vascular endpoint, described as a coronary artery bypass graft, MI, angioplasty, cardiac collapse resuscitation, fatal coronary endpoint, or VISP recurrent stroke. In total, seven loci reached or exceeded genome-wide significance ($p \leq 1.05 \times 10^{-7}$): four loci for AA recurrent stroke, two for AA composite endpoint, and one for EA composite endpoint. The most significant association was observed in the AA recurrent stroke analysis for methylation locus cg04059318, located in exon 1 of the *KLLN* gene and upstream of *PTEN* ($p = 4.52 \times 10^{-8}$). Studies have suggested that neuronal *PTEN* loss results in neuroprotection against ischemia acutely, but contributes to vascular cognitive impairment long-term. cg00076998, located upstream of *PIK3CB* was the most significant locus associated with the AA composite endpoint ($p = 7.87 \times 10^{-8}$). *PIK3CB* has been described to mediate pulmonary vascular smooth muscle proliferation, as well as being a key regulator in Akt cellular signaling. cg02365967, located in exon 35 of *HERC2*, was the only CpG locus reaching genome-wide significance in the EA analyses (composite endpoint, $p = 8.01 \times 10^{-8}$). *HERC2* is involved in the regulation and prevention of adverse arterial remodeling. In conclusion, our survival analyses identified seven statistically significant loci associated with time to VISP recurrent stroke or composite vascular endpoint, several of which were located in gene regions with plausible stroke and cardiovascular implications.

3235T

Determining the genetic basis of anthracycline-cardiotoxicity by transcriptomic response QTL mapping in a panel of induced cardiomyocytes. D.A. Knowles^{1,2}, C.K. Burrows³, J.D. Blischak³, K.M. Patterson³, D.J. Serie⁴, N. Norton⁵, C. Ober⁵, J.K. Pritchard^{1,6,7}, Y. Gilad^{8,9}. 1) Department of Genetics, Stanford University, Stanford, CA; 2) Department of Radiology, Stanford University, Stanford, CA; 3) Department of Human Genetics, University of Chicago, Chicago, IL; 4) Department of Health Sciences Research, Mayo Clinic, Jacksonville, USA; 5) Department of Cancer Biology, Mayo Clinic, Jacksonville, USA; 6) Department of Biology, Stanford University, Stanford, CA; 7) Howard Hughes Medical Institute, Stanford University, CA; 8) Department of Medicine, University of Chicago, Chicago, IL.

Anthracycline-induced cardiotoxicity (ACT) is a key limiting factor in setting optimal chemotherapy regimes, with almost half of patients expected to develop congestive heart failure given high doses. However, the genetic basis of sensitivity to anthracyclines remains unclear. We created a panel of iP-SC-derived cardiomyocytes from 45 individuals and performed RNA-seq after 24h exposure to varying doxorubicin dosages. The transcriptomic response is substantial: the majority of genes are differentially expressed and over 6000 genes show evidence of differential splicing, the later driven by reduced splicing fidelity in the presence of doxorubicin. We show that inter-individual variation in transcriptional response is predictive of in vitro cell damage, which in turn is associated with in vivo ACT risk. We developed an efficient linear mixed model, *suez*, which detects 447 response-expression quantitative trait loci (QTLs). Combining *suez* with our RNA splicing quantification algorithm LeafCutter we find 42 response-splicing QTLs. These molecular response QTLs are enriched for lower p-values in ACT genome-wide association and enable prediction of cellular damage, supporting the in vivo relevance of our map of genetic regulation of cellular response to anthracyclines. .

3236F

Enhancer-mediated enrichment of interacting JMJD3-DDX21 to *ENPP2* locus prevents R-loop formation and promotes transcription elongation. D. Argaud, M.C. Boulanger, G. Mkannez, P. Mathieu. Quebec Heart and Lung Institute, Université Laval, Quebec, Quebec, Canada.

Introduction: *ENPP2*, which encodes for the enzyme autotaxin (ATX), plays an important role in cancer and in chronic inflammatory disorders such as fibrosis, arthritis, and cardiovascular diseases. Growing evidence suggests that epigenetic processes control inflammation. To this effect, JMJD3, a H3K27me3 demethylase, participates to the regulation of several inflammatory genes. However, the molecular interplay between JMJD3 and other H3K27me3 demethylase such as UTX in regulating inflammation is still largely unknown. In this work, we investigated the role of H3K27me3 demethylase and its interactome on the regulation of the *ENPP2* locus. **Methods:** HEK293T cells were treated with LPS and functional characterization of gene expression was carried out. Chromatin immunoprecipitation coupled with qPCR (ChIP-qPCR) was used to evaluate the recruitment of JMJD3 and its interactome to the *ENPP2* locus. CRISPR-Cas9 was used to evaluate the functional role of a distant-acting enhancer. **Results:** In reporter assay, LPS increased the activity of promoter region of *ENPP2*. LPS also promoted the expression of *ENPP2*, through a NF- κ B pathway, involving the recruitment of p65 on two κ B consensus sequences located in the promoter. Simultaneously, the level of H3K27me3, a histone repressive mark, was removed by UTX. Knockdown of JMJD3 with small interfering RNA (siRNA) prevented LPS-mediated expression of *ENPP2*. JMJD3 controlled the expression of *ENPP2* in a nonenzymatic manner and was recruited to the transcription start site (TSS) and within the gene body. Mass spectrometry analysis revealed a novel interaction for JMJD3 with DDX21, a RNA helicase that unwinds R-loops created by hybridization between nascent RNA transcript and DNA template. ChIP re-ChIP experiments confirmed the co-recruitment of JMJD3-DDX21 as a protein complex at TSS. Upon LPS treatment, JMJD3 is necessary for DDX21 recruitment to TSS and within the gene body, allowing the resolution of aberrant R-loops. CRISPR-Cas9-mediated deletion of a distant-acting enhancer, located at 82kb from *ENPP2*, decreased the expression of *ENPP2* and lowered the recruitment of JMJD3-DDX21 complex at TSS and the progression of this complex through the gene body. **Conclusions:** UTX is necessary for the removal of H3K27me3, whereas enhancer-mediated enrichment of novel JMJD3-DDX21 interaction at *ENPP2* locus is necessary for transcription elongation via the suppression of aberrant R loops formation in response to inflammatory stimulus. .

3237W

Genetic and environmental effects on vascular endothelium gene regulation and cardiovascular traits. A.S. Findley¹, A.L. Richards¹, C. Petrin¹, A.S. Shanku¹, A. Alazizi¹, E. Doman¹, O. Davis¹, N. Hauff², Y. Sorokin², X. Wen³, R. Pique-Regi^{1,2}, F. Luca^{1,2}. 1) Center for Molecular Medicine and Genetics, Wayne State University, Detroit, MI; 2) Department of Obstetrics and Gynecology, Wayne State University, Detroit, MI; 3) Department of Biostatistics, University of Michigan, Ann Arbor, MI.

GWAS and large-scale eQTL studies, such as GTEx, have identified thousands of genetic variants associated with complex traits and gene expression. However, only a limited number of environmental factors are measured in these studies, potentially confounding genetic effect sizes when the underlying environment is heterogeneous. Measuring molecular phenotypes (gene expression, chromatin accessibility) in tightly controlled cellular environments provides a more tractable setting in which to study gene-environment interactions in the absence of other confounding variables. Here we have exposed human umbilical vein endothelial cells (HUVECs) from 17 healthy donors to 4 treatments (retinoic acid, dexamethasone, caffeine, and selenium) and appropriate vehicle-controls for 6 hours. We genotyped and performed RNA-seq and ATAC-seq to model genetic and environmental effects on gene regulation and chromatin accessibility in the vascular endothelium, a common site of pathology in cardiovascular disease. All four treatments induced significant alterations in the transcriptional and chromatin landscapes. We found that genes near regions of differentially accessible chromatin were more likely to be differentially expressed (OR = [3.41, 6.52], $p < 10^{-6}$). Additionally, we have identified transcription factor footprints in each condition and developed predictions for the effect of regulatory variants on transcription factor binding. Combining our treatment-specific predictions with GTEx eQTLs from the aorta, coronary, and tibial arteries, we identified 7,788 eGenes and observed an enrichment in eQTLs for variants predicted to disrupt transcription factor binding during treatment with dexamethasone, caffeine, and retinoic acid, indicating that these environmental conditions are latently present in GTEx samples. Indeed 71 genes have response eQTLs (reQTLs) (FDR = 10%). Using conditional allele-specific expression we validated 50% ($p=10^{-4}$ compared to chance) of the reQTLs that we were able to test. We observed a 3.35-fold enrichment for genes associated with complex traits in GWAS for our reQTL genes compared to genes without a reQTL ($p = 2.32 \times 10^{-9}$). Our approach leverages molecular phenotyping data in response to tightly controlled treatments to characterize the complex interplay between genes and environment and their role in complex traits, highlighting the importance of latent environmental exposures in large-scale datasets.

3238T

Trans-eQTL and GWAS associations at ZNF800 missense variant rs62621812 suggest an extensive regulatory role in adipose tissue. A.K. Iyengar¹, J.P. Davis¹, C.K. Raulerson¹, S.M. Brotman¹, A. Ko², P. Pajukanta^{2,3}, M. Laakso⁴, K.L. Mohlke¹. 1) Department of Genetics, University of North Carolina, Chapel Hill, NC; 2) Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, CA; 3) Molecular Biology Institute at UCLA, Los Angeles, CA; 4) University of Eastern Finland, School of Medicine, Kuopio, Finland.

Subcutaneous adipose *trans*-eQTL studies have identified variants associated with expression level of multiple distant genes (>5 Mb away), suggesting that they serve as master regulators of adipose gene expression. We examined trait associations and molecular and biological mechanisms of an adipose *trans*-eQTL signal represented by zinc finger protein ZNF800 variant rs62621812, encoding Pro103Ser, previously shown to be associated ($P \leq 5 \times 10^{-8}$) with expression level of nine distant genes in a microarray analysis of the Metabolic Syndrome in Men (METSIM) study ($n=770$). To assess the potential contribution of this signal to adipose-related traits, we explored GWAS results in the UK Biobank ($n>478,000$). rs62621812 was associated ($P \leq 5 \times 10^{-8}$) with 19 obesity-related traits including whole body impedance ($P=2 \times 10^{-30}$). To identify additional target genes, we analyzed METSIM adipose RNA-seq-based expression levels ($n=434$). rs62621812 was associated ($P \leq 5 \times 10^{-6}$) with five additional distant genes, and six of the total 14 target genes are known transcriptional regulators, which may expand the network of genes regulated by this signal. The Pro103Ser substitution was also associated with increased expression of ZNF800 in *cis* ($P \leq 3 \times 10^{-16}$). To detect possible molecular mechanisms for the *cis* association, we functionally annotated the variants in linkage disequilibrium ($r^2 > 0.6$, GoT2D EUR) with rs62621812 as well as the strongest GWAS variants ($P \leq 5 \times 10^{-8}$) for whole body impedance. The lead Pro103Ser substitution variant was predicted by Polyphen and SIFT to be benign. Of 30 other variants, two are located in regulatory elements as predicted by adjacent H3K27ac and H3K4me1 marks, interaction with the ZNF800 promoter in adipocyte Capture Hi-C data, and/or cross-species sequence conservation, suggesting that these variants may instead or also regulate expression of ZNF800. To further investigate the role of ZNF800 in adipose-specific gene regulation, we are using a CRISPR-Cas9 approach to knock out ZNF800 function in SGBS and SW872 preadipocytes. We have produced two independent clones that exhibit 50-60% decreased ZNF800 expression, and experiments are underway to further alter ZNF800 function and assess effects on target gene expression and adipocyte function. Determining the effects of this putative master regulator in adipose tissue could provide greater insight into the molecular mechanisms of *trans*-eQTLs and the biological mechanisms underlying obesity.

3239F

Large chromosome 4q25 intergenic deletion disrupts chromatin structure and leads to familial sinus node dysfunction. *H. Murata*^{1,2}, *P. Lindenbaum*¹, *S. Le Scouarnec*¹, *E. Baron*¹, *A. Rajalu*³, *F. Kyndt*^{1,3}, *JF. Deleuze*⁴, *H. Le Marec*^{1,3}, *V. Probst*^{1,3}, *R. Redon*^{1,3}, *JJ. Schott*^{1,3}. 1) l'institut du thorax, INSERM, CNRS, UNIV Nantes, Nantes, France; 2) Department of Cardiovascular Medicine, Nippon Medical School, Tokyo, Japan; 3) l'institut du thorax, CHU Nantes, Nantes, France; 4) Centre National de Recherche en Génomique Humaine, CEA, Evry, France.

Sinus node dysfunction (SND) refers to several conditions causing physiologically inappropriate atrial rates. Genetics of human SND remains poorly understood with few genes identified (*HCN4*, *SCN5A*, *GJA5*, *KCNQ1*, *MYH6* and *ANK2*) and further genetic heterogeneity. Here we report two families with SND segregating with an intergenic deletion associated with long-range *cis*-regulatory elements (CREs) of *PITX2*. We first applied exome sequencing to a 4-generation family presenting 16 patients with SND, atrial fibrillation and/or repolarization abnormalities that did not allow to identify the disease-causing gene within a 16.5-cM linkage interval ($Z_{max}=5.9$, $\theta=0$) on chr.4q25 although immunoblot of a muscle biopsy revealed a striking decrease in ANK-B expression. We then applied 30x whole genome sequencing (WGS) in 2 patients and identified 72 rare ($<0.1\%$) SNVs and a 15-Kb deletion using LUMPY in the candidate interval. This intergenic deletion, which was located between *PITX2* and *ANK2* and contained a CTCF motif, was segregating in all patients and was not found in 855 French control genomes. We also applied WGS on a second family presenting 5 patients with SND and negative for all SND genes and identified a 91-Kb deletion overlapping the initial deletion. CTCF functions as a genome organizer mediating genomic interactions between genes and CREs. To clarify a possible regulatory function of the 1.5-Mb intergenic region containing the deletion, we interrogated 3 epigenetic databases. High-throughput chromosome conformation capture (Hi-C) data (1) revealed that this intergenic region was placed in a topologically associating domain (TAD) containing *PITX2*. Chromatin interaction analysis by paired-end tag (ChIA-PET) data of CTCF in MCF-7 (2) suggested several possible functional chromatin loops in the TAD. Of them, a CTCF-mediated chromatin loop corresponding to a 300-Kb genomic region was clarified as a possible CRE using histone modification data from Roadmap. The region ranging between the intra-deletion CTCF motif and another distal motif was repressed in H3K27me3 profiles of fetal heart and H1 ESC-derived mesenchymal stem cells, while the region was activated in H3K27ac profiles of H1 ESC-derived mesoderm cells. Thus, this CRE may regulate expression of *PITX2* and/or *ANK2* in pacemaker cell differentiation. Ongoing experiments on patient's iPSC-derived cardiomyocytes will further characterize the disease mechanism. (1) Leung *et al.* Nature, 2015 (2) Li *et al.* Cell, 2011.

3240W

Dynamic effects of genetic variation on gene expression during cardiomyocyte differentiation. *K. Rhodes*¹, *R. Elorbany*², *B. Strober*³, *N. Krishnan*⁴, *A. Battle*^{3,4}, *Y. Gilad*^{1,2}. 1) Human Genetics, University of Chicago, Chicago, IL; 2) Genetic Medicine, University of Chicago, Chicago, IL; 3) Biomedical Engineering, Johns Hopkins University, Baltimore, MD; 4) Computer Science, Johns Hopkins University, Baltimore, MD.

Gene regulatory changes during developmental processes contribute to complex traits and diseases in humans. In order to fully understand the mechanisms through which regulatory variants generate phenotypic effects during development, it is necessary to study their effect *through time* as cells differentiate from a pluripotent state to a mature one. Such research has previously been challenging in humans due to ethical and practical limitations regarding the use of human fetal tissues. To address these challenges, we leveraged the ability to differentiate induced pluripotent stem cells (iPSCs) to cardiomyocytes *in vitro*. We differentiated 19 Yoruba iPSC lines to cardiomyocytes and collected RNA-sequencing data every 24 hours for 16 days. We combined this high-resolution time course dataset and existing genotype information to understand temporal patterns of gene regulation. First, we identified *cis*-eQTLs in each time step independently; we found ~100 genome-wide significant eGenes (eFDR = .05) per time step. We determined that eQTL summary statistics from proximal time points are more strongly correlated than distal time points, confirming that time is a primary factor in explaining gene regulatory changes in our data. We next developed a statistical model to identify *cis*-eQTLs whose effects change as a function of time throughout cardiomyocyte differentiation, or "dynamic eQTLs". Using this approach, we found 556 significant dynamic eGenes (eFDR = .01). We characterized the genomic and epigenomic contexts of dynamic eQTLs. We found that early-acting dynamic QTLs are enriched for iPSC enhancer and promoter marks while late-acting dynamic QTLs are enriched for enhancer and promoter marks from primary heart tissue. Lastly, we explored the potential role of these dynamic QTLs in the context of cardiovascular disease. The dynamic QTLs we identified reveal another level of genetically encoded regulation spanning developmental time and could impact cardiac traits and cardiovascular disease.

3241T

Dissecting NF- κ B function using multi-species, multi-tissue epigenomic comparisons. M.D. Wilson^{1,2}, A. Alizada^{1,2}, N. Khyzha^{3,4}, L. Wang^{1,2}, L. Antounians^{1,2}, A. Medina-Rivera^{1,5}, M. Liang^{1,2}, M. Khor^{1,2,4}, J.E. Fish^{3,4}. 1) Genetics and Genome Biology, SickKids Research Institute, Toronto, Ontario, Canada; 2) University of Toronto, Department of Molecular Genetics, Toronto, Canada; 3) University of Toronto, Department of Laboratory Medicine and Pathobiology, Toronto, Canada; 4) University Health Network, Toronto General Hospital Research Institute, Toronto, Canada; 5) Universidad Nacional Autónoma de México, International Laboratory for Research in Human Genomics, Juriquilla, Mexico.

The nuclear factor kappa-light-chain-enhancer of activated B cells (NF- κ B) transcription factor plays an ancient and prominent role in orchestrating inflammation and innate immunity. Genome-wide DNA binding assays of the human NF- κ B subunit RELA (p65) has revealed tens of thousands of NF- κ B binding sites and hundreds of target genes. However, the function of individual RELA binding sites and the extent to which NF- κ B occupancy has evolved within mammals are not well understood. Here we characterized the genome-wide binding of RELA in primary vascular endothelial cells (ECs) isolated from the aortas of human, mouse and cow. ECs were stimulated acutely with the pro-inflammatory cytokine tumor necrosis factor alpha (TNFA). We profiled RELA and JUN occupancy, open chromatin, select histone modifications, and RNA expression revealing distinct modes of RELA binding. We found ~5000 RELA sites conserved in all three species, ~2000 of which were also bound in multiple human cell types. This conserved core of NF- κ B binding sites were enriched for genes controlling and pro-inflammatory responses, human disease mutations, and were also prominent components of ~40 inflammation induced super-enhancers (SEs) common to several tissues. To gain insight into the function of individual conserved NF- κ B binding sites we focused on the inflammation-induced SE proximal to the monocyte recruiting chemokine *CCL2*, which we detected as a SE in all three species and across multiple cell types. We tested the functional significance of six conserved RELA binding sites comprising this SE using CRISPR/Cas9 genome editing. We found that only the most proximal upstream RELA binding site could abolish the induction of *CCL2* upon TNFA treatment. This site also contains a disease-associated variant that can modulate *CCL2* induction. Overall, our cross-species and cross-tissues analyses give new insights into NF- κ B biology, its pathogenesis, and function in inflammation-induced SEs in mammals.

3242F

Promoter interactome of human embryonic stem cell-derived cardiomyocytes connects GWAS regions to cardiac gene networks. M.-K. Choy^{1,3}, B.M. Javierre^{2,4}, S.G. Williams¹, S.L. Baross¹, Y. Liu¹, S.W. Wingett², A. Akbarov¹, C. Wallace^{5,6}, P. Freire-Pritchett^{2,7}, P.J. Rugg-Gunn⁸, M. Spivakov², P. Fraser^{2,9}, B.D. Keavney¹. 1) Division of Cardiovascular Sciences, University of Manchester, Manchester, UK; 2) Nuclear Dynamics Programme, The Babraham Institute, Cambridge, UK; 3) Clare Hall, University of Cambridge, Cambridge, UK; 4) Josep Carreras Leukaemia Research Institute, Campus ICO-Germans Trias I Pujol, Badalona, Barcelona, Spain; 5) MRC Biostatistics Unit, University of Cambridge, Cambridge, UK; 6) Department of Medicine, University of Cambridge, Cambridge, UK; 7) Division of Cell Biology, Medical Research Council Laboratory of Molecular Biology, Cambridge, UK; 8) Epigenetics Programme, The Babraham Institute, Cambridge, UK; 9) Department of Biological Science, Florida State University, Tallahassee, Florida, USA.

Long-range chromosomal interactions bring distal regulatory elements and promoters together to regulate gene expression in biological processes. By performing promoter capture Hi-C (PCHi-C) on human embryonic stem cell-derived cardiomyocytes (hESC-CMs), we show that such promoter interactions are a key mechanism by which enhancers contact their target genes after hESC-CM differentiation from hESCs. We also show that the promoter interactome of hESC-CMs is associated with expression quantitative trait loci (eQTLs) in cardiac left ventricular tissue; captures the dynamic process of genome reorganisation after hESC-CM differentiation; overlaps genome-wide association study (GWAS) regions associated with heart rate; and identifies new candidate genes in such regions. These findings indicate that regulatory elements in hESC-CMs identified by our approach control gene expression involved in ventricular conduction and rhythm of the heart. The study of promoter interactions in other hESC-derived cell types may be of utility in functional investigation of GWAS-associated regions.

3243W

Circulating miR-92a could be a possible biomarker of change in blood pressure: A four-year follow-up study. R. Fujii¹, H. Yamada², E. Munetsuna³, M. Yamazaki⁴, K. Ohashi⁵, H. Ishikawa⁴, K. Maeda⁴, C. Hagiwara¹, Y. Ando⁴, S. Hashimoto², N. Hamajima⁵, K. Suzuki¹. 1) Department of Preventive Medical Sciences, Fujita Health University School of Health Sciences, Toyoake, Aichi, Japan; 2) Department of Hygiene, Fujita Health University School of Medicine, Toyoake, Aichi, Japan; 3) Department of Biochemistry, Fujita Health University School of Medicine, Toyoake, Aichi, Japan; 4) Department of Clinical Biochemistry, Fujita Health University School of Health Sciences, Toyoake, Aichi, Japan; 5) Department of Healthcare Administration, Nagoya University Graduate School of Medicine, Nagoya, Aichi, Japan.

MicroRNAs (miRNAs) are small non-coding RNAs of ~25 nucleotides in length that play a crucial role in several pathophysiological conditions and diseases. Recent experimental studies have revealed that miRNAs are key regulators in developing hypertension via endothelial dysfunction. The association of miRNAs and hypertension in human has been investigated in cross-sectional studies, with few studies in longitudinal analysis. The aim of our study was to examine the relationship between the expression levels of circulating miRNAs and change in blood pressure (BP). A total of 229 residents who participated in the health-check up in both 2012 and 2016, in Hokkaido, Japan were finally analyzed. The study protocol and procedure was approved by the Ethics Review Committee of Fujita Health University. Written informed consent was obtained from all participants. The quantitative real-time polymerase chain reaction was used to determine the expression levels of nine circulating miRNAs that previously reported in cross-sectional studies. Log-transformed values of each miRNA were used in our statistical analysis. Mean blood pressure (MBP) was calculated using systolic blood pressure (SBP) and diastolic blood pressure (DBP). The % change in BP from 2012 to 2016 was calculated according to the following formula: % change of BP = $(BP_{2016} - BP_{2012}) / BP_{2012} \times 100$. Hypertension was defined as SBP ≥ 130 and/or DBP ≥ 85 and taking antihypertensive agents. The mean age (SD) of this population at the baseline survey in 2012 was 63.7 (8.5), and the proportion of male was 45.9%. In a cross-sectional analysis using the baseline data, miR-92a was nominally correlated with MBP after the Bonferroni correction ($P = 0.06$), and then selected for further analyses. The expression level of miR-92a was significantly associated with % change in MBP after adjusting for potential confounders with the standardized β of 0.180, $P = 0.04$), but was not an independent risk factor for incidence of hypertension ($P = 0.30$). We found that the circulating miR-92a level was associated with a significant change in MBP in a population-based study. The results suggested that the circulating level of miR-92a could be a biomarker of change in BP. Further experimental studies are needed to elucidate biological mechanisms underlying this relationship.

3244T

Identification and annotation of cardiac relevant long noncoding RNAs in iPSC-derived cardiomyocytes. M. D'Antonio¹, P. Benaglio², W.W. Greenwald³, A. D'Antonio-Chronowska¹, H. Matsui¹, E.N. Smith¹, K.A. Frazer^{1,2}. 1) Institute for Genomic Medicine, University of California, San Diego, La Jolla, CA; 2) Department of Pediatrics and Rady Children's Hospital, University of California San Diego, La Jolla, CA; 3) Bioinformatics and Systems Biology, University of California San Diego, La Jolla, CA.

Over 97% of the human genome is estimated to be non-coding; yet a large portion of these non-coding regions are still actively transcribed. Tens of thousands of these transcripts are long non-coding RNAs (lncRNAs), and while most have unknown function, some have been found to affect central processes in cardiac biology including metabolism, hypertrophy, conduction, differentiation, and proliferation. Un-annotated lncRNAs could therefore play important roles in human cardiac traits, disease, or development. To facilitate genome-wide identification and annotation of cardiac relevant lncRNAs, we have generated data from iPSC-derived cardiomyocytes (iPSC-CMs) from iPSCORE including: RNA-seq from 144 lines; NKX2-5 ChIP-Seq, H3K27ac ChIP-Seq, and GroSeq from 7 lines; and Hi-C from 13 lines. We initially examined whether patterns of lncRNA regulation were similar to those of protein coding genes, and observed that, as with protein coding genes, high NKX2-5 binding was correlated with high histone acetylation and transcription of lncRNAs in iPSC-CMs. We therefore designed a "cardiac functional prediction score" based on rank of the strength of nearby NKX2-5 peaks, H3K27ac peaks, and GRO-seq transcripts at the TSS (± 2 kb) of each lncRNA genome-wide. This score both recovered known cardiac relevant lncRNAs, including *MALAT1* and *CDKN2B-AS1* (*ANRIL*), and highlighted novel cardiac relevant lncRNAs, including a lowly expressed Gencode lncRNA (*RP11-57G10.8*) with unknown function located 34 kb upstream of *SIRT1*, and an unnamed lncRNA located in *MYLK3*. To identify potential targets of putative novel cardiac relevant lncRNAs, we integrated high-depth iPSC-CM HiC maps with gene expression data from the 144 iPSC-CM lines. We calculated the correlations between the colocalized lncRNA and protein-coding genes and found 22 loops with expression correlations significantly different from zero ($FDR < 0.05$), of which 18 were positively correlated and four were negatively correlated. Interestingly, the most significant association was between the 3' end of *PRDM16*, a transcription factor involved in dilated cardiomyopathy, and *LINC00982*, a lncRNA associated with cell proliferation and cell cycle progression. These analyses highlight similarities between the regulation of protein coding and lncRNA genes and provide a method for identifying and annotating tissue relevant lncRNAs.

3245F

External chromatin accessibility quantitative trait loci (caQTLs) are enriched for association with human GWAS data independently of internal caQTLs. *W.W. Greenwald¹, A. D'Antonio-Chronowska², H. Matsui², E.N. Smith³, K.A. Frazer^{2,3}.* 1) Bioinformatics and Systems Biology Graduate Program, University of California, San Diego, La Jolla, CA; 2) Institute for Genomic Medicine, University of California, San Diego, La Jolla, CA; 3) Department of Pediatrics and Rady Children's Hospital, University of California, San Diego, La Jolla, CA.

Epigenetic information has successfully been leveraged to identify causal variants and genes underlying human diseases. In particular, accessible chromatin has been shown to be highly enriched for trait- and disease-associated genetic variation in relevant cell types. Genetic variants that modify accessible chromatin (i.e. chromatin accessibility quantitative trait loci, caQTLs) are therefore promising candidate causal variants. These variants can be located within accessible chromatin peaks (internal caQTLs) or outside peaks (external caQTLs), but as the few caQTL studies performed to date have been biased toward internal caQTL discovery, the functional impact and disease-relevance of external caQTLs is largely unknown. To study this question, we have undertaken the largest caQTL study to date in 152 induced pluripotent stem cells (iPSCs) and their derived cardiomyocytes (iPSC-CMs). These cell lines come from 134 individuals in iPSCORE for which we have deep whole genome sequencing, gene expression data, and HiC reference maps. We identified ~900k and ~1m ATAC-peaks in iPSCs and iPSC-CMs, respectively – one of the largest sets of ATAC peaks to date. Using matrixQTL, we identified 4,297 ATAC peaks associated with genetic variation (caPeaks; FDR<0.05) in iPSC, and 4,018 in iPSC-CM. Although the majority of caPeaks were cell type specific, 835 of the caPeaks overlapped between the two cell types (OR=43.01, $p<10^{-320}$). Notably, we found 94% of lead caQTLs to be external caQTLs. To understand the disease relevance of caPeaks and caQTLs, we examined GWAS enrichment in iPSC-CM features using FGWAS on UK BioBank data for pulse rate. We observed 2.7 fold enrichment within caPeaks above all ATAC peaks (20 fold above background), and a 7.2 fold enrichment for caQTLs fold above background, even when jointly modeled with the ATAC peaks. As the majority of caQTLs were external, these variants are therefore a highly promising novel annotation that could improve prioritization of causal variants. Ongoing analyses include characterization of how cell-type, chromatin state, and 3D architecture modify how caQTLs and caPeaks are enriched for disease and trait genetic signals across a range of heart relevant GWAS studies. Overall, this study is the first to identify caQTLs in multiple cell types using samples with the same genetic backgrounds, and to show that caQTLs which are external to accessible chromatin regions are highly enriched for association with human GWAS data.

3246W

Susceptibility enhancers for coronary artery disease: From association to mechanism. *M.U. Kaikkonen¹, I. Selvarajan¹, A. Ravindran¹, P. Polonen², A. Toropainen¹, M. Heinaniemi², M. Laakso³, C. Romanoski⁴, M. Civelek⁵.* 1) A.I. Virtanen Institute, University of Eastern Finland, Kuopio, Kuopio, Finland; 2) School of Medicine, University of Eastern Finland, Kuopio, Finland; 3) Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland, Finland; 4) Department of Cellular and Molecular Medicine, University of Arizona, Tucson, USA; 5) Center for Public Health Genomics, Department of Biomedical Engineering, University of Virginia, Charlottesville, VA, USA.

Over 98% of SNPs associated with coronary artery disease are located in the noncoding sequences with yet undefined function. We and others have previously shown that a large fraction of associations affect gene regulation by perturbing enhancers, which are regulatory elements responsible for cell type-specific expression. Importantly, a significant number of susceptibility regions fall within enhancer elements specific to disease-relevant cell type and activation state. Still, a major challenge for understanding the biological function of susceptibility enhancers is the precise identification of their target genes. Here, we studied cell and stimulus-specific enhancers in all of the major cell types present in the atherosclerotic vascular wall; namely, endothelial cells, smooth muscle cells and macrophages. From these data, we discovered thousands of enhancers that overlap with coronary artery disease risk variants. Next, we sought to determine the enhancer target genes by 1) identifying the coding genes in physical contact with the susceptibility enhancers based on chromatin interaction data, 2) performing eQTL analysis in aorta and in the individual cell types, and 3) establishing enhancer-promoter links from correlation in enhancer and gene activity patterns across GRO-seq datasets. Altogether, hundreds of target genes were identified using the three methods, many exhibiting high cell-type specificity. The data collected will be used to identify networks of genes involved in disease pathogenesis and to select genes for functional studies. Altogether, we present here a comprehensive effort to identify and characterize the role of enhancer variants across disease relevant cell types to provide a mechanistic interpretation of the genetic basis of coronary artery disease.

3247T

Unprogrammed presentation number

3248F

Aberrant mechanosensing triggers acute aortic dissections in a mouse model with a low penetrant risk allele in *Myh11*^{R247C}. S. Wang, C. Kwartler, T. Zhang, A. Peters, D. Milewicz. University of Texas Health Science Center at Houston.

The major risk factors for acute aortic dissections are hypertension (HTN) and pathogenic variants in heritable thoracic aortic disease (HTAD) genes. We have determined that individuals with early onset aortic dissections without a family history or syndromic features have an increased burden of rare genetic variants of unknown significance (VUSs) in HTAD genes. Furthermore, a subset of the VUSs identified in these cases are confirmed to disrupt protein function but to a lesser extent and termed these "risk variants". We assessed the role of one of these risk variants by engineering a mouse model with a homozygous *Myh11* missense variant (*Myh11*^{R247C}) in the motor head of the smooth muscle myosin heavy chain. The mutated alteration decreases aortic contractility but does not cause aortic disease in mice. When hypertension is induced by feeding the mice L-NAME and high salt diet, blood pressures were both significantly increased within 2 weeks in wildtype (WT) and mutant mice. With induction of HTN, about 20% of mutant mice (4/22) die within 2 weeks but none of the WT mice (0/20) ($p < 0.05$). Necropsy identified pericardial tamponade due to a retrograde dissection in the aortic root as the cause of death in the mutant mice. Electron microscopy analyses found abnormal formation of focal adhesions (FAs) in the mutant mice compared with WT, leading to greater distortion of FAs in the mutant mice when HTN was induced. RNA sequencing and pathway analysis of the *Myh11*^{R247C} aortas identified the FA signaling as the major pathway altered in the mutant mouse with HTN. Phosphorylated FAK was increased in *Myh11*^{R247C} aortas at baseline compared with WT, and increased further in *Myh11*^{R247C} aortas with HTN induction. When smooth muscle cells (SMCs) were seeded on fibronectin-coated plates to engage FAs, levels of phosphorylated regulatory light chain were dramatically increased in *Myh11*^{R247C} SMCs compared with WT, presumably to compensate for the defective myosin motor. Apoptosis markers, Bid and Bax, were also increased. To determine if FAK signaling and elevated apoptosis were involved in aortic dissection in the mutant mice, either a FAK inhibitor or p53 inhibitor were administered 7 days before and during HTN induction. Both treatments completely prevent dissection deaths ($p \leq 0.05$ for both treatments). These data indicated aberrant mechanosensing through focal adhesions leading to increased apoptosis trigger aortic dissections in the *Myh11*^{R247C} mouse model.

3249W

Single cell profiling of cardiovascular development under hyperglycemic conditions reveals mechanistic links to congenital heart disease. J.E.

Gorzynski^{1,2}, S.C. Juneja², M.E. Slocumb³, S.L. Paige³, J.R. Priest². 1) Genetics, Stanford University, Stanford, CA; 2) Comparative Medicine, Stanford University, Stanford, CA; 3) Pediatric Cardiology, Stanford University, Stanford, CA.

Background: Maternal diabetes of any etiology (type 1, type 2, gestational) is known to increase the risk of congenital heart disease in offspring, an effect linked to elevated glucose levels in multiple clinical observations. However, the mechanism by which a hyperglycemic fetal environment resulting from increased maternal blood glucose alters cardiac development is unknown.

Methods: We performed single cell transcriptomics of explanted cardiac tissue from embryonic day 12.5 murine hearts. The hearts were exposed to physiologic conditions (normal glucose & insulin) or diabetic conditions (elevated glucose & insulin). Cell types were identified and segregated using previously validated transcriptional markers. Within each cell type cluster, control and experimental groups were compared to identify cell-type specific transcriptome changes in response to diabetic conditions. **Results:** We identified four cardiac specific cell types: cardiomyocytes, epicardial cells, endothelial cells and mesenchymal cells. In all cell types we observed an anabolic response to hyperglycemia (increased transcription of ribosomal and decreased transcription of mitochondrial genes). The diabetic group of cardiomyocytes and mesenchymal cells showed decreased expression of *ACTA2* (pval=6.56e-33, 2.25e-17 respectively). Additionally, endothelial cells in the diabetic group showed a decrease in *PFN1* (pval 7.33 e-23), a gene which codes for an actin binding protein. **Discussion:** Our novel data highlight the utility of single-cell transcriptome analysis in identifying cell-type specific responses to disease conditions. Genes involved in protein synthesis and cellular respiration were perturbed in all cell types, while changes in expression to genes responsible for cellular structure and contraction were unique to cardiomyocytes, mesenchymal cells and endothelial cells. This suggests a combinatorial effect of multiple cell types in mediating the emergence of congenital heart disease in the setting of maternal diabetes, representing multiple potential therapeutic targets for disease prevention.

655T

Juvenile hyaline fibromatosis is an autosomal recessive genetic disease.

Four cases report. J. Aparicio¹, M.L. Hurtado Hernandez², M. Barrientos Perez³, S. Chatelain Mercado⁴. 1) Genetics; 2) Cytogenetics; 3) Endocrinology, Hospital Para El Niño Poblano, San Andrés Cholula Puebla, Mexico; 4) Biotechnology, Metropolitan University of Mexico City.

Juvenile hyaline fibromatosis (JHF) is considered a non frequent genetic, autosomal recessive; characterized by connective tissue disorder characterized by multiple subcutaneous nodules, gingival hypertrophy, osteolytic lesions, and joint deformities (contractures). As one of the main clinical symptoms, are multiple subcutaneous nodules, articular contractures, and gingival hypertrophy after 6 months of life. After one year old the subcutaneous nodules appears more frequently among teenagers a great variety of deformations are observed with movement limitations. 70 cases with JHC approximately are reported in the literature. The hyaline fibromatosis juvenile (FHJ) is a rare genetic disease of connective tissue. It is characterized by multiple cutaneous nodules, hypertrophy, gingival, joint contractures and osteolytic injury FHJ is associated to genes that encodes the protein-2 morphogenesis (CMG2 or ANTXR2) located on the long arm of chromosome 4 (4q21.21). FHJ has an autosomal recessive Mendelian Inheritance pattern. The official name of this gene is "anthrax toxin receptor 2." Mutations in the ANTXR2 gene (also known as the CMG2 gene) cause FHJ. The ANTXR2 gene is therefore unchanged of producing a specific protein which is associated with the formation of small blood vessels (capillaries). It is believed that this particular ANTXR2 protein keeps the normal structure of basement membranes, which are tiny structures that separate and support cells in a great variety of tissues. The main clinical features of the disease are caused by the accumulation of a clear (hyaline) substance in different parts of the body. The nature of this substance is not clear, but its chemical composition is made up of a protein and sugar molecules. It is believed that mutations in the ANTXR2 gene disrupt the formation of basement membranes, was the hyaline substance leaks through and alters various structure tissues body.

656F

Whole exome sequencing to identify candidate genes associated with hereditary predisposition to UM.

M.H. Abdel-Rahman^{1,2,3}, K.M. Sample¹, B. Kelly⁴, D. Gordon⁴, P. Johansson⁵, R. Pilarski^{2,3}, G. Boru¹, T. Grosse¹, J.B. Massengill¹, D.D. Kinnamon², F.H. Davidoff¹, N. Hayward⁶, P. White^{4,6}, C.M. Ce-bulla^{1,3}. 1) Ophthalmology Department, The Ohio State University, Columbus, Ohio; 2) Division of Human Genetics, The Ohio State University, Columbus, Ohio; 3) The Ohio State University Comprehensive Cancer Center, Columbus, Ohio; 4) The Institute for Genomic Medicine, Nationwide Children's Hospital, Columbus, Ohio; 5) QIMR Berghofer Medical Research Institute, Brisbane, Australia; 6) Department of Pediatrics, The Ohio State University, Columbus, Ohio.

Objective: About 12% of uveal melanoma (UM) patients have features suggestive of hereditary cancer predisposition. Germline mutation in *BAP1* explains only 3-4% of these patients, suggesting the existence of other candidate genes. The goal of this study was to identify novel genes associated with hereditary predisposition to UM. **Methods:** Exome sequencing was carried out for 29 unrelated high-risk UM patients with no detectable mutations or deletions in *BAP1* and 10 of their relatives. The probands included 25 with familial UM, one with bilateral UM, one with congenital UM and 2 patients with strong family history of cancers observed in the BAP1-Tumor Predisposition. The relatives included 7 with UM, one with ovarian cancer and the unaffected parents of the congenital UM case. UM relatives were first degree in two instances and third degree in four. **Results:** We focused our initial analysis on genes with established association with hereditary predisposition to cancer, which are routinely tested in diagnostic laboratories. We identified coding pathogenic or potentially pathogenic variants in 5 cancer predisposition genes (*CHEK2*, *MLH1*, *PALB2*, *SMARCE1*, and *RET*) in 5 patients. In addition, 18 missense variants of uncertain significance (based on ACMG criteria) were identified in 11 patients. Biallelic inactivation of *PALB2* (somatic frameshift mutation and low RNA expression) and *MLH1* (loss of heterozygosity and loss of protein expression) were observed in the tumors from the respective patients. No evidence of biallelic inactivation of *CHEK2* or *SMARCE1* was observed in the tumors. Given the rarity of UM, we next assessed unreported or rare (MAF<0.0001) variants in potential cancer associated genes. Variants in cancer associated genes, mostly missense variants of uncertain significance, were identified in 24/29 patients. In 17 of those, more than one gene was impacted. Several of these variants were in genes in the DNA damage repair pathway. In one family, we confirmed functional alterations in three potential candidate tumor suppressor genes. **Conclusions:** The study provides limited/moderate evidence of gene/disease association of germline mutation in *PALB2* and *MLH1* with hereditary predisposition to UM. It also identifies several other potential candidate genes. The results suggest locus heterogeneity in predisposition to UM and potential multi-genic etiology in a subset of patients. Validation in a larger UM cohort is warranted.

657W

Clinical relevance of screening tools for detection of cancer predisposition syndromes in an Asian cohort of childhood tumors. S. Chan¹, W. Chew¹, N.D. Ishak¹, W. Lim², S. Li¹, S. Tan³, J. Teo², T. Shaw¹, K. Chang³, Y. Chen³, P. Iyer³, E. Tan³, M. Seng³, M. Chan³, A. Tan³, S. Low⁴, S. Soh³, A. Loh³, J. Ngeow¹. 1) National Cancer Centre Singapore, Singapore; 2) SingHealth Duke-NUS Institute of Precision Medicine (PRISM), Singapore; 3) KK Women's and Children's Hospital, Singapore; 4) National Neuroscience Institute, Singapore.

PURPOSE Detection of cancer predisposition syndromes (CPS) is a challenge for pediatric oncologists, especially in cohorts with unknown prevalence of germline mutations. We assessed the utility of two clinical screening checklists for CPS validated by genomic sequencing and concurrently profiled the prevalence and spectrum of germline mutations in our Asian cohort of pediatric solid tumors. **PATIENTS AND METHODS** We evaluated 102 patients under age 19 years recruited at our tertiary medical center over a period of 31 months. Patient records were reviewed against two published checklists and germline mutations in 100 cancer-associated genes were profiled through a combination of whole-exome sequencing and multiplex ligation-dependent probe amplification on blood-derived genomic DNA. **RESULTS** Fifty-four patients met clinical criteria for CPS on both checklists, of which 10 (18.5%) harbored germline mutations in associated genes identified by sequencing. Specificity of screening was markedly improved by combining criteria of both checklists without effect on sensitivity. The 10 children (9.8%) had pathogenic germline mutations in six known cancer predisposition genes: *TP53*, *DICER1*, *NF1*, *FH*, *SDHD* and *VHL*. *TP53* was most frequently mutated, affecting five children with adrenocortical carcinoma, sarcomas and diffuse astrocytoma. Multiple pathogenic germline mutations were found in two patients. Disparity in prevalence of germline mutations across tumor types suggested variable genetic susceptibility and implied potential contribution of novel susceptibility genes. A family history of cancer was observed in only 50% of children with pathogenic germline mutations. **CONCLUSION** Currently available CPS screening checklists are adequately sensitive to detect at-risk children and are relevant for clinical application. Our study showed that 9.8% of Asian pediatric solid tumors have a heritable component, consistent with cohorts from other populations.

658T

Novel *PTEN* gene variant associated with growing teratoma syndrome.

S.L. Cole¹, B.P. Tullius^{1,2}, S. Aradya³, V. Triano¹, E.C. Huang⁴, T. Sanchez⁵, A. Pawar⁶, S.P. Shankar¹. 1) Division Genomic Medicine, Department of Pediatric, UC Davis Health, Sacramento, CA; 2) Division of Hematology/Oncology/BMT, Nationwide Children's Hospital, Columbus, OH; 3) Invitae, San Francisco, CA; 4) Department of Pathology and Laboratory Medicine, University of California, Davis Medical Center, Sacramento, CA; 5) Department of Radiology, University of California, Davis Medical Center, Sacramento, CA; 6) Department of Pediatrics, University of California, Davis Medical Center, Sacramento, CA.

Pathogenic variants in *PTEN* have been reported to cause autosomal dominant cancer predisposition syndromes including Bannayan-Riley-Ruvalcaba syndrome, Cowden syndrome 1, Lhermitte-Duclos syndrome and Macrocephaly/autism syndrome. We report the **first** association of a novel heterozygous pathogenic variant c.685_686dupTC (p.Gly230Glnfs*27) in *PTEN* in a 2 year 4 month old female, presenting with ovarian germ cell tumor (OGCT) complicated by growing teratoma syndrome (GTS). Our patient, the youngest reported thus far, presented with urinary urgency and abdominal pain for the preceding 5 months, that acutely worsened on the day of presentation. An abdominal CT scan revealed a soft tissue mass occupying the pelvis. AFP was elevated to 3057.1 ng/mL and beta-HCG was 2554 mIU/mL. She underwent gross total resection of the left adnexal mass accompanied by a left salpingo-oophorectomy. Histologically, the tumor consisted predominantly of malignant teratoma - adenocarcinoma from the mature teratoma admixed and choriocarcinoma consistent with a malignant OGCT staged as FIGO Ib. After surgery, the patient received 3 cycles of adjuvant chemotherapy consisting of cisplatin, etoposide, and bleomycin. During her treatment, her serum HCG became undetectable 1.5 months after resection, and AFP dropped to normal levels at 4 months after resection. Seven months after completion of her chemotherapy, a surveillance CT scan identified a new mass at the original site. Tumor markers AFP, beta-HCG, as well as CA125 were not elevated. The patient underwent surgical resection. Pathology showed a mature teratoma with predominantly intestinal/mucinous differentiation. Retinoblastoma 1 (pRB) protein immunostaining was negative. The second tumor gave rise to concern for GTS. She was referred to genetics where growth parameters were noted at > 2 standard deviations above the mean and scattered nevi, and a hypopigmented macule were identified. Overgrowth syndrome was suspected, and genetic testing was obtained. An overgrowth panel revealed, a novel heterozygous pathogenic variant c.685_686dupTC (p.Gly230Glnfs*27) in *PTEN*. Parental testing is pending. Immature germ cell tumors are rare in the pediatric population and the emergence of GTS an even more rare complication following the treatment of this type of malignancy. The *PTEN* finding changes her management and we recommended screening for *PTEN* associated tumors per NCCN Clinical Practice Guidelines in Oncology.

659F

Assessing the prevalence and clinical implications of secondary findings for hereditary cancer predisposition among patients with chronic kidney disease. E. Groopman¹, M. Marasa¹, S. Sanna-Cherchi¹, K. Kiryluk¹, D. Goldstein^{2,3}, A. Gharavi^{1,2}. 1) Department of Medicine, Division of Nephrology, Columbia University, New York, NY, USA; 2) Institute for Genomic Medicine, Columbia University, Hammer Health Sciences, 1408, 701 West 168th Street, New York, NY, USA; 3) Department of Genetics & Development, Columbia University, Hammer Health Sciences, 1702, 701, West 168th Street, New York, NY, USA.

The broadening use of whole-exome sequencing (WES) necessitates assessment of not only diagnostic yield but also the rate of secondary findings across clinical indications. The latter include hereditary cancers, whose detection among patients referred for non-cancer indications can prompt further testing and shape care. WES is increasingly applied for patients with chronic kidney disease (CKD), for whom secondary findings for hereditary cancer have high clinical relevance, given that immunosuppression is a first-line therapy for many CKD subtypes and in renal transplantation. Yet, the prevalence of pathogenic variants in germline cancer predisposition genes among CKD patients has not been assessed. We assessed the frequency of secondary findings for hereditary cancers among 2,521 patients exome sequenced for all-cause CKD at Columbia University Medical Center. We analyzed WES data for pathogenic variants in a set of 148 known germline cancer predisposition genes, including all the ACMG 59 cancer predisposition genes and those deemed medically actionable germline cancer genes by ClinGen. We examined yield and implications for clinical care. The cohort had median age 46y; 43.5% were female and 48.2% self-identified as non-white Europeans. Overall 4.4% of cases had pathogenic variants, spanning 46 of the 148 hereditary cancer genes tested. Yield did not differ significantly between pediatric and adult (age > 21y) cases (OR=1.28 [0.67, 2.67], P=0.55) or Europeans and non-White Europeans (OR=0.96 [0.64, 1.42], P=0.85). 31 (1.2%) individuals had a pathogenic variant in an ACMG 59 cancer gene: findings in *BRCA1/2* (N=14) and *PMS2* (N=4) were most frequent. A further 0.7% had pathogenic variants in other highly actionable cancer genes, including *ATM* (0.36%), *CHEK2* (0.12%), and *PALB2* (0.04%). In addition, 1.9% had lower penetrance cancer risk alleles, including *APC* p.I1307K (1.2%), *HOXB13* p.G84E (0.2%), and *MITF* p.G312K (0.2%). Beyond indicating subspecialty referral, the genetic diagnosis would directly inform nephrologic care, including immediately shaping therapy for the 95 cases receiving immunosuppression (for their primary nephropathy or renal transplant). Our results reveal a novel utility of genomic sequencing to yield findings which while not explicative for the primary test indication, nonetheless inform its management, encouraging expanded use of WES in CKD and supporting similar studies in other disease cohorts.

660W

Exome sequencing in high breast and ovarian cancer incidence families which lack detectable mutations in established cancer susceptibility genes. P.A. Kenny, M.A. Girtman, J.A. Deviley, D.R. Meier, C.R. Vos, C.S. Richmond, K.A. Lofgren. Kabara Cancer Research Institute, Gundersen Medical Foundation, La Crosse, WI.

At our center, a standard 25 gene panel has proved useful for the identification of pathogenic germline mutations in many families with a high penetrance of breast and ovarian cancer, yet there remain several families with very high cancer incidence but from whom these assays have not identified pathogenic alleles. To attempt to define the genetic mechanism of susceptibility in such cases, are conducting germline whole exome sequencing in affected individuals. Here we describe the identification of strong candidate susceptibility variants in two such families. Both variants cause loss-of-function of DNA repair genes not previously implicated in breast cancer, but which share properties with genes with known roles in maintenance of genome integrity. In one individual, with a personal history of ovarian cancer and two independent breast tumors, we identified a novel point mutation in *POLD2* which introduces a mutation at a residue which is biochemically conserved in all animal species thus far sequenced. In the second individual, a frameshift mutation was detected in *ERCC4*. In both cases, the variants found have not been observed in large exome datasets (e.g. ExAC) or reported in ClinVar. Our study highlights the potential utility of whole exome sequencing as a discovery tool in those families with high cancer incidence yet which lack mutations in known hereditary cancer predisposition genes.

661T

APC testing in pediatric patients. S.C. Knapke, L. Susswein, E. Webb, K.J. Postula, Y. Wang, R.T. Klein, K.S. Hruska. GeneDx, Gaithersburg, MD.

Background: Familial Adenomatous Polyposis (FAP) is a genetic syndrome characterized by extensive colorectal polyposis and high risk for colorectal cancer due to pathogenic variants in *APC*. Additional cancers, benign tumors and clinical features are also part of the phenotype for individuals with FAP. In many cases, FAP presents in childhood or adolescence. Data are currently lacking, however, with regards to the utility of *APC* gene testing in pediatric patients with personal histories suggestive of FAP. **Purpose:** This descriptive analysis aims to identify common indications for *APC* gene testing and calculate the overall and indication-specific yields in pediatric patients with FAP-related features. **Methods:** Test results and clinical history were reviewed for pediatric patients undergoing hereditary cancer testing that included *APC*. Individuals were included if they were <18yrs at testing and reported a personal history of FAP-related features. Individuals who had targeted variant testing were excluded. Descriptive statistics were employed. **Results:** Among 186 pediatric patients with a personal history of FAP-related features, 12 (6.5%) were found to have a pathogenic or likely pathogenic variant (PV) in *APC*. The most common indications for testing included colorectal adenomatous polyps/polyposis (30), CHRPE (17), medulloblastoma (16), desmoid (15), pilomatixoma (11), colorectal cancer (9) and hepatoblastoma (8). Yield of testing was highest in individuals with colorectal adenomatous polyps/polyposis (7/30, 23.3%), fibroma/fibromatosis (2/12, 16.7%) and medulloblastoma (2/16, 12.5%). No *APC* PVs were identified in individuals with CHRPE, pilomatixoma, colorectal cancer, desmoid or hepatoblastoma. The *APC* yield was highest for individuals whose testing included *APC* only (7/51, 13.7%), followed by a large multigene colorectal cancer panel (6/56, 10.7%). **Conclusions:** The approach to, indications for, and yield of *APC* gene testing in pediatric patients is highly variable. These results support the pursuit of genetic counseling and consideration of *APC* testing for pediatric patients with diagnoses including adenomatous polyps/polyposis, medulloblastoma and fibromas. Although no PVs were identified in individuals with CHRPE or hepatoblastoma, *APC* testing should still be considered for these indications based on previously published data. These results also suggest that testing for *APC* only or a multigene colorectal panel can be considered with comparable yield.

662F

An essential role for *BRCA2* in ovarian development and function. E. Levy-Lahad^{1,2}, A. Weinberg-Shukron^{1,2}, M. Rachmiel³, P. Renbaum¹, S. Gulsuner^{4,5}, T. Walsh^{4,5}, O. Lobel¹, A. Dreifuss⁶, A. Ben-Moshe⁷, S. Zeligson¹, R. Segel^{1,2}, T. Shore⁶, R. Kalifa⁶, M. Goldberg⁷, M.-C. King^{4,5}, O. Gerlitz², D. Zangen^{2,8}. 1) Shaare Zedek Medical Center, Jerusalem, Israel; 2) Faculty of Medicine, The Hebrew University of Jerusalem, Hadassah Medical School, Jerusalem, Israel; 3) Pediatric Endocrinology Clinic, Assaf Harofeh Medical Center, Zerifin, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel; 4) Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA; 5) Department of Genome Sciences, University of Washington, Seattle, WA; 6) Department of Developmental Biology and Cancer Research, IMRIC, Faculty of Medicine, The Hebrew University of Jerusalem, Jerusalem, Israel; 7) Department of Genetics, Alexander Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Israel; 8) Division of Pediatric Endocrinology, Hadassah Hebrew University Medical Center, Jerusalem, Israel.

Genetic characterization of ovarian dysgenesis patients has revealed critical pathways in ovarian development, but the process remains incompletely understood. In two sisters of non-consanguineous Ethiopian origin with XX-ovarian dysgenesis, whole exome sequencing revealed compound heterozygosity for nonsense truncating *BRCA2* mutations. Compound heterozygotes had reduced *BRCA2* mRNA (3.5-fold, $p=2.29e-06$) and reduced *BRCA2* protein (7-fold, $p=6.7e-12$). DNA damage response was impaired, with pronounced chromosomal breakage and failure of RAD51 recruitment to DNA damage sites. The *BRCA2* nonsense mutations are modeled in *Drosophila* by knockout of the *BRCA2* ortholog. *Dmbrca2* null flies (*Dmbrca2*^{KO}) were sterile, with gonadal dysgenesis in both sexes. Ovarioles in female *Dmbrca2*^{KO} had extensive disintegration of their egg chambers before developing to stage 10, with cell envelope destruction, DNA condensation, and extensive apoptosis. Testes from *Dmbrca2*^{KO} males were smaller and underdeveloped, with aberrant DNA location, lack of actin structures and abnormal expression of cleaved caspase-3. The patients' phenotype was significantly milder than the Fanconi anemia D1 syndrome previously described with biallelic *BRCA2* mutations. Beyond the major presentation of ovarian dysgenesis, the sisters (ages 15y and 20y) had microcephaly with normal intelligence; one sister had survived leukemia at age 5y following standard therapy. Originally there was no family history of breast or ovarian cancer. When the mother was identified as a *BRCA2* carrier, surveillance ultrasound revealed stage III ovarian cancer. These results reveal a novel role for *BRCA2* in ovarian development. Homologous recombination, in which *BRCA2* plays a critical role, is important for both DNA damage repair and for normal meiosis. Several recently identified genetic bases of ovarian dysgenesis are genes encoding proteins involved in meiotic processes such as homologous recombination and chromosome segregation (e.g. *HFM1*, *MCM8*, *MCM9*, *SYCE1*, *STAG3*, *SPIDR*). Our findings thus indicate an emerging concept of ovarian dysgenesis as a consequence of defects in meiotic processes, many of which are also important for DNA repair. This suggests that chromosomal breakage assays may be a useful screening tool in ovarian dysgenesis. Increased chromosomal breakage could suggest a DNA repair defect as the underlying genetic basis, as well as indicate possible cancer predisposition requiring surveillance.

663W

Germline lysine-specific demethylase 1 mutations confer susceptibility to multiple myeloma. S. Lipkin¹, X. Wei¹, M. Calvo-Vidal¹, S. Chen¹, M. Revuelta¹, V. Joseph², C. Snyder³, C. Vachon⁴, J. McKay⁵, D. Jaybaylan¹, R. Waller⁶, R. Niesviszky¹, O. Landgren², S. Slager⁷, J. Churpek⁸, L. Godley⁹, J. Garber⁹, K. Anderson⁹, M. Daly¹⁰, M. Armutov¹, C. Dumontet¹, H. Lynch¹, N. Camp⁹, K. Offit¹, R. Klein¹⁰, H. Yu¹, L. Cerchietti¹. 1) Weill Cornell Medicine, New York City, NY; 2) Memorial Sloan-Kettering Cancer Center, New York, New York; 3) Creighton University, Omaha, Nebraska; 4) Mayo Clinic, Rochester, Minnesota; 5) Cancer Research Center of Lyon, Lyon, France; 6) University of Utah, Salt Lake City UT; 7) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts; 8) University of Chicago, Chicago, Illinois; 9) Dana-Farber Cancer Institute, Boston, Massachusetts; 10) Icahn School of Medicine, New York, New York.

We identified familial and early-onset multiple myeloma kindreds with truncating mutations in lysine-specific demethylase 1 (LSD1/KDM1A), an epigenetic transcriptional repressor that primarily demethylates histone H3 on lysine 4 and regulates hematopoietic stem cell self-renewal. In addition, we found higher rates of germline truncating and predicted deleterious missense KDM1A mutations in patients with multiple myeloma unselected for family history compared with controls. Both monoclonal gammopathy of undetermined significance (MGUS) and multiple myeloma cells have significantly lower KDM1A transcript levels compared with normal PCs. Transcriptome analysis of multiple myeloma cells from KDM1A mutation carriers shows enrichment of pathways and MYC target genes previously associated with myeloma pathogenesis. In mice, antigen challenge followed by pharmacologic inhibition of KDM1A promoted PC expansion, enhanced secondary immune response, elicited appearance of serum paraprotein, and mediated upregulation of MYC transcriptional targets. These changes are consistent with the development of MGUS. Collectively, our findings show that KDM1A is the first autosomal dominant multiple myeloma germline predisposition gene providing new insights into its mechanistic roles as a tumor suppressor during post-germinal center B-cell differentiation. We will provide updated data on mouse models of Kdm1A mutation as well.

664T

An atypical late-onset case of homozygous ATM variant supported by a complex molecular mechanism of mosaic uniparental isodisomy. S. Nambot^{1,2,3}, V. Goussot⁴, A. Fievet^{5,11}, R. Boidot⁴, L. Arnould⁴, N. Vulquin⁶, R. Arjmand⁷, V. Derangere⁴, J. Albuissou⁴, F. Ghiringhelli⁸, N. Foray⁹, JP. De Villartay¹⁰, G. Pierron⁵, M. Chevarin², A. Baurand¹, D. Stoppa-Lyonnet¹¹, A. Vitobello^{2,12}, F. Tran-Mau Them^{1,2,12}, C. Philippe^{1,2,12}, C. Thauvin-Robinet^{1,2,3,12}, Y. Duffourd^{1,2}, L. Faivre^{1,2,3,12}. 1) Centre de Génétique et Centre de Référence Maladies Rares «Anomalies du Développement de l'Interrégion Est», Hôpital d'enfants CHU Dijon Bourgogne, Dijon, France; 2) Inserm UMR 1231 GAD «Génétique des Anomalies du Développement», Université de Bourgogne, Dijon, France; 3) Fédération Hospitalo-Universitaire Médecine Translationnelle et Anomalies du Développement (FHU TRANSLAD), CHU Dijon Bourgogne et Université de Bourgogne-Franche Comté, Dijon, France; 4) Département de biologie et pathologie des tumeurs, Centre Georges François Leclerc, Dijon, France; 5) Department of Genetics and Biology of Cancers, Inserm UMR 830, Institut Curie, PSL Research University, Paris 75248, France; 6) Département de radiothérapie, Centre Georges François Leclerc, Dijon, France; 7) Service de Neurologie, CHU Dijon, Dijon, France; 8) Département d'oncologie médicale, INSERM LNC U1231, Centre Georges François Leclerc, Dijon, France; 9) INSERM UMR 1052, Centre de Recherche en Cancérologie de Lyon Groupe de Radiobiologie, LYON, France; 10) INSERM UMR 1163, Laboratoire Dynamique du Génome et Système Immunitaire Université Paris Descartes - Sorbonne Paris Cité, Institut Imagine, Paris, France; 11) Institut Curie, Service de Génétique, Paris, France; 12) UF Innovation en diagnostic génomique des maladies rares, CHU Dijon Bourgogne, Dijon, France.

Introduction: Ataxia telangiectasia (AT) is a rare autosomal-recessive disorder caused by variants in the AT-mutated (*ATM*) gene. It is generally characterized by childhood-onset cerebellar ataxia, telangiectasia, oculomotor apraxia, cancer predisposition, particularly lymphoid tumors, immunodeficiency, and hyper-radiosensitivity. The phenotypic studies show a continuous spectrum from severe classical childhood-onset AT to a mild adult-onset disorder, depending on the residual kinase activity. Germline *ATM* heterozygosity is estimated to 0.5%-1% of the general population, and provides a 2 to 4 fold increase in breast cancers risk. **Materials and Methods:** We present the atypical case of a 69-year-old man with breast cancer diagnosed at 65 years. Diagnostic gene panel test in DNA extracted from blood detected a c.7271T>G missense *ATM* variant with an allelic frequency of 90%, confirmed by Sanger sequencing.

Results: Reverse phenotyping showed an attenuated AT phenotype, as previously reported with this variant, including essential tremor since childhood, signs of cerebellar and proprioceptive ataxia, absence of telangiectasia and oculomotor apraxia. Additional tests revealed an inversion of the chromosome 7 in 4 out of 50 metaphases, compatible with a late-onset AT, minor elevation of immunoglobulins, and normal rate of alpha foeto-protein. Tumor sequencing revealed an allelic frequency for the variant at 85%, and at 69% in the peritumoral tissue (allelic frequency in DNA extracted from fibroblasts and saliva in progress). Of note, the proband's daughter is non-carrier of the variant, confirming the existence of wild type allele in some germinal cells. After excluding other causes of mosaicism (sample contamination, recent transfusion, leukemia, bone marrow transplant), SNP-array was performed and evidenced an isodisomy of the whole 11q chromosome. These molecular results permitted to adapt the therapy of the patient, because of an indication of radiotherapy for bone metastasis. Functional studies of the patient's cells showed a strong spontaneous genomic instability with partial and delayed activation of the *ATM* protein, in favor of a moderate radiosensitivity as compared in homozygous *ATM* carriers, with a high risk of radiation-induced complications, leading to a fractionated and spread radiotherapy. **Conclusion:** The diagnosis work-up of a sporadic male breast cancer have led to the identification of a late onset AT supported by a mosaic uniparental isodisomy.

665F

A mouse model to study tissue-specific response to Brca1 loss. *J.O Olayiwola, S.G Rao, J.G Jackson.* Department of Biochemistry and Molecular Biology, Tulane University School of Medicine, New Orleans, LA.

Female *BRCA1* mutation carriers have an 85-90% lifetime risk of developing breast cancer by age 70. Although all cells in the body of a carrier are heterozygous for *BRCA1*, tumors overwhelmingly develop in the breast. Presently, there is no satisfying mechanistic explanation for this tissue specificity. To address this, we created a mouse model of tet-inducible Cre recombinase and combined with a tdTomato fluorescent protein reporter allele. Using adult mice with all the required alleles, and control mice lacking one or two alleles, we administered water or doxycycline ad libitum and then tested for Cre expression and tdTomato expression following recombination. In mice with all required alleles given doxycycline, fluorescence microscopy showed virtually all cells underwent recombination. Very low levels of leakiness as shown by infrequent tdTomato expression was present in tissues of mice administered water. Thus, we show a system of temporally controlled Cre induction in all tissues of adult mice. To investigate and compare responses of different tissues to loss of *Brca1*, we bred our doxycycline inducible Cre mice to *Brca1^{F22-24F22-24}* p53^{-/-} male mice to generate inducible Cre, *Brca1^{F22-24F22-24}* p53^{-/-} mice. Doxycycline administration to these mice led to *Brca1* recombination in all tissues examined as evidenced by excision of exons 22 through 24; which holds the second BRCT domain. Preliminary studies in a small cohort have shown that mice can survive up to 3 months post *Brca1* recombination, the recombination is depicted in mammary, intestine, liver and tail DNA of this mice. Currently, we are examining acute cellular responses and the p53 pathway response in tissues, as well as mouse survival in response to *Brca1* deletion.

666W

Evidence of reduced penetrance in BRCA2 truncating variant c.658_659del (p.Val220Ilefs*4). *N. Singh, E. Mundt, B. Roa, S. Manley, K. Bowles.* Myriad Genetic Laboratories, Inc., Salt Lake City, UT.

Background: The cancer risks associated with pathogenic variants in *BRCA2* have been well studied. Although these cancer risks are known to vary, it can be difficult to assess variant specific penetrance estimates. Here we present evidence of lower cancer risks associated with truncating variant c.658_659del (p.Val220Ilefs*4) based on a laboratory specific classification tool, Pheno analysis, with support from published literature. Pheno analysis is a computerized statistical tool used to provide evidence for variant classification. **Methods:** All evidence of pathogenicity for the variant was evaluated both at the time of initial classification and over time as new data became available. Pheno calculates a variant specific score based on the personal and family cancer histories of each proband with the variant and compares those scores to matched controls. Pheno has been thoroughly validated and tested showing PPVs and NPVS of >99.5%. **Results:** *BRCA2* c.658_659del (p.Val220Ilefs*4) creates a frameshift inserting four novel amino acids and predicts premature protein truncation leading to an initial classification of a pathogenic variant (PV). Although a truncating variant in the middle of the gene is expected to confer loss of function and typical *BRCA2* cancer risks, the Pheno analysis shows that the cancer histories for multiple unrelated probands with this specific variant are less severe than the matched controls with other *BRCA2* PVs. In addition, this variant has been observed in trans with other *BRCA2* PVs in many individuals with Fanconi anemia. The collective evidence support this variant having reduced penetrance, as it does not appear to be associated with embryonic lethality as expected with some highly penetrant *BRCA2* PVs. **Conclusions:** New evidence for the *BRCA2* c.658_659del (p.Val220Ilefs*4) variant indicates the cancer risks associated with this variant are lower than expected for a pathogenic *BRCA2* variant. Given this updated information, our laboratory has revised the classification of this variant to "special interpretation" to alert providers of the reduced penetrance associated with this variant so they can make patient specific decisions regarding screening and management. This example of a reduced penetrance truncating variant highlights the importance of both developing tools capable of showing penetrance differences as well as regularly reviewing variant classifications and their underlying evidence.

667T

A *RAD51C* mutation causing HBOC is expanded in the Newfoundland population likely due to a founder effect. S. Werdyani¹, L. Dawson^{1,2}, K. Smith¹, S. Hansford^{1,2}, R. Ndikumana³, T. Young¹, D. O'Rielly¹. 1) Discipline of Genetics, Faculty of Medicine, Memorial University, St. John's, NL, Canada; 2) Department of Gynecology and Obstetrics, Eastern Health, St. John's, NL, Canada; 3) Department of Pathology, Eastern Health, St. John's, NL, Canada.

Background: Extended families from the founder population of Newfoundland (NL) provide an excellent source for genomic investigation. The main objective of this study was to identify deleterious genetic variation(s) that lead to hereditary breast and ovarian cancer (HBOC) in *BRCA1/2*-negative cases from NL. **Methods:** Five women from 4 extended NL families were diagnosed with HBOC and enrolled in the study. Genomic DNA was extracted from whole blood and sequenced by an Illumina® NGS multigene panel covering HBOC susceptibility genes. Haplotype analysis was performed using 7 microsatellite markers flanking the *RAD51C* variant on chromosome 17q22. TaqMan genotyping was performed on 907 controls and 62 unrelated ovarian cancer cases from the NL population. RNA was extracted from peripheral blood of the proband and her unaffected mother, then it was reverse transcribed into cDNA and sequenced. Sanger sequencing of tumor tissue was also performed.

Results: Five women from 4 extended NL families that were *BRCA1/2*-negative carried the *RAD51C* c.571+4A>G mutation (rs587780257). This mutation was detected in 9 additional individuals including 2 women with HBOC, 1 with pancreatic cancer, 4 unaffected (one of whom had her ovaries removed at age 37); and 2 unaffected males. In contrast, the variant was absent in 15 unaffected family members. Haplotype analysis identified ~4 Mbp region that encompasses the c.571+4A>G mutation, and we find that this disease-associated haplotype is shared across all 4 families, suggesting a shared common ancestor. Of the 907 ethnically-matched controls, 3 were heterozygous for the mutation, estimating a MAF of 0.002 in NL, which is 100X higher compared to dbSNP (0.00002). Furthermore, the mutation was not detected in 62 unrelated ovarian cancer cases from this population. RT-PCR in blood revealed that c.571+4A>G causes aberrant splicing (skipping of exon 3, or exons 3 and 4) resulting in a truncated Rad51C protein in carriers. Sequencing of tumor tissue (from 2 female mutation carriers from 2 separate families) revealed loss of heterozygosity. **Conclusion:** Our study suggests that *RAD51C* c. 571+4A>G variant is likely pathogenic and associated with breast, ovarian, and possibly pancreatic cancer. This mutation is also amplified in the NL population, probably due to a founder effect. Further investigations including gene and protein expression profiling are ongoing.

668F

Multilocus inherited neoplasia alleles syndrome (MINAS): New cases of double heterozygosity for pathogenic cancer predisposition gene variants revealed by whole genome sequencing. J. Whitworth¹, P. Smith¹, J.E. Martin¹, R. Casey¹, K. Carss^{2,3}, J. Stephens^{2,3}, K. Stirrups^{2,3}, C. Penkett^{2,3}, R. Mapeta², S. Ashford^{2,3}, K. Megy^{2,3}, R. Armstrong^{1,2,4}, J. Barwell⁵, J. Cook⁶, A.B. Skytte⁷, M. Tischkowitz^{1,2}, E. Maher^{1,2}, NIHR-BioResource Rare Diseases Consortium. 1) Department of Medical Genetics, University of Cambridge and NIHR Cambridge Biomedical Research Centre and Cancer Research UK Cambridge Centre, Cambridge Biomedical Campus, Cambridge, CB2 0QQ, United Kingdom; 2) NIHR BioResource, Cambridge University Hospitals, Cambridge Biomedical Campus, Cambridge, CB2 0QQ, United Kingdom; 3) Department of Haematology, University of Cambridge, NHS Blood and Transplant Centre, Cambridge Biomedical Campus, Cambridge, CB2 0PT; 4) East Anglian Medical Genetics Service, Cambridge Biomedical Campus, Cambridge, CB2 0QQ, United Kingdom; 5) Department of Clinical Genetics, University Hospitals of Leicester, Leicester Royal Infirmary, Leicester, LE1 5WW, United Kingdom; 6) Department of Clinical Genetics, Sheffield Children's NHS Foundation Trust, Sheffield, S5 7AU, United Kingdom; 7) Department of Clinical Genetics, Aarhus University Hospital, Aarhus, DK-8200, Denmark.

Pathogenic variants in cancer predisposition genes (CPGs) such as *BRCA1*, *BRCA2*, *MSH2*, *MLH1*, *VHL*, *RET* etc. are characteristically associated with specific tumor phenotypes and germline genetic testing has previously focused on small groups of genes associated with those phenotypes. However, diagnostic investigation using large CPG panel assays and genome wide sequencing approaches such as WES and WGS is becoming increasingly prevalent. Consequences of this shift include the detection of pathogenic variants in individuals with an atypical tumor phenotype as well as identification of cases who harbor variants in two or more CPGs that would individually be considered to confer sufficient risk to prompt genetic counselling. This phenomenon of MINAS (Multiple Inherited Neoplasia Allele Syndrome; see Whitworth et al JAMA Oncology, 2016) can pose a number of clinical and diagnostic challenges (e.g. whether combinations of CPG variants will act synergistically and increased penetrance and/or atypical phenotypes). We have investigated the frequency of MINAS in a variety of patient cohorts. Among 440 individuals with multiple primary tumours without a genetic diagnosis, whole genome sequencing and variant assessment (based on ACMG guidelines) of 83 CPGs identified 3 patients with MINAS including combinations of variants affecting *FH1MAX*, *FLCN/CHEK2* and *PMS2/BMPRI1A*. In addition, a parent child dyad with wild type gastrointestinal stromal tumor both harbored pathogenic variants in *SDHA* and *PALB2*. We have compiled details of 122 cases of MINAS in the published literature and will present an assessment of its consequences for clinical management of the proband and their family. The phenotypic effects of particular aberrant gene combinations are more predictable if information regarding other occurrences of it are available. We have therefore established a database for sharing MINAS cases within the Leiden Open Variant Database environment (<https://databases.lovd.nl/shared/diseases/04296>).

669W

Exploring the range of renal lesions in DICER1 syndrome. *M. Apellaniz-Ruiz^{1,2}, G.M. Vujanac³, P.S. Thorner^{4,5}, J.R. Priest⁶, W.D. Foulkes^{1,2,7,8}*. 1) Department of Human Genetics, McGill University, Montréal, QC, Canada; 2) Lady Davis Institute, Segal Cancer Centre, Jewish General Hospital, Montréal, QC, Canada; 3) Department of Pathology, Sidra Medicine, Doha, Qatar; 4) Department of Pediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, ON, Canada; 5) Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada; 6) Minneapolis, Minnesota, USA; 7) Department of Medical Genetics, Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada; 8) Program in Cancer Genetics, Departments of Oncology, McGill University, Montreal, Quebec, Canada.

Introduction: DICER1 syndrome is a rare pediatric tumor predisposition syndrome caused by germline (typically truncating) mutations in *DICER1* and characterized by an array of benign and malignant conditions including pediatric cystic nephroma (pCN), anaplastic sarcoma of the kidney (ASK) and Wilms tumor. pCN is a benign multicystic tumor lacking blastemal and solid elements, differentiating it from cystic partially differentiated nephroblastoma (CPDN) and cystic Wilms tumor (cWT). Malignant progression of pCN to ASK has rarely been reported and is supported by the presence of *DICER1* mutations in both lesions. The involvement of *DICER1* mutations in cystic renal lesions other than CN is unclear. Thus, we studied the pathologic features and *DICER1* mutation status in a series of uncommon pediatric renal lesions. **Methods:** A series of 14 FFPE kidney tumor samples including CN (n=2), CPDN (n=6), cWT (n=2) and other lesions (n=4) was collected. Tumor and normal DNA was screened for *DICER1* variants using a Fluidigm array covering exons and exon-intron boundaries. Sanger sequencing was used to validate the variants. Immunohistochemical staining for WT1 was performed. **Results:** We identified 4 samples with *DICER1* mutations fitting the classic two-hit model: germline truncating and somatic hotspot RNase IIIb mutations. Two were pCN, as expected. The third case was a 5-year old male with a multicystic renal lesion. It was piecemeal biopsied, and despite the numerous cysts, the presence of solid areas composed of undifferentiated stroma lacking blastema ruled out pCN and CPDN, and favoured a stromal WT. The fourth case was a 9-month old female with a multiloculated cystic lesion. Several foci of small primitive cells with WT1 nuclear expression were present in the septa and were interpreted as blastema. Based on this, the lesion was categorized as CPDN. In both these latter cases, the first "hit" was in the germline. **Conclusions:** The diagnosis of a cystic kidney tumor should prompt consideration of testing for DICER1 syndrome, even if the lesions do not fit the classical description of DICER1-associated lesions. Both molecular genetics and specialist pathology review are essential to refine the diagnosis and to offer appropriate management to patients with uncommon lesions, and their relatives.

670T

Leukemia in a patient with 15q overgrowth syndrome. *E.E. Bodle¹, R. Gupta², A.M. Cherry², L. Muffly², M. Manning^{1,2}*. 1) Department of Pathology, Stanford University, Stanford, CA; 2) Division of Medical Genetics, Stanford University, Stanford, CA; 3) Division of Blood and Marrow Transplant, Stanford University, Stanford, CA.

Overgrowth syndromes are rare genetic conditions which present as global or segmental hyperplasia and are often associated with increased risk of malignancy. Trisomy of the terminal portion of 15q which includes the IGF1R gene, produces a rare overgrowth phenotype that has been termed 15q overgrowth syndrome (15q OGS). Upregulation of IGF1R has long been implicated in oncogenesis of multiple cancer types, including acute leukemias, and has been shown to render cells more susceptible to other transforming events. To date, too few cases of 15q OGS syndrome have been reported to identify any cancer predisposition. We present a case of a 34 year-old female with intellectual disability, macrocephaly, and subtle dysmorphic features who was diagnosed with mixed phenotype acute leukemia (lymphoid and myeloid). Prior to initiation of therapy she was referred to medical genetics for further evaluation and was identified as having a chromosomal translocation resulting in a partial trisomy of chromosome 15q, consistent with 15q OGS. A review of the literature for cases of malignancy in individuals with increased copy number of 15q revealed only one other reported patient. Given the small number of reported cases, we cannot rule out an increased risk of cancer associated with this chromosomal overgrowth syndrome. Although concerns have been raised regarding treatment feasibility in the setting of chromosomal disorders, the reported patient underwent successful treatment by hematopoietic stem-cell transplant.

671F

Assessing the modifying capacity of *CD36* in Lynch syndrome compared to that observed in familial adenomatous polyposis. *T. Connor*^{1,2,5}, *C. Oldmeadow*⁴, *B. Talseth-Palmer*⁴, *P. Pockney*^{4,5,6}, *R. Thom*⁴, *R. Scott*^{2,3}. 1) Junior Medical Officer Unit HNELHD, John Hunter Hospital, Newcastle, Australia; 2) Hunter Area Pathology Services, Lookout Road, New Lambton 2305; 3) Medical Genetics School or Biomedical Sciences and Pharmacy Newcastle; 4) Hunter Medical Research Institute Lookout Road, New Lambton; 5) University of Newcastle Newcastle; 6) Division of Surgery, John Hunter Hospital, Newcastle.

Genetic variants associated with Familial Adenomatous Polyposis (FAP) and Lynch Syndrome (LS) have long been studied to determine their role in the pathogenesis of these two diseases. From studies of FAP it became quickly recognised that a genotype phenotype correlation existed. For LS a more complex association has emerged that appears to be specific to which of four genes is implicated in the disease. Notwithstanding there appears in LS evidence that disease expression may be modified by either genetic and or environmental factors. The genotype/phenotype correlations observed in FAP are approximations and are not highly correlated. For example, the mutational hotspot at codon 1309 in *APC* is frequently reported as being associated with an extremely severe phenotype yet not all individuals who harbour the codon 1309 variant present with severe disease. For LS no genotype/phenotype correlations have been described, probably as a result of DNA mismatch repair being necessary for the maintenance of genomic stability. Nevertheless, patients who carry the same mutation in one of the four MMR repair genes can have vastly different ages of disease onset. An explanation for such variation in phenotype in both FAP and LS is the presence of modifier genes that can influence the age of disease diagnosis and the presenting phenotype. A number of modifier genes has been reported for LS but little has been forthcoming for FAP. Recently, we have shown that *CD36* has been implicated, in mouse models of FAP, to be a modifier of disease, as judged by an increase in tumor multiplicity in this mouse model. This association has recently been shown to influence the degree of symptomatic polyposis in a large cohort of FAP patients but did not address the association with colorectal cancer. In the current study we have examined the role of a number of single nucleotide polymorphisms (SNPs) in *CD36* in LS to determine whether the modifying effect of variants in this gene are directly associated with colorectal cancer development in a population of patients known to be at high risk of this disease. The identification of modifiers of disease risk in FAP and LS will have implications for surveillance and prophylactic measures that need to be taken to reduce the risk of these patients presenting with advanced disease.

672W

Synonymous but not silent: A synonymous *VHL* mutation confers susceptibility to pheochromocytomas in a four-generation family. *S.K. Flores*¹, *Z. Cheng*¹, *A. Jasper*¹, *R.W. Tothill*², *P.L.M. Dahia*¹. 1) Department of Medicine, Division of Hematology/Oncology, UT Health San Antonio, San Antonio, TX; 2) Peter McCallum Cancer Centre, Melbourne, Victoria, Australia.

Pathogenic mutations in the von Hippel-Lindau (*VHL*) gene predispose individuals to VHL disease comprising of renal cell carcinomas, pheochromocytomas (PCCs), hemangioblastomas of the central nervous system and other manifestations. Clinical presentation varies remarkably. In VHL disease type 2, PCC risk is higher, with type 2C manifesting PCC only. We report on a four-generation family with a history of PCCs in a pattern consistent with autosomal dominant inheritance. The proband developed a unilateral PCC at age 32. Several of her family members, including her brother, father, paternal aunt and cousin, have also developed unilateral PCCs. Whole exome sequencing of her germline DNA revealed a heterozygous, synonymous mutation (c.414A>G, p.P138P) in *VHL* exon 2. No other candidate genes were identified. Sanger sequencing showed that the mutation segregated with the PCC phenotype in the family. The variant was not observed in population databases (ExAC) whereas ClinVar had 3 entries with conflicting interpretations (2 uncertain significance, 1 likely pathogenic). Nanostring-based Pheo-Type profiling of the proband's archival PCC was consistent with a VHL subtype. In PCC from another relative, loss of the WT allele was observed at the variant locus. PCC cDNA analysis showed absence of the full length VHL transcript and presence of the shorter transcript lacking exon 2 in contrast to other PCCs which expressed both transcripts. Leukocyte cDNA analysis of carrier and WT relatives supported this finding. Another synonymous *VHL* mutation, located at the exon-intron border, has been reported to induce exon 2 skipping and thus alternative splicing and is associated with tumors. *VHL* encodes a tumor suppressor with ubiquitin ligase activity for degradation of hypoxia inducible factors (HIFs). *VHL* exon 2 is critical for the HIF binding domain; predominant expression of the shorter isoform leads to elevated HIF targets associated with oncogenesis. Most genetic screening workflows exclude synonymous variants. Our findings show that synonymous variants in coding regions of susceptibility genes should be taken into consideration as they may have alternative splicing consequences. Based on our findings, the c.414A>G variant is pathogenic and carriers should undergo routine follow-up for early detection. Although this family's clinical profile suggests VHL type 2C disease, further surveillance is recommended as the consequences of this variant are not yet fully defined.

673T

Accurate clinical interpretation of moderate risk breast cancer genes requires inclusion of the significant modifying effects of common genomic variants and family history. P.A. James^{1,3,4}, N. Li², S. McInerney¹, S. Rowley², D. Goode², R. Scott², L. Devereux², N. Grewal², I.G. Campbell^{2,3}. 1) Parkville Familial Cancer Centre, Peter MacCallum Cancer Centre, Melbourne, Victoria, Australia; 2) Cancer Genomics Laboratory, Peter MacCallum Cancer Centre, Melbourne, Victoria, Australia; 3) Sir Peter MacCallum Department of Oncology, Melbourne University, Melbourne, Victoria, Australia; 4) Genomic Medicine, Royal Melbourne Hospital, Parkville, Melbourne, Victoria, Australia; 5) Hunter Area Pathology Service, Newcastle, NSW, Australia; 6) LifePool Study, Peter MacCallum Cancer Centre, Melbourne, Victoria, Australia.

Rare variants in several genes are proposed to confer a moderate risk of breast cancer; typically an increased risk between 1.5 and 4 fold. Many of these genes appear on commercial testing panels but the utility of testing for these moderate effects in clinical practice remains unproven. In a large breast cancer case-control study we examined sequence data for 14 known or proposed moderate genes, along with the polygenic risk score and family history to determine whether incorporation of modifying effects could better identify women at actionable levels of breast cancer risk. Method: 5125 women affected by invasive breast cancer, (*BRCA1/BRCA2* excluded), and 4239 healthy controls were screened for variants in potential breast cancer risk genes, and a polygenic risk score (PRS) made up of 74 GWAS-reported SNPs using a custom targeted capture panel. Loss of function and pathogenic missense variants in established risk genes (*PALB2*, *CHEK2*, *ATM*) or proposed candidates (*BRIP1*, *BLM*, *MRE11A*, *NBN*, *FANCM*, *RAD51C*, *RAD51D*, *RAD50*, *BARD1*, *XRCC2*, *ATR*), were analysed together with an individual's PRS and breast cancer family history (FHx) by multiple logistic regression. A contrasting group of very high risk genes (*TP53*, *PTEN*, *STK11*, *CDH1*) were also examined. Results: Analysis included 517 rare pathogenic variants in 9364 women (6.6% cases, 3.9% controls). Polygenic risk and FHx showed the expected association with breast cancer: PRS, HR 1.52 (95% CI 1.46-1.58) per SD, Breast cancer FHx, OR_{adj} 1.42 (1.35-1.50). The risk conferred by pathogenic variants in 14 established and proposed moderate genes, but not in rare high risk genes, was significantly modified by the PRS, with no risk observed associated with a low PRS background (low quintile) OR=0.7 (0.5-1.2), and clinically actionable risk associated with the same variants and high PRS, (high quintile) OR=3.7 (2.5-5.6). Defining 'high risk' as a 30% lifetime breast cancer risk, 16% of carriers of pathogenic variants in candidate moderate risk genes would be high risk, 30% moderate risk, and for 54% their estimated risk was not significantly different to population levels. Conclusion: Important SNP-modifying effects affect many suggested moderate risk breast cancer genes, potentially contributing to the conflicting reports of risk for these genes. Combined interpretation of rare variants with PRS data and FHx can significantly improve the accurate identification of high risk individuals.

674F

A retrospective cohort study investigating age and gender specific presentations of Li-Fraumeni Syndrome. S.P. MacFarland¹, K. Zelle¹, J. Long², D. McKenna², J. Powers², B.W. Katona², S.M. Domchek², K.L. Nathanson², G.M. Brodeur¹, K.N. Maxwell². 1) Children's Hospital of Philadelphia, Philadelphia, PA., USA; 2) University of Pennsylvania, Philadelphia, PA, USA.

Introduction: Patients with Li-Fraumeni Syndrome (LFS) have an up to 90% lifetime cancer risk. Although previously considered a pediatric cancer predisposition syndrome, adults with LFS are increasingly ascertained via multi-gene panels. We sought to investigate age- or gender-specific genotype or phenotype correlations in LFS. **Methods:** Patients were identified by query of the electronic medical record from 2007-2017 at the Children's Hospital of Philadelphia and the University of Pennsylvania. Medical records review confirmed the presence of a pathogenic *TP53* mutation, family relationships, and personal and family cancer histories. **Results:** Ninety-eight individuals from 63 families with LFS were identified. Of the 63 families, 60% fit Chompret criteria and 8% had known de novo mutations. Of remaining families (n=20), 85% had LFS-like syndrome, breast cancer under age 40, multiple primary cancers or a pediatric cancer. Comparing 68 females to 30 males, 72% versus 57% had any cancer and 56% vs 67% had an LFS core cancer. Non-LFS core cancers were more common in males, 9% vs 27% (p=0.03). In order to study LFS diagnosed in children versus adults, 25 index cases of LFS under age 18 (pediatric probands) and 33 index cases of LFS over age 18 (adult probands) were identified. There were no significant genotype differences between pediatric and adult probands (DNA binding domain missense in 52% vs 65%, tetramerization domain missense in 10% vs 16% and truncating mutations in 38% vs 19% of pediatric vs adult probands, respectively). Pediatric and adult probands' family histories fit LFS criteria at similar rates (24% vs 30% for Classic and 62% vs 55% for Chompret). Nearly all probands presented with cancer including 96% of pediatric and 94% of adult probands with LFS core cancers found in 88% of pediatric probands and 79% of adult probands. Only non-LFS tumors were marginally more common in adult vs pediatric probands (36% vs 12%, p=0.06). **Conclusion:** Though sample size was limited and included only patients from tertiary care medical centers, this collaborative retrospective study is uniquely positioned to make adult versus pediatric comparisons. In this cohort, LFS-core cancers occurred at similar rates in adult and pediatric probands, and non-LFS cancers were more common in males and adult probands. A prospective longitudinal cohort study is ongoing to enable further genotype-phenotype studies in adult versus pediatric LFS.

675W

Challenges in clinical variant classification in *TP53*: A representative case series for *TP53* c.1000G>C; p.G334R. J. Powers¹, K. Zelle², B. Levin³, T. Slavin⁴, V. Joseph⁵, J. Weitzel⁶, K. Offit⁶, J. Stopfer⁶, J. Garber⁶, F. Couch⁷, K. Nathanson¹, S. Domchek¹, K. Maxwell¹. 1) University of Pennsylvania, Philadelphia, PA; 2) Children's Hospital of Philadelphia, Philadelphia, PA; 3) MD Anderson Cancer Center at Cooper, Camden, NJ; 4) City of Hope, Duarte, CA; 5) Memorial Sloan Kettering Cancer Center, New York, NY; 6) Dana Farber Cancer Institute, Boston, MA; 7) Mayo Clinic, Rochester, MN.

Introduction: Li-Fraumeni syndrome (LFS) is a cancer predisposition syndrome caused by germline mutations in the *TP53* gene. While *TP53* is well-characterized, challenges in variant classification remain due to the predominance of missense mutations, limited research in domains beyond the DNA-binding domain, and phenotypic heterogeneity. In this case series, we present data supporting pathogenicity of a rare *TP53* tetramerization domain missense variant with discordant clinical classifications, c.1000G>C; p.G334R. **Methods:** A retrospective, multi-institutional case review of individuals with *TP53* c.1000G>C; p.G334R was performed in conjunctions with literature and public data review. **Results:** In a research based sequencing study of 278 early onset breast cancer patients, an Ashkenazi Jewish proband with early onset breast cancer and sarcoma was found to carry the *TP53* c.1000G>C; p.G334R variant. Familial analysis indicated the variant segregated with component LFS cancers in third degree relatives. The proband's paternal first-cousin, with early onset breast cancer and glioblastoma, underwent clinical testing confirming c.1000G>C; p.G334R, which was classified as a variant of uncertain significance (VUS). Nine additional c.1000G>C; p.G334R probands were ascertained. Of the ten total cases, five met Chompret, and two met LFS-like criteria. The variant segregated with cancer in third and fourth degree relatives in two families. Of the eight carriers with ancestry data, all reported Ashkenazi Jewish descent; whereas no carriers were identified in 1154 Ashkenazi Jewish control chromosomes or in gnomAD. *In silico* prediction models were consistent with a deleterious effect. Literature review revealed cell-type specific reduction in transactivation capability. Of nine CLIA-CAP laboratories for which variant calls were obtained, five laboratories classify c.1000G>C; p.G334R as a VUS and four classify as likely pathogenic or pathogenic. Loss of heterozygosity and haplotype analyses are pending and will be reported in the near future. **Conclusion:** Our data suggest that *TP53* c.1000G>C; p.G334R is a rare Ashkenazi Jewish founder mutation leading to a Li-Fraumeni-like syndrome. This study demonstrates challenges in variant classification even in well-characterized genes such as *TP53*. Centralized collection and deep phenotyping of multiple families in combination with gene-specific expertise may aid in variant classification and clinical management.

676T

Beyond Lynch syndrome: Thinking about CMMRD. R. Quero^{1,2}, M. Manotas^{1,2}, A. Gómez². 1) Pontificia Universidad Javeriana, Bogotá D.C., Colombia; 2) Hospital Universitario San Ignacio.

Lynch syndrome (LS [MIM 609310]) is an adult-onset hereditary cancer syndrome characterized by gastrointestinal and genitourinary cancers. It is caused by monoallelic germline mutations in *MLH1*, *MSH2*, *MSH6*, *PMS2* (MMR genes) or *EPCAM*. In contrast, biallelic mutations in MMR genes result in a distinct disease defined as Constitutional Mismatch Repair Deficiency Syndrome (CMMRD). The proband was referred to Medical Genetics due to colorectal cancer at 39 years old, a daughter with medulloblastoma, who died at 3 years old, and maternal family history of colorectal and gastric cancers. As suspected, a heterozygous germline mutation in *MLH1* gene was found (c.790+1G>A), confirming LS. The patient and his wife reported that their deceased daughter was diagnosed and followed-up as neurofibromatosis type 1 (NF1), due to café-au-lait spots, and the medulloblastoma was considered related to it. At that moment, a complete pedigree of the proband's wife was made, finding a maternal aunt with colorectal cancer at 35 years old and a common ancestor with the proband. The patient's wife was found with the same mutation in *MLH1*. This couple is undergoing surveillance for LS and has a 5-year-old daughter who is awaiting assessment to define evaluation for this specific mutation. CMMRD is characterized by early-onset malignancies: hematological, central nervous system tumors, colorectal and other cancers, typically seen in LS patients at a later age, and features reminiscent of NF1. In this case, it is presumed that the deceased daughter had this syndrome and the evaluation of the alive daughter is expected to determine if she is a *MLH1* mutation biallelic carrier. It is important to suspect and diagnose this rare disease since there are proposals of surveillance protocols for early tumor detection and potential therapeutic interventions.

677F

Phenotypic variability in familial FH whole gene deletion. *C. Quindipan¹, S. Thomas², A. Martinez³, M. Fox³, S.C. Saitta^{1,4}.* 1) Center for Personalized Medicine, Children's Hospital Los Angeles, Los Angeles, CA; 2) Division of Hematology/Oncology, Children's Hospital Los Angeles, Keck School of Medicine, University of Southern California, Los Angeles, CA; 3) Invitae, San Francisco, CA; 4) Department of Pathology, Keck School of Medicine, University of Southern California, Los Angeles, CA.

FH encodes fumarate hydratase, a Krebs cycle enzyme that acts as a tumor suppressor. Germline pathogenic variants in *FH* are seen in autosomal dominant hereditary leiomyomatosis and renal cell cancer (HLRCC) and autosomal recessive fumarate hydratase deficiency. We report a 12-year-old female who presented with abdominal distension, respiratory distress, and altered mental status. A CT scan showed a left adrenal gland mass with pheochromocytoma confirmed by histopathology of the resected tumor. The patient has normal intellectual function and no concern for a metabolic disorder. A 3-generation family history was unremarkable for cancer but the patient's mother required a hysterectomy at age 47 for uterine fibroids and heavy menstrual bleeding. A 14-gene Paraganglioma-Pheochromocytoma Panel identified a pathogenic deletion of the entire coding sequence of the *FH* gene in the proband's peripheral blood. Chromosome microarray then showed a 1.34 Mb deletion of chromosome 1q43 including *FH*. None of the other 10 genes in the deleted interval are associated with human disease. All first-degree family members were tested and the deletion was found only in the proband and in her mother; the unaffected sibling was deletion negative. None of the maternal extended family members are available for testing. Only 1% of patients with pheochromocytoma have germline *FH* mutations. Uterine leiomyomata instead are present in almost every female with HLRCC, presenting with dysfunctional uterine bleeding, though rarely becoming malignant. HLRCC can include multiple cutaneous leiomyomas on the abdomen, shoulders, and extremities, seen in the 2nd-4th decades. Renal tumors are observed in 15% of individuals with HLRCC including type 2 papillary renal cell carcinoma, collecting-duct carcinomas, and Wilms tumor. Neither the patient nor her mother is currently affected with cutaneous or renal tumors. This is, to our knowledge, the first-known familial *FH*-related disorder in a pediatric patient and her parent, presenting with distinct phenotypes. Tumor surveillance for both individuals will now include screening for pheochromocytomas and renal cancer. In addition, abdominal ultrasounds will be performed on the proband every two years to screen for uterine fibroids. This case underscores the importance of family testing for germline cancer predisposition genes and its relevance to developing an appropriate care plan.

678W

Co-occurrence of neurofibromatosis Type 1 (NF1) and Type 2 (NF2): Common downstream effects on the RAS pathway with implications for tumor burden and potential treatment. *S. Krishnamurthi, J. Fisher, J. Mann.* Kaiser Permanente, Clovis, CA.

NF1 and NF2 are two highly penetrant autosomal dominant phakomatoses with an incidence of 1:3000 and 1:33,000 respectively. Somatic mosaicism is known feature of both disorders. The protein products of the associated genes, neurofibromin and merlin, are functionally unrelated but act as tumor suppressors through modulation of the RAS/RAC oncogenic pathways. Neurofibromin is a RAS GTPase activating protein (GAP) which maintains RAS in the inactive GDP form and its loss leads to abnormal activation of RAS and a downstream oncogenic kinase, PAK1, which is involved in several aspects of oncogenesis. Merlin also directly inactivates PAK1 and its dysfunction causes abnormal activation of PAK1. Thus, NF1 and NF2 are genetically distinct disorders with similar phenotypic outcomes resulting from a common effector site at the molecular level. This creates implications for tumor burden in cases of co-occurrence of these two conditions but also provides an opportunity for targeted therapy to combat both disorders simultaneously with use of PAK1 blockers. We report a case of a 54-year female clinically diagnosed at an early age with NF1 with numerous neurofibromas, café au lait macules and lisch nodules who later presented in her 50's with unilateral hearing loss. MRI revealed bilateral vestibular schwannomas meeting clinical criterion for NF2. She has no known family history of NF1 or NF2. Molecular testing for *NF1*, *NF2*, *LZTR1* and *SMARCB1* showed only a pathogenic splicing alteration in the *NF1* gene. Tumor testing has not yet been undertaken. Co-occurrence of NF1 with NF2 has been rarely reported before with a germline variant in one disorder and a somatic mosaicism at a segmental level in the other. About a third of simplex cases of NF2 are found to be mosaic. Two independent germline variants leading to both conditions together in the same individual has never been reported before to our knowledge. Regardless, for any individual meeting criteria both disorders ensuring optimal ongoing surveillance and management, and appropriate recurrence risk counseling is an essential consideration. PAK1 blockers may provide a unique treatment option in the future.

679T

A de novo BRCA2 mutation identified in a woman with breast cancer and strong maternal family history. R. Drouin^{1,2}, W. Pirjamali³, L. Provencher^{4,5}, M. Letourneau^{1,5}, C. Bouffard⁶, S. Strom², H. Gao². 1) Laval University, Faculty of Medicine, Quebec City, Quebec, Canada; 2) CHUL-CHUQ Université Laval, CHUL, Division of Medical Genetics, Dept of Pediatrics, Quebec City, Quebec, Canada; 3) Fulgent Genetics, Temple City, CA, USA; 4) CHUL-CHUQ Université Laval, Centre des Maladies du Sein Deschênes-Fabia, Hôpital du Saint-Sacrement, Dep of Surgery, Quebec City, Quebec, Canada; 5) CHUL-CHUQ Université Laval, Hôpital Hôtel-Dieu de Québec, Division of Radio-Oncology, Dept of Specialized Medicine, Quebec City, Quebec, Canada; 6) Laboartoire de recherche transdisciplinaire en génétique, médecines et sciences sociales, Division of Genetics, Dept of Pediatrics, FMSS, Université de Sherbrooke, Sherbrooke, Quebec, Canada.

A 45-year-old woman was referred to Medical Genetics for investigation of her breast cancer, an infiltrating ductal carcinoma diagnosed at the age of 43. Maternal family history revealed seven cases of breast cancer. The tumor was positive for estrogen and progesterone receptors and negative for amplification of HER2/NEU. The patient had a recurrence risk score of 24 (intermediate risk, Genomic Health / Oncotype DX) and was tested for the following genes at Fulgent Genetics: *APC*, *ATM*, *AXIN2*, *BARD1*, *BMPR1A*, *BRCA1*, *BRCA2*, *BRIP1*, *CDH1*, *CHEK2*, *EPCAM*, *GALNT12*, *GREM1*, *HOXB13*, *MLH1*, *MLH3*, *MRE11A*, *MSH2*, *MSH6*, *MUTYH*, *NBN*, *NF1*, *PALB2*, *PMS2*, *POLD1*, *POLE*, *PTEN*, *RAD50*, *RAD51C*, *RAD51D*, *SMAD4*, *SMARCA4*, *STK11*, *TP53* and *XRCC2*. **Design/Method.** Retrospective case review of a patient with a diagnosis of breast cancer tested with a cancer gene panel at Fulgent Genetics. Genomic DNA from the submitted sample is barcoded and enriched for the coding exons of targeted genes using hybrid capture technology. Prepared DNA libraries are sequenced using a Next Generation Sequencing (NGS) technology. Sequencing results are aligned to the reference sequence and variants are detected in regions with sufficient coverage. In general, 100% of coding regions and splicing junctions are covered to at least 50x by NGS or by Sanger sequencing. Locus specific databases, literature searches, and other molecular biological principles are used to classify variants. To minimize false positive results, any variants that do not meet internal quality standards are confirmed by Sanger sequencing. All genes listed were evaluated for large deletions and/or duplications. Deletions or duplications identified by NGS are confirmed by an orthogonal method (qPCR or MLPA). **Results.** A pathogenic truncating mutation was detected in *BRCA2* (c.9435_9436del p.Ser3147Cysfs*2). Parental testing was negative, indicating a de novo mutation. The parent identity testing will be done to confirm that the parents are the parents and that there was no sample swap. **Conclusions.** These results confirm that whenever it is possible, parents should be tested to determine whether a mutation has been inherited or occurred de novo to better tailor the genetic counseling.

680F

Analysis of actionable adult-onset disease risk findings in newborn genomic sequencing. O. Ceyhan-Birsoy^{1,2}, S. Fayer², J.M. Murry^{2,4,5}, K. Machin^{2,4,5}, C.A. Genetti⁶, T. Schwartz², M.S. Lebo^{2,4,5,7}, T.W. Yu^{5,6,8}, P.B. Agrawal^{5,6,9}, R.B. Paradi^{8,10}, R.C. Green^{3,5,7}, H.L. Rehm^{2,4,5,7,11}, I.A. Holm^{5,6}, A.H. Beggs^{5,6}. 1) Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY; 2) Laboratory for Molecular Medicine, Partners HealthCare Personalized Medicine, Cambridge, MA; 3) Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Boston, MA; 4) Department of Pathology, Brigham and Women's Hospital, Boston, MA; 5) Harvard Medical School, Boston, MA; 6) Division of Genetics and Genomics, The Manton Center for Orphan Disease Research, Boston Children's Hospital, Boston, MA; 7) The Broad Institute of MIT and Harvard, Cambridge, MA; 8) Department of Neurology, Boston Children's Hospital, Boston, MA; 9) Division of Newborn Medicine, Boston Children's Hospital, Boston, MA; 10) Department of Pediatric Newborn Medicine, Brigham and Women's Hospital, Boston, MA; 11) Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA.

Currently, there is ongoing debate about the risks and benefits of returning adult-onset disease risk findings from genomic sequencing (GS) to children. The BabySeq Project is a randomized controlled trial that explores the use of GS in the clinical care of newborns. Healthy newborns from the well-baby nursery and ill newborns from intensive care units (ICUs) were enrolled and half were randomized to receive GS in addition to standard of care. All families in the GS arm received a report that returned risk and carrier status for childhood-onset diseases as well as pediatrics-related pharmacogenomics findings. In addition, 114 families were consented to receive actionable adult-onset disease risk information for their newborns. To identify genes associated with diseases known to present only during adulthood and for which early knowledge could improve health outcomes, 113 genes that were possibly associated with adult-onset cardiovascular, neurological and endocrine disorders, and hereditary cancers, including the 59 genes recommended by ACMG to be analyzed for secondary GS findings, were curated for validity of disease association, earliest reported age of onset, penetrance, and actionability. Genes that may present during childhood, including 53 ACMG recommended genes, were excluded, as they were assessed during the initial childhood-onset disease risk analysis. Seven genes were determined to present during adulthood only, have at least moderate penetrance, and be medically actionable. *PMS2* and *EPCAM* were excluded from further analysis, as the majority of their pathogenic variation was not expected to be detected by exome sequencing. GS data were filtered and interpreted to identify pathogenic and likely pathogenic variants in *BRCA1*, *BRCA2*, *MLH1*, *MSH2*, and *MSH6* genes. 3/114 (2.6%) newborns had actionable adult-onset disease risk findings. Two newborns that were healthy at the time of enrollment had pathogenic variants in *BRCA2* and *MSH2*, associated with hereditary breast and ovarian cancer and Lynch syndrome, respectively. One newborn with a congenital heart disease enrolled from the ICU had a pathogenic *BRCA2* variant. Testing of parental samples identified parents who also harbored these variants and therefore risk for disease. Results were returned and families were counseled and referred to oncology specialists as appropriate. We continue to explore the medical, behavioral and economic impacts of our GS results on the newborns and their families.

681W

Germline *MSH2* c.-82G>C promoter variant found in Newfoundland

patients with mismatch repair deficiency. A.L. Jacobson¹, A.S. McFaddin¹, M. Aronson², A. MacMillan³, P. Matthews⁴, G. Boughton¹, B.A. Fernandez⁵, R.H. Kim⁶, U. Tabori⁷, B.H. Shirts¹, E.Q. Konnick¹, C.C. Pritchard¹. 1) Department of Laboratory Medicine, University of Washington, Seattle, WA; 2) Zane Cohen Centre for Digestive Diseases, Mount Sinai Hospital, Toronto, ON, Canada; 3) Provincial Medical Genetics Program, Eastern Health, St. John's, NL, Canada; 4) James Paton Memorial Regional Health Centre, Gander, NL, Canada; 5) Discipline of Genetics, Memorial University of Newfoundland, St. John's, NL, Canada; 6) Mount Sinai Hospital, Department of Medicine, University of Toronto, Toronto, ON, Canada; 7) Division of Haematology Oncology, The Hospital for Sick Children, Toronto, ON, Canada.

Introduction: Pathogenic variants in mismatch repair (MMR) genes cause Lynch syndrome and constitutional mismatch repair deficiency (CMMRD). Very few pathogenic promoter variants have been identified in MMR genes. We identified a rare homozygous germline *MSH2* promoter variant (c.-82G>C) in a child suspected to have CMMRD. The patient had an ultra-mutated brain tumor with *MSH2/MSH6* loss by immunohistochemistry (IHC) that was unexplained by any other variant, prompting a comprehensive review of the variant. This variant is not reported in gnomAD and is reported once in ClinVar as a variant of uncertain significance (ID 231091). **Methods:** We identified carriers of the *MSH2* c.-82G>C germline variant among a series of 10,466 patients ascertained for clinical cancer predisposition testing using a comprehensive NGS cancer panel which includes promoter sequences of the MMR genes. Both germline and paired MMR deficient tumor samples were tested in a subset of cases. **Results:** In our clinical series, we identified four additional patients heterozygous for the *MSH2* c.-82G>C germline variant (Table 1). All patients originated from Newfoundland and were not known to be related. Three of 4 patients had Lynch syndrome-related cancers (2 colorectal, 1 endometrial). We detected IHC loss of *MSH2/MSH6* in all 3 tumors. We also detected *MSH2* promoter methylation in one tumor in which sufficient material was available for this testing (1 of 1 methylated); additional tumor material is being requested for the other *MSH2* c.-82G>C germline carriers to test for *MSH2* methylation. One heterozygous carrier did not have a personal history of cancer and was tested for family history of cancer. **Conclusion:** Our results support that the *MSH2* c.-82G>C variant is likely pathogenic and a likely cause of Lynch syndrome and CMMRD in Newfoundland patients. We hypothesize that this promoter variant may predispose to *MSH2* gene silencing via methylation. Additional work, including segregation studies and additional *MSH2* methylation testing, will help to further characterize this variant.

Table 1					
Case	Clinical history	MMR IHC	Germline MMR	Zygosity	Somatic MMR
1	Brain tumor, sibling with brain tumor	<i>MSH2/6</i> loss	<i>MSH2</i> c.-82G>C	hom	None
2	Endometrial cancer	<i>MSH2/6</i> loss	<i>MSH2</i> c.-82G>C	het	<i>MSH2</i> methylation
3	Colorectal cancer, Amsterdam II family	<i>MSH2/6</i> loss	<i>MSH2</i> c.-82G>C	het	<i>MSH2</i> c.2459-12A>G <i>MSH6</i> c.3261dup
4	Colorectal cancer	<i>MSH2/6</i> loss	<i>MSH2</i> c.-82G>C	het	<i>MSH2</i> c.1147C>T (stop)
5	No personal history of cancer	N/A	<i>MSH2</i> c.-82G>C	het	N/A

682T

VHL Information Sharing International Consortium (VISION): A ClinGen expert panel to evaluate VHL gene-specific criteria for variant interpretation. D.I. Ritter¹, C. Badduke², M. Anderson³, K. Clark⁴, R.H. Giles⁵, M. Griffith⁶, O. Griffith⁷, C. Horton⁸, K. Hruska⁹, H.C. Kang¹⁰, C. Kesserwan¹¹, K. Krysiak¹², L. Kujan¹³, M. Luo¹⁴, J. Machado¹⁵, E. Maher¹⁶, K. McGoldrick¹⁷, R. Patel¹⁸, T. Peseran¹⁹, T. Stockley²⁰, A. Tirosh²¹, Y. Wang²², S.E. Plon²³, R.H. Kim²⁴, ClinGen VISION VHL Expert Panel. 1) Baylor College of Medicine, Houston, TX, USA; 2) University Health Network, Princess Margaret Cancer Center, Canada; 3) Invitae, San Francisco, CA, USA; 4) McDonnell Genome Institute, Washington University, St. Louis, MO, USA; 5) University Medical Center, Utrecht, NL; 6) Ambry Genetics, Aliso Viejo, CA; 7) GeneDx, Gaithersburg, MD, USA; 8) St. Jude's Children's Research Hospital, Memphis, TN, USA; 9) Children's Hospital of Philadelphia, Philadelphia PA, USA; 10) Prevention Genetics, Marshfield WI, USA; 11) University of Cambridge, Cambridge, UK; 12) National Institutes of Health, Bethesda, MD, USA.

Von Hippel Lindau (VHL) disease is a hereditary tumor predisposition syndrome with multiple subtypes and has a prevalence of ~1/36,000. To date, no concerted effort has been established to share VHL clinical or variant information or develop VHL gene specifications. The Clinical Genome Resource (ClinGen) has convened an Expert Panel (EP) to assess VHL gene-specific criteria for variant interpretation using the American College of Medical Genetics and Genomics (ACMG) guidelines. The VHL Information Sharing International Consortium (VISION) is composed of 27 members across multiple clinical and research disciplines. Our first aim is to review the ACMG evidence codes to determine the needed specifications for VHL, informed by the ClinGen Sequence Variant Interpretation (SVI) guidance. To accomplish this, VISION has divided into three sub-teams as follows: Computational, Functional, Splicing (Team V), Segregation, Allelic, de novo (Team H) and Population-related criteria (Team L). Thus far, we have completed Step 1 of the ClinGen EP Application and are in the process of Step 2 (developing VHL gene specific criteria specifications). We have focused on preliminary specifications for 6 codes (PVS1, PM1, PP1, PP3, BA1 and BS1) and intend to complete all codes that need assessment by Spring 2019. These rules will be piloted on a test set of at least 75 VHL variants. A second aim of VISION is to encourage sharing of VHL clinical and variant information through contribution of VHL variants to ClinVar and consolidation of publicly available variants. To this end, members of VISION collaborated with the crowd-sourced database CIViC (Clinical Interpretations of Variants in Cancer). CIViC possesses the unique capacity of associating deep clinical information with precise genomic information, critical components of VHL variant interpretation. Over 450 VHL variants representing over 1300 cases with clinical phenotyping using HPO terms were deposited into CIViC to date and will provide a dataset for EP use. We have imported this, and additional CIViC VHL curations, into a variant-based pre-curation format that can be reviewed prior to submission into the ClinGen Variant Curation Interface (VCI), where it can be incorporated into the EP variant curation. Collectively, the VISION consortium will work towards facilitating a culture of VHL information sharing, and develop gene specific rules to help resolve challenging VHL variants in an ongoing manner.

683F

Identifying genetic polymorphisms conferring inverse risk in Alzheimer's and cancer using multivariable association analysis. G. Pathak, N. Phillips. Department of Microbiology, Immunology & Genetics UNT Health Science Center, Fort Worth, TX.

Worldwide, age demographics are shifting due to the significant expansion of the elderly population. Not surprisingly, this has resulted in an increased prevalence of aging-associated diseases, including Alzheimer's and cancer. Large population studies have reported that individuals with Alzheimer's have a lower risk of cancer history, and individuals with a history of cancer have a lower-than-expected prevalence of Alzheimer's. We hypothesize that the inverse relationship between Alzheimer's and cancer is attributed to genetic variants conferring opposite effects in cancer and Alzheimer's. Therefore, to identify these SNP markers for cancer subtypes and Alzheimer's, a frequentist's multinomial regression (MR) was applied to compare the three disease phenotypes- Alzheimer's, breast and prostate cancer- against controls. This approach increases the power to detect relative risk to Alzheimer's over cancer. QC-filtered genotype data were pooled for of all three disease-affected individuals and conducted PCA analysis to remove population stratification outliers. MR tests were conducted using Trinculo in a total of 4824 individuals with all individuals over the age of 50 and adjusting for age, sex, and first two principal components are covariates. A suggestive threshold of combined $p < 10^{-4}$ using likelihood ratio test for association of variants across all categories was used to identify significant SNPs. Genetic variants were filtered based on odds-ratio, variants with odds ratio less than 1 in Alzheimer's and higher than 1 for both cancers, and variants conferring high risk in cancer ($OR > 1$) and low risk towards cancer ($OR < 1$). The following genic SNPs exhibited significantly different allelic distributions when comparing all four phenotypes and conferred opposite risk effects when comparing odds ratios for Alzheimer's to odds ratios for cancer (both breast and prostate): rs207906 (XRCC5), rs11893547 (SMYD5), rs2075650 (TOMM40), rs405509 (APOE), rs8106922 (TOMM40), rs224519 (FOXP2). The next for this approach is to stratify the population by sex and identify gender-specific SNP associations for Alzheimer's and cancer. Exploring genetic variant-dependent bidirectional effects in cancer and Alzheimer's holds the potential to identify possible off-target effects of underlying pathology providing a protective effect of one disease over the other.

684W

Systematic study of chromosomal instability across human cancers. B.N. Lasseigne, S.J. Cooper, G.M. Cooper, R.M. Myers. HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA.

Genomic instabilities (GIN) like chromosomal instability (CIN, altered chromosome number and structure) are molecular signatures linked to gross genomic alterations, are an enabling characteristic of the hallmarks of cancer, represent distinct etiologies, and are associated with drug efficacy and patient prognosis. However, studies of copy number burden use different methods like number of breakpoints, number of genes with copy number variation (CNV), percent of bases with CNV, and functional aneuploidy scores. Because there is not consensus on which are most biologically and clinically relevant, our goal was to evaluate CIN metrics across human cancers. Using The Cancer Genome Atlas (TCGA, $n=33$ cancers) with other public data as validation, we demonstrate CIN metrics are not equivalent and occur at different frequencies in cancers. For example, in clear cell kidney carcinoma (KIRC) the number of breakpoints and number of genes with CNV are correlated ($p=0.20$), but represent distinct biological properties like homologous repair defects and selective pressure for CNVs with driver genes, respectively. CIN metrics also had varying associations with clinical characteristics (e.g., number of breakpoints or number of genes with CNV and reduced patient survival in KIRC; Cox $p=3.62 \times 10^{-7}$ and 0.45, respectively). We also identified molecular features consistently associated with each CIN metric (e.g., 8,831 CpGs associated with number of genes with CNV in KIRC; linear regression with patient age as a covariate, BH-adjusted $p < 0.05$). GO-term, transcription factor occupancy, and proximity to tissue-specific eQTL enrichments for each metric identified pathways and feature sets suggesting mechanisms of disease etiology and progression and potential therapeutic strategies (e.g., targeting purinoceptors in KIRC tumors with a low number of genes with CNV). Additionally, we hypothesized because CIN is linked to widespread DNA hypomethylation and characteristic gene expression changes, it can be inferred from these and other data types. We predicted CIN from each TCGA data type, consistently building models with favorable performance characteristics (e.g., LASSO regression of KIRC high/low number of breakpoints predicted from gene expression, area under curve > 0.91). Our study improves CIN signature interpretation, models CIN in other data types, and provides insight into the role of CIN in human cancer etiology, progression, and potential treatment strategies. .

685T

Biological networks modulating chemotherapy response in ovarian cancer.

D.G. Topouza¹, J. Choi¹, S. Nesdoly², Q.L. Duan^{1,2}. 1) Department of Biomedical and Molecular Sciences, Queen's University, Kingston, Ontario, Canada; 2) Queen's School of Computing, Queen's University, Kingston, Ontario, Canada.

Ovarian cancer has the lowest survival rate of all common gynaecological cancers. The gold standard of care for ovarian cancer is surgery and platinum-based chemotherapy; however, more than a third of patients are not responsive to this treatment. The underlying mechanisms of platinum therapy resistance remain poorly understood. In this study, we analyzed RNA sequence data from 190 ovarian serous cystadenocarcinoma patients (The Cancer Genome Atlas) to detect gene expression networks and genomic variants associated with platinum chemotherapy response. Responders (N=111) were cancer-free for one year after the completion of chemotherapy, whereas non-responders (N=79) had a recurrence of cancer less than six months following chemotherapy completion. The aims of this project were to: (i) identify differentially expressed genes using univariate analysis; (ii) identify and characterize gene co-expression networks using multivariate analysis; and (iii) associate genomic variants with platinum chemotherapy response. Preliminary analysis identified 56 differentially expressed genes between platinum responders and non-responders. These included genes contributing to known ovarian cancer regulatory pathways such as the *PI3K-AKT-mTOR* and integrin signaling pathways, as well as novel genes contributing to immune cell regulation pathways. Moreover, we identified a co-expression network associated with chemotherapy response, which was enriched for genes in known cancer development pathways including serine/threonine kinase activity and cyclin D regulation. In addition, this network of genes was enriched for cellular responses to oxidative stress, which is known to be involved in chemotherapy resistance. On-going analysis includes association of tumor and germline variants with chemotherapy response. Our study of the genes and networks associated with chemotherapy response will implicate the underlying mechanisms of platinum therapy resistance. We hope that this will ultimately facilitate genetic testing, promote novel drug development, and improve the efficacy of chemotherapy.

686F

Genetic risk factors for high-risk human papillomavirus (HPV) infection and persistence among women of African ancestry.

C.A. Adebamowo^{1,2,3,4}, S.N. Adebamowo^{1,2}, A.A. Adeyemo⁵ for the H3Africa ACCME Research Group. 1) Department of Epidemiology and Public Health, University of Maryland School of Medicine, Baltimore, MD, USA; 2) University of Maryland Greenbaum Comprehensive Cancer Center, University of Maryland School of Medicine, Baltimore, MD, USA; 3) Institute of Human Virology, University of Maryland School of Medicine, Baltimore, MD; 4) Institute of Human Virology Nigeria, Abuja, Nigeria; 5) Center for Research on Genomics and Global Health, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, USA.

Background: Genetic factors may influence the susceptibility to high-risk human papillomavirus (hrHPV) infection and persistence. We conducted a genome-wide association study (GWAS) to identify variants associated with hrHPV infection and persistence among women of African ancestry. **Methods:** Participants were 522 Nigerian women evaluated at baseline and 6 months follow-up visits for HPV. HPV was characterized using SPF₂₅/LiPA₁₀. hrHPV infection was positive if at least one carcinogenic HPV genotype was detected in a sample provided at the baseline visit and persistent if at least one carcinogenic HPV genotype was detected in each of the samples provided at the baseline and follow-up visits. Genotyping was done using the Illumina Multi-Ethnic Genotyping Array (MEGA) and imputation was done using the African Genome Resources Haplotype Reference Panel. Association analysis was done under additive genetic models adjusted for age, HIV status and the first 3 principal components of the genotypes. **Results:** The mean ($\pm$ SD) age of the study participants was 38 ($\pm$ 8) years, 46% were HIV negative, 24% were hrHPV positive and 10% had persistent hrHPV infections. The GWAS yielded 71 and 107 variants associated with hrHPV infection and persistence, respectively, with suggestive genome-wide significance [p value $< 1.0 \times 10^{-3}$]. The top variants (rs149473200, rs147344426, rs151071053 and rs572823632) associated with hrHPV infections are in or near *KLF12* [all $p = 1.5 \times 10^{-4}$]. *KLF12* is an important regulator of gene expression during carcinogenesis, and it is highly expressed by human immune cells. In several studies, it was associated with HPV and cervical cancer. The top variants associated with persistent hrHPV infections were 3 variants upstream of *AL450244* (rs143668247, rs12740341, rs2502139), a cluster of 5 intronic variants in *JPH2* (rs116834259, rs74358070, rs11452236, rs79032354, rs6130527) and 2 intronic variants in *GPR39* (rs16832308, rs62167448) [all $p < 7.0 \times 10^{-7}$]. These variants are in enhancer histone marks in several tissues and alter the motifs *Ets*, *GATA*, *NR4A*, *RXRA*, which have been associated with several cancers of the female reproductive system and leiomyosarcomas. **Conclusions:** This exploratory GWAS yielded risk loci that provide clues to the pathogenesis of hrHPV infection and persistence in women of African ancestry. Given the limited sample, larger discovery and replication studies are under way to further characterize the reported associations.

687W

Classical HLA alleles are associated with persistent high-risk human papillomavirus (HPV) infection in African women. *S.N. Adebamowo^{1,2}, A.A. Adeyemo³, C.A. Adebamowo^{1,2,4,5} for the H3Africa ACCME Research Group.* 1) Department of Epidemiology and Public Health, University of Maryland School of Medicine, Baltimore, MD, USA; 2) University of Maryland Greenebaum Comprehensive Cancer Center, University of Maryland School of Medicine, Baltimore, MD, USA; 3) Center for Research on Genomics and Global Health, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA; 4) Institute of Human Virology, University of Maryland School of Medicine, Baltimore, MD, USA; 5) Institute of Human Virology Nigeria, Abuja, Nigeria.

Background: Persistent high-risk human papillomavirus (hrHPV) infection is a necessary cause of cervical cancer. However, the genetic factors underlying its risk are not well understood. We hypothesized that immunogenetic variation plays a role in HPV infection and persistence. Therefore, we conducted a study of classical HLA alleles and their association with hrHPV persistence.

Methods: We characterized HPV infection using SPF₂₈/LiPA₁₀ in 522 Nigerian women at baseline and at 6 months follow-up visits in 2014. hrHPV infection was positive if at least one carcinogenic HPV genotype was detected in a sample provided at the baseline visit and persistent if at least one carcinogenic HPV genotype was detected in each of the samples provided at the baseline and follow-up visits. Genotyping was performed with the Illumina Multi-Ethnic Genotyping Array. Classical HLA alleles were imputed from genotypes in the MHC region using the HLA genotype imputation with attribute bagging (HIBAG) algorithm. HLA association tests were conducted under additive genetic models adjusted for age and HIV status. *P*-values were adjusted for multiple comparisons with the false discovery rate (FDR) method. **Results:** The mean ($\pm$ SD) age of the study participants was 38 ($\pm$ 8) years, 46% (239/522) were HIV negative, 24% (125/522) were hrHPV positive at baseline and 10% (51/522) had persistent hrHPV infections. HLA-C*04:01 was the most prevalent MHC I allelic variant (27% of hrHPV negative, 31% of hrHPV positive and 32% of persistent hrHPV positive women), while HLA-DQA1*01:02 was the most prevalent MHC II allelic variant (40% of hrHPV negative, 38% of hrHPV positive and 32% of persistent hrHPV positive women). HLA-C*17*01 was significantly, positively associated with persistent hrHPV infections (OR 2.94; 95% CI: 1.34 – 6.43; adjusted *P*-value 0.04). HLA-DQA1*02:01 showed borderline positive association with persistent hrHPV infections (OR 2.45; 95% CI: 1.21 – 4.96; adjusted *P*-value 0.07). HLA-DQA1*02:01 has been shown to be associated with HPV seropositivity in other populations. **Conclusions:** This study, the first to investigate HLA alleles and persistent hrHPV risk in African women, has identified important risk alleles that merit further study. Our findings provide new insights into the risk factors for hrHPV infection in women of African ancestry.

688T

Functional network-based mapping identifies disease genes in polygenic loci unresolved by fine-mapping of cutaneous melanoma GWAS. *M. Arto-mov^{1,2}, A. Loboda^{2,3}, H. Tsao^{1,2}, M.N. Artyomov⁴, M.J. Daly^{1,2}.* 1) Massachusetts General Hospital, Boston, MA; 2) Broad Institute, Cambridge, MA; 3) ITMO University, St.Petersburg, Russia; 4) Washington University, St. Louis, MO.

Background: GWAS in combination with fine-mapping is the major methodology for studying complex traits. Having made great success in identification of associated loci, GWAS still has limited power to detect associated genes. Numerous associations observed in GWAS imply a limited number of causal functional pathways. Hence, functional interactions network (e.g. Protein-Protein Interactions, Co-expression) could serve a valuable source of information for resolution of polygenic loci. We illustrate this approach with mapping of cutaneous melanoma (CM) GWAS results that identifies novel candidate genes. **Methods:** CM GWAS results from 15,990 CM cases and 26,409 controls (PMID:26237428) were used for construction of the candidate genes list. InWebIM protein-protein interaction network reference (PMID:27892958) was used for assessing functional connectivity. For gene mapping validation we used exome-sequencing dataset of 273 CM cases matched to 7,629 controls (PMID:29522175). **Results:** By fitting beta-uniform distribution to observed *p*-values distribution we created a score for each gene, based on signal (beta-distributed) to noise (uniform) ratio. Genes found within the same locus received equal scores. Using modified prize-collecting Steiner tree algorithm we identified functional module containing genes with the greatest proportion of association signal. We hypothesize that non-relevant for melanoma gene removal from reference network and repeated search for functional module should result in module's cumulative score decrease to be the score of removed gene. By filtering out genes with score drop much smaller than expected and keeping genes with score drop below expectation we selected one candidate gene from each polygenic locus - *ARNT*, *RAD23B*, *CCND1*. With selecting right genes we expect to improve association signal observed in the list of all GWAS genes. We estimated total burden of rare (MAF<0.01%) protein-truncating variants in all GWAS genes using exome sequencing dataset (*p*=0.045; OR=2.12) and compared to the same gene list lacking likely non-relevant genes (*p*=0.016; OR=2.45), gene list lacking likely relevant genes (*p*=0.33; OR=1.42) expectedly contains noise variation. **Conclusions:** Validation shows that predicted non-relevant genes carry small to none signal in rare variation, while relevant genes improve cumulative association signal. These results implicate novel melanoma risk genes and offer usage of this methodology for other complex traits.

689F

Genetic association between *IFITM* gene polymorphisms and cervical cancer susceptibility. T. Chang¹, Y. Yang^{1,2}, Y. Lee^{1,2}, T. Chen¹, W. Lin¹, C. Lin¹. 1) Mackay Memorial Hospital, New Taipei, Taiwan; 2) Taipei Medical University, Taipei, Taiwan.

A small percentage of women with cervical HPV infection progress to cervical cancer, indicating that host genetic makeup could play an important role. Interferon induced transmembrane (IFITM) proteins are highly inducible by interferon and may be crucial factors in carcinogenesis and progression of multiple cancers. This study aims to investigate if variants of the *IFITM* genes are associated with cervical cancer in Taiwanese women. The *IFITM1* rs77537847 A/G, *IFITM2* rs1059091 G/A and rs2031090 G/A, and *IFITM3* rs3888188 C/A polymorphisms were genotyped in 300 cervical squamous cell carcinoma (CSCC) patients and 300 age/sex matched healthy controls. The presence and genotypes of HPV in CSCC patients were also determined. We found no significant association between any polymorphisms or haplotypes examined and CSCC risk. In addition, no significant association was observed between HPV-16 positive CSCC patients and controls. Our findings provide no support for the hypothesis that specific polymorphisms of *IFITM* genes are associated with CSCC susceptibility in Taiwanese women.

690W

Genetic ancestry and the risk of cutaneous squamous cell carcinoma.

H. Choquet¹, J. Yin¹, T.J. Hoffmann^{2,3}, Y. Banda², M.N. Kvale², C. Schaefer¹, N. Risch^{1,2,3}, M. Asgari^{4,5}, E. Jorgenson¹. 1) Kaiser Permanente Northern California Division of Research, Oakland, CA, USA; 2) Institute for Human Genetics, University of California San Francisco, San Francisco (UCSF), CA, USA; 3) Department of Epidemiology and Biostatistics, UCSF, CA, USA; 4) Department of Dermatology, Massachusetts General Hospital, Boston, MA, USA; 5) Department of Population Medicine, Harvard Pilgrim Health Care Institute, Harvard Medical School, Boston, MA, USA.

Purpose: Cutaneous squamous cell carcinoma (cSCC) is one of the most common malignancies in non-Hispanic white individuals. The risk of cSCC is particularly elevated in individuals of northern and western European origin, likely due to fair pigmentation phenotypes (e.g. blue eye, pale skin). Here, we investigate the association between genetic ancestry and cSCC risk in individuals of European origin, and evaluate whether pigmentation phenotypes and known cSCC-associated loci attenuate the ancestry associations.

Methods: To examine the contribution of genetic ancestry in cSCC risk, we conducted a genetic ancestry analysis in 7,701 cSCC cases and 60,166 controls of non-Hispanic white race/ethnicity from the Kaiser Permanente Northern California Genetic Epidemiology Research in Adult Health and Aging cohort. cSCC cases were identified based on electronic pathology records and systematized nomenclature of medicine codes and were clinically confirmed. We used logistic regression analysis to examine the association of ancestry principal components (PC) with cSCC case/control status, adjusting for age and sex as covariates. We then ascertained whether ancestry associations were explained by predicted pigmentation phenotypes and known cSCC-associated loci. Pigment phenotypes were predicted using HliPlex-S, a program that utilizes known pigment genetic variants to generate probabilities of different pigment phenotypes (i.e. for eye, hair, and skin). **Results:** The first two principal components (PC1 and PC2) differentiate geographical clines across Europe, and both were strongly associated with cSCC risk ($P_{PC1}=1.0 \times 10^{-66}$; $P_{PC2}=2.9 \times 10^{-49}$), suggesting that individuals of northern and western European ancestry had greater risk of cSCC. The genetic ancestry associations were attenuated, but still highly significant, after including HliPlex-S pigmentation phenotypes in the model ($P_{PC1}=6.3 \times 10^{-23}$; $P_{PC2}=1.6 \times 10^{-23}$). Finally, we incorporated fourteen known cSCC genetic risk variants into the ancestry and HliPlex-S model, which further attenuated, but did not eliminate the ancestry signal ($P_{PC1}=7.3 \times 10^{-22}$; $P_{PC2}=6.2 \times 10^{-21}$). **Conclusions:** Our findings reveal that northern and western European ancestry is associated with a greater risk of cSCC in non-Hispanic white populations. Pigmentary traits and cSCC-associated loci explain some of the ancestry associations. However, additional cSCC risk factors that correlate with European ancestry remain to be discovered.

691T

Comparative analysis of the expression of PARKIN, APC and Ki67 in colorectal polyps and adenocarcinomas. C.C.V. da Silva-Camargo, W.A.C.C. Cecílio, L. Noronha, V.S. Sotomaior. Pontifícia Universidade Católica do Paraná, Curitiba, Paraná, Brazil.

Colorectal carcinoma remains the 3rd most common cause of cancer in Brazil. The wide spread of screening programs combined with the evolution of molecular biology, applied on the clinical practice, have raised the success rates in the treatment of such neoplasia. Based on the current knowledge of the sequence adenoma-adenocarcinoma, we aimed to analyze comparatively the immune-expression of the tumor suppressor gene *PARK2* (and its protein, PARKIN) with *APC* and *Ki67*, in samples of colorectal polyps and adenocarcinomas. We included neoplastic polyps, classified as low dysplasia adenoma, high dysplasia adenoma, sessile serrated polyps, and hyperplastic polyps. Through techniques of tissue microarray, we reviewed by immunohistochemistry, 284 polyps from 222 patients, besides 73 colorectal samples of adenocarcinoma. Due to the fact that more than half of the patients with polyps were found to have more than one lesion, we decided to elaborate a multilevel linear regression statistical model, in order to avoid interpretation bias. The association among quantitative variables was performed by the Person's correlation test, while the comparison between groups was assessed either by the t-test of Student or analysis of variance. There was no significant difference on variables related to patient (sex, age, multiplicity of polyps) and the tissue expression of the proteins. Considering the hierarchical multilevel structure, we only found association to APC and the right side of the colon and type of polyp (sessile serrated). Regarding the level of differentiation, there were found significant correlation ($p < 0.05$) to PARKIN and APC towards the group of carcinomas. The opposite was seen to Ki67, *i.e.*, higher levels in polyps with elevated grade of dysplasia. Such findings have similarities in literature, many authors support that PARKIN and APC act on synergism over the colorectal tissue. Although, there are no studies showing overexpression of such proteins in adenocarcinomas, compared to neoplastic polyps, as we have seen here. This findings point out that PARKIN and APC might be involved in later mechanisms of tumor progression control, in the pathway of carcinogenesis.

692F

Nicotinamide induces mitochondrial-mediated apoptosis through oxidative stress in human cervical cancer HeLa cells. Y. Feng^{1,2}, C. Jiang¹, Z. Fang¹, Z. Zhang¹, W. Jiang¹ (Corresponding Author)¹. 1) Medical Genetics, Zhongshan Medicine School, Sun Yat-sen University, Guangzhou, Guangdong Province, China; 2) Department of Endocrinology, Guangzhou Women and Children's Medical Center, Guangzhou, Guangdong Province, China.

Aims: Nicotinamide participates in energy metabolism and influences cellular redox status and modulates multiple pathways related with both cellular survival and death. Recent studies have shown that it induced proliferation inhibition and apoptosis in many cancer cells. However, little is known about the effects of nicotinamide on human cervical cancer cells. We aimed to evaluate the effects of the indicated concentrations nicotinamide on cell proliferation, apoptosis and redox-related parameters in HeLa cells and investigated the apoptotic mechanism. **Materials and Methods:** After the treatment of the indicated concentrations nicotinamide, HeLa cell proliferation was evaluated by the CCK-8 assay and the production of ROS (reactive oxygen species) was measured using 2',7'-Dichlorofluorescein diacetate. The apoptotic effect was confirmed by observing the cellular and nuclear morphologies with fluorescence microscope and apoptotic rate of HeLa cell apoptosis was measured by flow cytometry using Annexin-V method. Moreover, we examined the mitochondrial membrane potential by JC-1 method and measured the expression of apoptosis related genes using qRT-PCR and immunoblotting. **Key findings:** Nicotinamide restrained the HeLa cell proliferation and significantly increased the accumulation of ROS and depletion of GSH at relatively high concentrations. Furthermore, nicotinamide promoted HeLa cell apoptosis via the intrinsic mitochondrial apoptotic pathway. The mechanism may related G6PD inhibited by Nicotinamide. **Significance:** Our study revealed that nicotinamide induced the apoptosis through oxidative stress and intrinsic mitochondrial apoptotic pathways in HeLa cell. The results emerge that nicotinamid may be an inexpensive, safe and promising therapeutic agent or a neoadjuvant chemotherapy for cervical cancer patients, as well useful to find new drugs for cervical cancer therapy.

693W

Array comparative genomic hybridisation of familial prostate cancer tumours identifies a recurrent copy number gain on chr19p13.3 encompassing the *EEF2* gene. *L.M. FitzGerald¹, K. Raspin¹, J.R. Marthick¹, R.C. Malley^{1,2}, S. Donovan³, J.L. Dickinson¹.* 1) University of Tasmania, Tasmania, Australia; 2) Royal Hobart Hospital, Hobart, TAS, 7000, Australia; 3) Hobart Pathology, Hobart, TAS, 7000, Australia.

In a bid to discover genomic features associated with prostate cancer (PrCa) development and progression, copy number variations (CNVs) have been studied in tumour samples. Early comparative genomic hybridisation (CGH) studies led to the identification of many chromosomal regions of loss and gain, with a small number of these shown to be consistent across studies and, significantly, some suggested to be associated with PrCa progression. More recently a small number of studies have applied dense genome-wide SNP array platforms to fresh-frozen prostate tumours to identify recurrent CNVs, however these platforms are not suitable for the more widely available formalin-fixed, paraffin-embedded (FFPE) tumour samples. With the aim of replicating or identifying novel recurrent prostate tumour CNVs and elucidating the underlying genes involved, we applied the Agilent Oligonucleotide array-based CGH (aCGH), with both genome-wide and custom probes, to 12 FFPE prostate tumour DNA samples from a single Tasmanian family, PcTas9. Analysis of these data revealed as little as two to tens of CNVs present in each tumour, the majority of which were gains. In addition, several recurrent CNVs were identified, the most common of which was present on chromosome 19p13.3 and contained only two genes, including *EEF2*. *EEF2* has recently been shown to be overexpressed in various cancer types, including PrCa where it has been suggested to be associated with Gleason score. Currently, we are performing *EEF2* gene and protein expression studies in tumour samples from PcTas9 and additional Tasmanian familial and sporadic PrCa cases, and will investigate whether these are associated with Gleason score. Our findings to date suggest that increased *EEF2* protein levels observed in prior prostate tumour studies may be due to a copy number gain on 19p13.3. To test this hypothesis, gene and protein expression studies are currently underway and results will be presented at the meeting.

694T

Pathway-analysis of MAPK/ERK SNPs in gastric cancer. *P. Gonzalez-Hormazabal¹, M. Musleh², M. Bustamante³, J. Stambuk⁴, R. Pisano⁴, V.G. Castro¹, E. Lanza¹, H. Valladares², J. Rojas².* 1) Human Genetics Program, Institute of Biomedical Sciences, University of Chile; 2) Department of Surgery, University of Chile Clinical Hospital; 3) Department of Surgery, School of Medicine at Eastern Campus, University of Chile; 4) San Juan de Dios Hospital.

Background: Chile has one of the world's highest mortality rate from gastric cancer. Activation of MAPK/ERK pathway is key in cancer cell proliferation. Aim: We aimed to assess the association with gastric cancer (GC) of polymorphisms in MAPK/ERK pathway in a case-control study of 242 gastric adenocarcinoma patients and 192 controls. Methods: DNA samples were genotyped using Infinium Global Screening Array (Illumina) and we obtained genotypes of 21 SNPs belonging to *ARAF*, *BRAF*, *GRB2*, *HRAS*, *MAP2K1*, *MAP2K2*, *MAPK1*, *MAPK3*, *RAF1*, *SHC1*, *SOS1* and *SOS2*. Estimates of association under the additive model was obtained from logistic regression, adjusting by population stratification (PC1 and PC2 from PCA analysis implemented in PLINK 1.90). A p-value of 0.0023 was considered significant (Bonferroni correction). Results: Three SNPs was associated with GC: *MAPK1* rs6928 p=0.002, OR=1.57, *RAF1* rs3729931 p=0.002, OR=1.53, and *MAPK1* rs2283792 p=0.002 OR=1.55. An interaction between *H.pylori* infection and *MAP2K2* rs3500912 is suggested (p=0.039). All the associated SNPs are eQTL according to GTEx portal. Conclusions: These findings suggest that *MAPK1* rs6928, *MAPK1* rs2283792, and *RAF1* rs3729931 polymorphisms are associated with GC, and the risk of *MAP2K2* rs3500912 is modified by *H.pylori* infection.

695F

Impact of next generation sequencing on the analysis of familial breast cancer in African populations.

Y. Hamdi¹, M. Boujemaa¹, N. Mighri¹, N. Mejr², O. Jaidane³, S. Ben Nacer⁴, S. Labidi², J. Ben Hassouna², J. Ayari¹, H. El Benna², H. Bouazizi³, B. Mezghanni³, A. Haddaoui¹, K. Rahhal³, H. Boussen², M.S. Boubaker¹, S. Abdelhak¹, PEC Consortium. 1) Institute Pasteur of Tunis, Tunis, Tunisia; 2) Abderahmane Mami Hospital, Ariana, Tunisia; 3) Salah Azaiez Hospital, Tunis, Tunisia; 4) Military hospital, Tunis, Tunisia.

A family history of breast cancer (BC) has long been thought to indicate the presence of inherited genetic events that predispose to this disease. Actually, the major challenge of BC molecular genetics is to identify causal variants. Because of important differences in genetic patterns between ethnicities, association studies and haplotype analysis in different populations can help to overcome this challenge. Our present genome-wide-scanning was applied with Affymetrix-Genome-Wide-Human-SNP-Array-6.0 in order to determine allelic frequencies of the known BC associated variants in 135 Tunisians representing the general-population. We constructed and phased haplotypes carried by the studied Tunisian-population. Moreover, we assessed the biological effect of the putative functional variants by conducting eQTL-assays. We also compared BC associated-loci between different HapMap-populations by comparing LD-blocks and by conducting Principle-Component-Analysis. Our haplotype analysis showed that the Tunisian-population is more predisposed to the following at risk loci: 2p24, 4q21, 6q25, 9q31, 10q26, 11q13, 14q32. Four polymorphisms specific to the Tunisian population have also been identified including one on the *ESR1*-gene. Furthermore, we characterized a *BRCA1*-variant that can be considered as a causal polymorphism. In addition, we conducted Whole-Exome-Sequencing in 11 BC patients. In a first BC family, no deleterious mutation have been identified on known BC genes. However, functional annotation showed that *MMS19*, *POLK* and *KAT6* are novel breast cancer candidate genes that seems to be inherited in a family specific model. In a second BC family, we identified 2 cis-deleterious mutations on *BRCA2*. This cis-double heterozygosity have not been reported before and seems to be associated with a distinctive clinical phenotype. In this first WES application on Tunisian BC patients and by combining association, functional and bioinformatics analysis we highlighted the impact of next generation sequencing technologies on the identification of novel breast cancer candidate genes and novel genetic patterns such as the *BRCA2*-cis double heterozygosity. Further genetic analysis on additional African BC cases may bring new insights into the biological mechanisms of breast carcinogenesis which will provide a better clinical management of patients from under-represented populations. .

696W

Long non-coding RNA p10247, high expressed in breast cancer (lncRNA-BCHE), is correlated with metastasis.

T. Hayano¹, K. Zheng², Y. Yang³, L. Wei⁴, Y. Zhang⁵, M. Pilar Piñero Pereda⁶, H. Nakaoka⁷, Q. Li⁸, I. Baragan⁹, Y. Lu¹⁰, L. Tamagnone¹¹, I. Inoue¹², X. Li¹³, J. Luo¹⁴, H. You¹⁵. 1) Department of Systems Bioinformatics, Graduate School of Medicine Yamaguchi University, Ube, Yamaguchi, Japan; 2) Department of Endocrine and Breast Surgery, The First Affiliated Hospital of Chongqing Medical University, Chongqing, China; 3) Department of Medical Genetics, School of Basic Medical Sciences, Peking University, Beijing, China; 4) Key Laboratory of Birth Defects and Reproductive Health of the National Health and Family Planning Commission, Chongqing Population and Family Planning Science and Technology Research Institute, Chongqing, China; 5) Investigación Clínica y Traslacional del Cáncer, Virgen de la Victoria Hospital, IBIMA, Málaga, Spain; 6) Division of Human Genetics, National Institute of Genetics, Mishima, Japan; 7) The First Affiliated Hospital, Jinan University, Guangzhou, China; 8) Physiology and Pharmacology Department, Karolinska Institute, Snola, Sweden; 9) Department of Oncology, University of Torino, Candiolo, Italy; 10) State Key Laboratory of Agrobiotechnology, College of Biological Sciences, Beijing Advanced Innovation Center for Food Nutrition and Human Health, China Agricultural University, Beijing, China; 11) Affiliated Cancer Hospital & Institute of Guangzhou Medical University, Guangzhou, China.

Background: Breast cancer (BC) is the most prevalent malignance among women. Metastasis is one of the major obstacles to BC treatment and is responsible for more than 90% of BC related deceases. Therefore, it is necessary to better understand the genetic and molecular mechanisms leading to BC metastasis. Involvement of Long non-coding RNAs (lncRNAs) in BC has been proposed given their abnormal expression during the occurrence, recurrence and progression of BC. However, the pathophysiological contribution of lncRNAs in BC and specifically in the metastatic stage are still unclear. **Methods:** We examined the differential expression of lncRNAs between five pairs of invasive BC and their corresponding adjacent tissues by Agilent human lncRNA + mRNA Array V3.0. A differentially expressed lncRNA was further validated in 56 additional samples of BC. We also conducted knockdown and over-expression analysed of the lncRNA. **Results:** We identified a differential increase in the expression of 231 lncRNAs and a differential decrease in the expression of 625 lncRNAs in the BC. One of the differentially expressed lncRNA (lncRNA-BCHE) was remarkably upregulated (fold change more than 20, p value = 0.029) in the 56 samples of BC. Knockdown of lncRNA-BCHE inhibited the proliferation of MDA-MB-231 cells in vitro. Over-expression of lncRNA-BCHE promoted the migration and invasion of BC cells. **Conclusion:** All these data suggested that this novel lncRNA is a potential prognostic marker and a target gene therapy in BC.

697T

An integrative approach for mapping the mutation landscape of DNA repair genes and pathways in prostate cancer. C. Hicks¹, T. Mamidi¹, J. Wu¹, E. Nicklow¹, O. Sartor², E. Ledet², W. Liu¹. 1) Department of Genetics, 533 Bolivar Street. Louisiana State University Health Sciences Center, School of Medicine, New Orleans, LA. 70112; 2) Department of Medicine and Urology, Tulane University, School of Medicine, 150 S Liberty St, New Orleans, LA 70112.

Background: Prostate Cancer (PC) is the most diagnosed cancer and the second leading cause of cancer-related deaths among men in the United States. Pathogenic variants in genes involved in the DNA repair process play a role in the pathogenesis of PC. The recent surge of high-throughput sequencing of PC genomes has led to an expanding molecular classification of PC. However, to date, sequence data has not been optimally leveraged and integrated with transcriptome data to map the mutation landscape of DNA repair genes and pathways in indolent and aggressive PC. Here we integrated sequence information with gene expression data to map the mutation landscape of DNA repair genes, molecular networks and biological pathways in indolent and aggressive PC. Our working hypothesis is that genomic aberrations and alterations in genes, molecular networks and biological pathways involved in the DNA repair process could lead to measurable changes affecting therapeutic decision making. **Methods:** We have addressed this hypothesis as follows: We evaluated 311 genes involved in the DNA process for somatic mutations using TCGA data involving 190 men with indolent PC (Gleason score <7) and 304 men with aggressive PC (Gleason score = 8-10). We leveraged mutation information with gene expression data on the same patients and 52 matched normal to discover signatures of mutated genes. We performed network and pathway analysis to generate an atlas of interactive maps and pathways enriched for mutations and their functional connectivity. **Results:** This unbiased approach revealed gene signatures, molecular networks and biological pathways enriched for mutations prevalent in indolent and aggressive PC. Among the mutated genes discovered, 25 were associated with indolent PC and 35 were associated with aggressive PC, with 11 genes common to both disease states. Several genes including *BRCA2*, *BRCA1*, *ATM*, *CHEK2* and *PALB2* and pathways (DNA repair and DNA mismatch repair) contained mutations predicted to be deleterious. The clinical implications of these results is that mutations could be used to identify patients at high risk of developing lethal PC. **Conclusions:** The results show that somatic mutations in genes and pathways mediating the DNA-repair processes are prevalent among men with indolent and aggressive PC and tend to be higher in the latter. Future work will involve both germline and somatic mutations in genes and pathways among men with PC in multiple ethnic populations.

698F

Genotype/phenotype assessment in male *BRCA1/2* pathogenic variant carriers. S. Hiraki, L.R. Susswein, P.D. Murphy, M.L. Marshall, K.S. Hruska, R.T. Klein. GeneDx, Gaithersburg, MD.

Background: Pathogenic variants in *BRCA1* and *BRCA2* are known to increase the risk of breast and ovarian cancer in women; however, male *BRCA1/2* pathogenic variant carriers are also at increased cancer risk, including prostate and breast cancers. Ovarian cancer risks are known to vary by variant location in *BRCA1/2*, and there is some suggestion that genotype-phenotype correlations also exist for prostate cancer. In this study, we delineated the spectrum of cancers in male *BRCA1/2* pathogenic/likely pathogenic variant (PV) carriers, and investigated whether prostate cancer risks differed by variant location. **Methods:** We reviewed the results of all men tested for *BRCA1/2* at GeneDx as part of a multi-gene panel of 2 to 64 genes between August 2013 and April 2018. Self-reported cancer histories and type of PV identified were reviewed. Based on the location of a majority of PVs, putative *BRCA1/2* prostate cancer cluster regions (PCCR) were defined, and the relative risk of prostate cancer for variants inside versus outside these regions was calculated. Men with more than one PV in any gene were excluded from all analyses, and those with exon-level deletion/duplications were excluded from PCCR analyses. **Results:** Among 353 PVs identified, 63% were in *BRCA2* and 37% were in *BRCA1*. A personal history of cancer was reported in 169/222 (76%) *BRCA2* PV carriers and 65/131 (50%) *BRCA1* PV carriers. Prostate cancer was reported in 23% (52/222) of *BRCA2* PV carriers and 12% (15/131) of *BRCA1* PV carriers. We identified putative PCCRs in *BRCA2* (c.3847-c.6468) and *BRCA1* (c.4035-c.4357). Comparing men who carried a PV inside versus outside the *BRCA2* PCCR, 28/83 (34%) versus 23/134 (17%) had prostate cancer, translating to a relative risk of 1.97 ($p < 0.01$). Comparing men who carried a PV inside versus outside the *BRCA1* PCCR, 4/8 (50%) versus 8/104 (8%) had prostate cancer, but the small sample size of this cohort precluded further analyses. **Conclusions:** Prostate cancer is the most frequently reported cancer in men with *BRCA1/2* PVs. Prostate cancer diagnoses appeared to be more common in men with PVs located within *BRCA2* c.3847-c.6468, defining a potential PCCR. A *BRCA1* PCCR may also exist, however, larger numbers are needed to confirm this association. Further delineation of genotype-phenotype correlations for cancer risk may contribute to more personalized risk assessment and management for male *BRCA1/2* PV carriers.

699W

The incidence of occult ovarian neoplasia and cancer in *BRCA1/2* mutation carriers after the bilateral prophylactic salpingo-oophorectomy (PBSO): A single center prospective study. R. Janavicius^{1,2}, G. Januska¹, P. Jukna¹, V. Rudaitis^{1,2}. 1) VUH Santaros Klinikos, Vilnius, VNO, Lithuania; 2) State Research Institute Innovative Medicine Center, Vilnius, Lithuania.

Objective: to evaluate the incidence of occult neoplasia in specimen collected during PBSO and to determine the significance of this operation in *BRCA1/2* mutation carriers. **Methods:** Between January 2010 and October 2016 a total of 564 new germline *BRCA1/2* mutation positive women were identified at Hereditary Cancer Competence Center Unit of Hematology, Oncology and Transfusion Medicine Center, in Vilnius University Hospital Santaros Klinikos (VULSK). All these patients had undergone oncogenetic counselling, genetic testing, and corresponding cancer risk reduction management in the same institution. Patients were prospectively included in password protected electronic Patients Surveillance System Registry (ONCOGEN). As part of ovarian risk reduction management process, laparoscopic prophylactic bilateral salpingo-oophorectomy (PBSO) was performed for 71 eligible women, who opted for this procedure and were included in this perspective study. Fifty nine women (83.1%) were *BRCA1* and 12 (16.9%) were *BRCA2* mutation carriers; study patients harboured 22 different *BRCA1/2* germline mutations. **Results:** STIC was diagnosed in seven (9.85%; 95% CI 4.58 - 19.26) women and occult ovarian cancer (OC) was found in four (5.6%; 95% CI 1.80 - 14.03) women. The median age of women that were diagnosed with STIC or OC at the time PBSO was 45.9 years (± 6.68 ; $P=0.7314$), the youngest patient being 42 and the oldest 64 years. All 11 women who were diagnosed with STIC or OC had *BRCA1* mutations. Interestingly, STIC was detected in 6 carriers (out of 27) of exclusively *BRCA1* c.4035delA mutation (22.2%; 95% CI 10.26 - 41.10), a common Baltic founder mutation. The difference between STIC detected in c.4035delA (6/27) and other mutations carriers (1/44) was statistically significant ($P=0.0105$). The mean follow up time for women being diagnosed with ovarian cancer or STIC was 3.72 years (± 1.24). The median operating time to perform a laparoscopic PBSO was 45 minutes (± 9.94) ranging from 25 to 65 minutes. No surgical complications occurred during this operation. **Conclusion:** The detection rate of precursor-only lesions (STIC) in our study (9.85%) was slightly higher than previously reported (5-8%). We found statistically significant enrichment of *BRCA1* c.4035delA mutation carriers in STIC group, which may indicate that c.4035delA mutation carriers may have an increased tumorigenesis rate in fallopian tubes, which can also explain higher STIC detection rate in our series. Grant SEN18/2015.

700T

An association study of genetic variants located in DHSs with the risk of gastric cancer in Chinese population. Y. Jiang^{1,2}, F. Yu^{1,2}, T. Wang^{1,2}, Q. Qi^{1,2}, H. Ma^{1,2}, Z. Hu^{1,2}, G. Jin^{1,2}. 1) Department of Epidemiology and Biostatistics, Nanjing Medical University, Nanjing, Jiangsu, China; 2) Jiangsu Key Lab of Cancer Biomarkers, Prevention and Treatment, Collaborative Innovation Center for Cancer Medicine, Nanjing Medical University, Nanjing, Jiangsu, China.

Objective: Deoxyribonuclease I hypersensitive sites (DHSs) are markers of regulatory elements across genome. Genetic loci of diseases or traits identified genome-wide association studies (GWAS) are enriched in DHSs. We explored whether genetic variants in DHSs played an important role in susceptibility to gastric cancer (GC). **Method:** DHSs were defined on the basis of data from 125 tissues and cell lines of ENCODE. Genetic variants in DHSs were evaluated the association with GC risk in 3771 cases and 5426 controls from 4 existing GWAS of Chinese population. Expression quantitative trait loci (eQTL) were analyzed by using genotype and mRNA data from GTEx (170 normal stomach tissues) and TCGA (412 GC tissues) and differential expression analysis was based on TCGA (413 GC tissues and 32 normal tissues). According to the results of association, eQTL, differential expression, GC risk related loci were selected for subsequent validation in 6222 GC cases and 7510 controls by Sequenom and TaqMan assays. **Results:** We found a variant at 7q22.1 was significantly associated with GC risk (per allele: OR=0.85, 95%CI: 0.81-0.90, $P=2.56 \times 10^{-3}$). The genotypes were significantly related to the expression level of *SEMA3D* ($P=7.16 \times 10^{-3}$), and expression of *SEMA3D* was down-regulated in GC tissues (paired samples: $P=1.50 \times 10^{-5}$, unpaired samples: $P=8.97 \times 10^{-5}$). **Conclusion:** These results indicate that the genetic variant in DHSs on chromosome 7q21.11 may modulate susceptibility to gastric cancer through regulating the expression of *SEMA3D*, which further highlighted the important role of genetic variants in DHSs in the development of GC.

701F

The analysis on the mechanism of apoptosis induced by inhibiting G6PD activity in HeLa cells. *W. Jiang¹, Y. Feng^{1,2}, X. Lin¹, C. Jiang¹, P. Du¹, H. Luo¹, R. Hassan¹, J. Yan¹.* 1) Department of Medical Genetics, Zhongshan Medicine School, Sun Yat-sen University, Guangzhou, Guangdong Province, China; 2) Department of Endocrinology, Guangzhou Women and Children's Medical Center, Guangzhou, Guangdong Province, China.

Aims: Glucose-6-phosphate dehydrogenase (G6PD) is a critical rate-limiting enzyme for pentose phosphate pathway (PPP). 5-phosphate ribose that is the nucleotide component and NADPH that is not only to maintain the glutathione (GSH) reduction state, but also function as for biosynthesis are the main metabolites in PPP. G6PD plays an important role in regulating the metabolic reprogramming of tumor cells, which is involved in the processes of apoptosis, proliferation and cell-cycle progression. But the mechanisms remains elusive. We aim to assess the effects of DHEA and nicotinamide on G6PD activity and redox status in HeLa cells and explored the mechanism of DHEA-induced apoptosis of HeLa cells, searching for a new potential cancer therapeutic target. **Methods:** After the treatment of the indicated concentrations DHEA, HeLa cell proliferation was evaluated by the CCK-8 assay and the production of ROS (reactive oxygen species) was measured using 2',7'-Dichlorofluorescein diacetate. The apoptotic effect was confirmed by observing the cellular and nuclear morphologies with fluorescence microscope and apoptotic rate of HeLa cell was measured by flow cytometry using Annexin-V method. Moreover, the mitochondrial membrane potential was examined by JC-1 method and the expression of apoptosis related genes were analyzed using qRT-PCR and immunoblotting. **Key findings:** DHEA inhibited G6PD activity to induce HeLa cell apoptosis via the intrinsic mitochondrial apoptotic pathway. The mechanisms related with the increased intracellular ROS levels, decreased GSH content. G6PD knockdown was associated with cell-cycle arrest during the G0/G1 or S phase, which promoted cell apoptosis. **Significance:** Our study demonstrated that DHEA inhibiting G6PD could induce ROS-mediated apoptosis, indicating the important role of G6PD in HeLa cells of redox regulation to suppress the cervical cancer and provided a new evidence that inhibition of G6PD activity may be useful for cervical cancer therapy.

702W

The genomic landscape of recurrent ovarian cancer reveals mutational processes and functional mechanisms that drive chemoresistance. *M.R. Jones¹, S.G. Coetzee¹, N.M. Gull¹, A.L.P. Reyes¹, J.T. Plummer¹, B.D. Davis¹, S. Chen¹, J.P.B. Govindavar², J. Rutgers², J. Lester², K. Lawrenson³, D.J. Hazelett¹, B.P. Berman^{1,4}, B. Karlan³, S.A. Gayther^{1,4}.* 1) Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, CA; 2) Dept of Pathology and Laboratory Medicine, Cedars-Sinai Medical Center, Los Angeles, CA; 3) Women's Cancer Program, Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA; 4) Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA.

Seventy percent of high grade serous ovarian cancer (HGSOC) patients are diagnosed with late stage disease when 5-year survival rates are ~30%. Patients initially respond to chemotherapies, but tumor recurrence occurs in 80% of cases leading to chemotherapy resistant disease that is ultimately fatal. More than 15% of serous OC patients carry a germline mutation in BRCA1/2, which account for approximately 30% of the heritability of HGSOC. The genomes of HGSOC tumors are highly unstable; BRCA1/2 mutant tumors are deficient in homologous recombination repair while tumors from patients without BRCA1/2 mutations harbor high numbers of foldback inversions. To identify the genomic mechanisms underlying chemoresistance in HGSOC we studied a cohort of 34 patients (13 BRCA1/2 mutation carriers; 21 patients with no identified genetic predisposition) with paired tumor specimens that were chemotherapy naive (from primary debulking surgery) as well as >1 recurrent tumor (chemotherapy resistant). We generated a genomic landscape of the changes that occur as a result of chemoresistance in each tumor (n=73) with whole genome sequencing, whole genome bisulfite sequencing and RNA-Seq. Somatic single nucleotide variant, Indel and structural variant (SV) analysis reveal monoclonal and polyclonal seeding of recurrent tumors by populations of clones from the primary tumor carrying mutations allowing them to survive the purifying selection process of chemotherapy. An range in increased mutational load in the recurrent tumors in each patient (1,069-15,913 SNVs) indicates diverse mutational mechanisms between patients. Mutational signatures in the primary tumor were dominated by BRCA and Age signatures, while recurrent tumors frequently differed from the matched primary and were enriched for Signatures 8 (Breast Cancer; C>A) and 12 (Liver Cancer; T>C). The genes most frequently mutated in recurrent tumors were LINC00969, RUNX1T1, and SEPT9; the candidate ovarian tumor suppressor gene. Two patients (non-BRCA mutation carriers) showed a unique sub-population of clones that were able to leave the peritoneal cavity and metastasize to the brain. Mutations identified across recurrent tumors identify pathways that can be targeted using existing therapies in platinum-resistant patients. To improve survival for women with HGSOC, discovering novel therapeutic opportunities for platinum-resistant disease remains a top priority.

703T

Early lifestyle factors are associated with an increased risk of atopic disorders in lymphoid cancer families. S.J. Jones^{1,2}, S. Stroschein³, D. Liu⁴, J. Spinelli⁵, J.M. Connors⁶, A. Brooks-Wilson^{2,3}. 1) Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada; 2) Canada's Michael Smith Genome Sciences Centre, British Columbia Cancer Agency, Vancouver, Canada; 3) Department of Biomedical Physiology and Kinesiology, Simon Fraser University, Burnaby, Canada; 4) Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, Canada; 5) Cancer Control Research Program, British Columbia Cancer Agency, Vancouver, Canada; 6) Centre for Lymphoid Cancers, BC Cancer Agency, Vancouver, Canada.

The etiology of lymphoid cancers is multifactorial. Research suggests that infectious diseases and immune disorders may influence the risk of lymphoid cancer. The hygiene hypothesis proposes that a lower infectious burden during early life prevents the immune system from maturing optimally and may cause adult-onset immunogenic disorders. We evaluated birth order and sibship size as indicators of early lifestyle in 192 families with a history of lymphoid cancer. Data was collected from 443 lymphoid cancer cases and 987 unaffected siblings among 341 lymphoid affected sibships. The relationship between lifestyle factors and risk of lymphoid cancer, allergies, and autoimmune conditions were examined using a linear regression with generalized estimating equation to accommodate correlated family data. Odds Ratios (ORs) were adjusted for age and sex. Birth order and sibship size were independently associated with risk of lymphoid cancer. Increasing birth order was inversely associated with risk of cancer for all lymphoid cancers collectively (OR=0.79; 95% CI:0.74-0.85), and separately for multiple myeloma (OR = 0.35; 95% CI:0.15-0.77), non-Hodgkin lymphoma (OR=0.65; 95% CI:0.53-0.79) and chronic lymphocytic leukemia (OR=0.73; 95% CI:0.54-0.98). That is, individuals with an earlier birth order position had a higher risk of lymphoid cancer than later born siblings. Effects of birth order were seen in smaller families with 2 or 3 lymphoid cases but were not observed in families with 4 or more affected. Birth order was also inversely associated with risk of allergies (OR=0.83, 95% CI:0.68-0.99) but not autoimmune disease. The associations between lymphoid cancer and atypical immune function have been explored in small cohort and large population studies, however, there is limited evidence in the familial context. The inverse relationship between birth order and risk of disease has been documented for allergies and some lymphomas, but no association has been reported for autoimmune diseases. The observed inverse relationship between birth order and sibship with risk of lymphoma is supportive of the hygiene hypothesis, and that childhood exposure to infectious disease may play a role in the lymphatic malignancies. The familial nature of these cancers implies a role of shared genetic and/or environmental factors. These effects may be modified by lifestyle factors that correlate with birth order and could identify protective lifestyle factors, even in the familial context.

704F

Mendelian randomization study of pulmonary function and lung cancer risk using novel genetic instruments developed in the UK Biobank cohort. L.

Kachuri¹, M. Johansson², P. Brennan³, S.R. Rashkin⁴, R.E. Graff⁵, N.C. Emami¹, N.E. Caporaso⁶, M.T. Landi⁷, D.C. Christiani⁸, P. Vineis⁹, G. Liu⁶, G. Scelo², X. Wu⁷, D. Albanes³, M.C. Aldrich⁸, A. Tardón⁹, G. Rennett¹⁰, C. Chen⁶, G.E. Goodman¹¹, J.A. Doherty¹², J.K. Field¹³, M.D. Teare¹⁴, L.A. Kiemeny¹⁵, S.E. Bojesen¹⁶, S. Zienoldiny¹⁷, A. Haugen¹⁷, S. Lam¹⁸, L. Le Marchand¹⁹, M.B. Schabath²⁰, E.J. Duell²¹, J.D. McKay²², C. Relton²³, R. Martin²⁴, G. Davey Smith²⁵, P.C. Haycock²⁶, C.I. Amos²⁷, J.S. Witte¹, R.J. Hung²⁴ On Behalf of the OncoArray Lung Cancer Consortium. 1) Department of Epidemiology & Biostatistics, University of California at San Francisco, San Francisco, CA, United States; 2) International Agency for Research on Cancer, Lyon, France; 3) Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA; 4) Departments of Epidemiology and Environmental Health, Harvard TH Chan School of Public Health, Boston, MA, USA; 5) MRC/PHE Centre for Environment and Health, Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, United Kingdom; 6) Ontario Cancer Institute, Princess Margaret Cancer Center, Toronto, ON, Canada; 7) Department of Epidemiology, Division of Cancer Prevention and Population Sciences, The University of Texas MD Anderson Cancer Center, Houston, TX, USA; 8) Department of Thoracic Surgery and Division of Epidemiology, Vanderbilt University Medical Center, Nashville, TN, USA; 9) University of Oviedo and CIBERESP, Faculty of Medicine, Campus del Cristo, Oviedo, Spain; 10) Clalit National Cancer Control Center at Carmel Medical Center and Technion Faculty of Medicine, Haifa, Israel; 11) Fred Hutchinson Cancer Research Center, Seattle, WA, USA; 12) Department of Population Health Sciences, Huntsman Cancer Institute, Salt Lake City, Utah, USA; 13) Roy Castle Lung Cancer Research Programme, University of Liverpool Cancer Research Centre Institute of Translational Medicine, University of Liverpool, Liverpool, UK; 14) School of Health and Related Research, University of Sheffield, Sheffield, UK; 15) Radboud Institute for Health Sciences, Radboud University Medical Centre, Nijmegen, The Netherlands; 16) Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Denmark; 17) The National Institute of Occupational Health, Oslo, Norway; 18) BC Cancer Agency, Vancouver, BC, Canada; 19) Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI, USA; 20) Department of Cancer Epidemiology, H. Lee Moffitt Cancer Center & Research Institute, Tampa, FL, USA; 21) Unit of Nutrition and Cancer, Catalan Institute of Oncology (ICO-IDIBELL), Barcelona, Spain; 22) Population Health Science, Bristol Medical School, MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK; 23) Institute for Clinical and Translational Research, Baylor College of Medicine, Houston, TX, USA; 24) Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, ON, Canada.

Individuals with impaired lung function have an increased risk of developing lung cancer. However, since impaired lung function is also an indicator of obstructive pulmonary disease, and a consequence of tobacco smoking, it is unclear whether the observed relationship with lung cancer is causal. To investigate this hypothesis, we developed genetic instruments for each lung phenotype using genome-wide analyses full UK Biobank cohort (UKB), and applied these novel instruments in Mendelian randomization (MR) analyses of 24 studies from the OncoArray collaboration. Lung function was measured as forced expiratory volume in 1 second (FEV₁), forced vital capacity (FVC), and FEV₁/FVC, based on reproducible spirometry (within 0.15 L) from European ancestry subjects. GWAS of 15.5 million variants for FEV₁ (n=371,892), FVC (n=369,830), and FEV₁/FVC (n=367,764) identified over 400 independently associated loci (p<5×10⁻⁸; r²<0.05) for each trait. Heritability estimated using LD-Score regression was 18% for FEV₁ and FVC, and 14% for FEV₁/FVC. After filtering indels and palindromic SNPs, the following instruments were developed: FEV₁ (374 SNPs, 6.1% of variation explained), FVC (318 SNPs, 4.9%), and FEV₁/FVC (431 SNPs, 8.9%). Causal odds ratios (OR) per 1 standard deviation (FEV₁, FVC) and 1 unit (FEV₁/FVC) increase were estimated using summary statistics from 18,806 lung cancer cases (adenocarcinoma: n=7,039; squamous carcinoma: n=4,550) and 14,220 controls (>80% European ancestry). Maximum-likelihood MR estimates indicated that genetic determinants of reduced FEV₁ were associated with increased risk of lung cancer overall (OR=1.26, 1.05-1.50), lung adenocarcinoma (OR=1.32, 1.05-1.67), and squamous cell carcinoma (OR=1.28, 0.97-1.69). These findings were supported by MR Robust Adjusted Profile Scores (OR lung: 1.21, 1.03-1.43; OR adeno: 1.27, 1.01-1.59), and the direction of the causal effect was confirmed by MR Steiger. There was little evidence of heterogeneity among FEV₁ causal effect estimates (Cochrane's Q p-value>0.05). Despite the high genetic correlation between FEV₁ and FVC (r_g=0.77), no statistically significant associations with lung cancer were observed for FVC or FEV₁/FVC. Our findings suggest that FEV₁ capture distinct pathways that may be implicated in lung cancer susceptibility. However, these results should be interpreted with caution since pleiotropic effects cannot be ruled out, and the contribution of smoking-mediated pathways remains to be clarified.

705W

Genetic polymorphisms in the *SPC25*, *LOC285762* and *NPY* genes are associated with colorectal advanced adenoma in a Korean population.

S. Kang¹, J. Lee², S. Chung¹, E. Shin², E. Choe¹, S. Choi¹. 1) Seoul National University Hospital Gangnam Center, Seoul, Korea, Republic of; 2) DNA Link, Inc., Seoul, Korea, Republic of.

Colorectal cancer (CRC) is a leading cause of morbidity and mortality in the developed countries. Most CRCs develop from adenomas in a well-described adenoma-carcinoma sequence. Cancer risk of adenoma with advanced features (size more than 1cm, villous elements or high grade dysplasia, and adenoma more than 3) is particularly high compared with non-advanced adenoma. We performed the first genome-wide association study (GWAS) on susceptibility to advanced adenoma in a Korean population. A GWAS was conducted among 7,948 healthy individuals; among these cases, the test set comprised 625 cases and 600 controls, and the replication set comprised 283 cases and 243 controls. Advanced adenoma was diagnosed by colonoscopy during comprehensive medical check-ups, and single-nucleotide polymorphisms (SNPs) related to advanced adenoma were detected with the Affymetrix Axiom™ KORV1.1-96 Array. In total, 6 SNPs were identified in three SNP aggregates in the test set ($P < 10^{-3}$, within 200 kb) after adjusting for age and sex. All the SNPs were replicated in the replication set ($P < 0.05$). Two SNPs were near the *SPC25* gene, two SNPs were near the *LOC285762* gene, and two SNPs were near the *NPY* gene. The top SNP associated with colorectal advanced adenoma was rs78127544, located near the *LOC285762* gene [combined set, P -value = 1.809×10^{-4} , odds ratio = 1.556 (95% confidence interval: 1.235-1.961)]. We found 6 novel SNP alleles associated with advanced adenoma that implicated three gene loci, the *SPC25*, *LOC285762*, and *NPY* gene, as being involved in the underlying genetic susceptibility to colorectal advanced adenoma. Our results could provide new genetic insights into the development of advanced adenoma in Asian populations.

706T

Chromosome 10q24.32 variant increases risk of non-melanoma skin cancer and Bowen's disease in arsenic exposed population. M.G. Kibriya¹, F. Jasmine¹, B.L. Pierce¹, M. Argos², M.T. Islam³, M. Rakibuz-Zaman³, A. Ahmed³, H. Shahriar³, T. Islam³, J. Graziano⁴, C.R. Shea⁵, M. Kamal⁶, H. Ahsan^{1,7,8}.

1) Public Health Sciences, University of Chicago, Chicago, IL, USA; 2) Division of Epidemiology and Biostatistics, School of Public Health, University of Illinois, Chicago, IL, USA; 3) University of Chicago Research Bangladesh, Dhaka, Bangladesh; 4) Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, USA; 5) Department of Dermatology, University of Chicago, IL, USA; 6) Department of Pathology, Bangabandhu Sheikh Mujib Medical University (BSMMU), Dhaka, Bangladesh; 7) Cancer Research Center, University of Chicago, Chicago, IL, USA; 8) Institute of Population and Precision Health, Biological Sciences Division, University of Chicago, Chicago, IL, USA.

Background: Non-melanoma skin cancers (NMSC) are the most prevalent malignancy in the US, with an estimated 2 million new diagnoses each year. NMSC includes basal cell carcinoma (BCC) and squamous cell carcinoma (SCC). Bowen's disease (BD) is actually squamous cell carcinoma in-situ. While ultraviolet (UV) light exposure and skin sensitivity are known risk factors for NMSC especially in Caucasians, arsenic (As) exposure through contaminated drinking water or diet may be a risk factor in some populations. Arsenical skin lesion (ASL) is a hallmark of As toxicity. From population-based genome-wide investigations in As exposed individuals, we previously identified six SNPs in chr10q24.32 region (rs1046778, rs11191527, rs11191659, rs4290163, rs4919694 and rs9527) associated with ASL and As metabolism. We also identified a number of CNVs in germline DNA predisposing individuals to develop ASL. In this study, we prospectively examine if those genomic markers are associated with NMSC and BD in As-exposed Bangladeshi population. **Material:** From a large and long-term (96 months) prospectively followed-up cohort, we randomly selected 2171 subjects (m=1032, f=1139) at enrollment and genotyped their whole blood DNA samples at baseline on Illumina SNP chips to generate DNA copy number. We previously reported the SNPs and CNVs that were associated with ASL in this study population. Participants were followed up every 2 years for a total of 8 years, especially for the development of NMSC and BD. Among these subjects, the clinically suspected NMSC were subjected to skin biopsy. Cases were defined as BD, SCC or BCC. **Result:** Out of the 2171 subjects, a total of 416 (M=301, F=115) developed ASL during follow-up and 70 of them (M=37, F=33) were biopsied for suspected BD or NMSC. Histopathology showed that 31 had BD, 2 had SCC and 14 had BCC. Keratosis with dysplasia was seen in 16 and the rest 7 had other dermatological lesions. Individuals with variant allele for SNP rs11191659 near the gene *AS3MT* showed significant association with NMSC and BD cases combined (OR 2.45, 95%CI 1.13 – 5.33). This association remained unchanged (OR 2.44, 95%CI 1.12 – 5.33) even after adjustment for age, gender, and baseline arsenic exposure. None of the other five 10q24.32 SNPs, and none of the CNVs previously showing associations with ASL showed association with NMSC. **Conclusion:** Genomic variation in chromosome 10q24.32 region may have association with NMSC risk in a Bangladeshi population.

707F

Small RNA sequencing of 441 normal prostate transcriptomes reveals prostate cancer risk-SNP eQTL associations with miRNA expression. N.B. Larson, S.K. McDonnell, S. Riska, Z. Fogarty, Y. Liu, S. Baheti, M.S. DeRyck, L.S. Tillmans, D.J. Schaid, S.N. Thibodeau. Mayo Clinic College of Medicine and Science, Rochester, MN.

Large genome-wide association studies have to date identified 284 single-nucleotide polymorphisms (SNPs) associated with increased risk of prostate cancer (PRCA). Many of the identified variants are located in intergenic regions of the genome and presumed to be regulatory in nature. Expression quantitative trait loci (eQTL) analyses have been successful in linking PRCA risk SNPs to genes through evidence of transcriptional dysregulation, and prostate-specific eQTL studies have begun to unravel connections between PRCA risk SNPs and protein-coding mRNAs. However, less work has been directed toward non-coding RNAs, including miRNAs. Recent evidence demonstrates that pri-miRNA transcripts are susceptible to the same cis-acting regulatory effects as other long RNAs, and PRCA risk SNPs may confer systemic impacts on the prostate transcriptome through miRNA expression dysregulation. In this study, we performed small RNA sequencing on a large set of normal prostate tissue samples with available genome-wide genotyping and RNA-Seq expression data to investigate the role of miRNAs in the genetic risk of PRCA. Sequencing was performed on an Illumina HiSeq 4000 instrument using NEBNext® Multiplex Small RNA Library Prep Kit with sample multiplexing at up to 48 samples per lane. Sequencing results were processed using the CAP-miRSeq miRNA workflow and trimmed mean of M (TMM) normalization, identifying 605 expressed mature miRNAs (≥ 5 reads in $>25\%$ of samples) across 441 samples that passed quality control. We then performed a systematic cis-eQTL association study focused on PRCA risk genetics by testing all SNPs in high LD ($r^2 \geq 0.5$) with PRCA risk SNPs, consisting of 5264 total tests (2608 unique SNPs and 82 unique miRNAs). Under a gene-based significance threshold of $FDR < 0.05$, we identified two significant cis-eQTL associations: rs1058319 with the miR-941 family ($P = 4.6E-09$, $B = -0.187$) on chromosome 20, and single-base deletion chr10:104317808:1_CA with mir-1307-3p ($P = 1.6E-05$, $B = 0.105$). Follow-up analysis of transcriptome-wide protein-coding trans-eQTL results for rs1058319 also indicated significant differences in effect estimates between miR-941 target and non-target genes (Wilcoxon $P = 0.0057$) based on TargetScan, with results consistent with the expression silencing mechanism of miRNAs. Future studies aim to further integrate small- and long-RNA expression datasets to better elucidate the complex transcriptomic dysregulation associated with PRCA genetics.

708W

Novel pathogenic variants for high-risk breast cancer identified from whole exome sequences of Korean BRCAX cases. J.-Y. Lee¹, J. Kim², J.W. Lee², S.W. Kim², J. Sung¹. 1) Seoul National University Graduate School of Public Health, Seoul, Korea, Republic of; 2) Asan Medical Center, Seoul, Korea, Republic of; 3) Daerim St. Mary's Hospital, Seoul, Korea, Republic of.

Introduction Genome-wide association studies (GWAS) of BRCAX cases (high-risk cases without pathogenic mutation on *BRCA1/2*) have found several novel candidate loci associated with breast cancer. Common markers identified from GWASs have inborn limitations in that they have limited impacts and the markers are not likely the causative variants. In this study, we aimed to find rare variants associated with BRCAX using whole exome sequencing (WES) data of high-risk breast cancer cases in Korea. **Materials and Methods** We included 172 high-risk breast cancer cases without pathogenic mutation on *BRCA1/2*—early onset (<40), with strong family history, with past medical history of other cancer, or bilateral cases from the Korean Hereditary Breast Cancer (KOHBCA) study. A WES analysis was performed with Nextera Rapid Capture Exome and Expanded Exome Kit. WES data were processed using BWA, Picard, and GATK, according to GATK Best Practices. We scanned through all exonic regions and prioritized variants with Variant Effect Predictor (VEP). First, we excluded common variants (minor allele frequency $>0.1\%$ in the 1KG, ExAC, or gnomAD) and benign variants predicted by SIFT, PolyPhen2, and CADD (<15). Then, variants predicted to be involved in protein truncation were retained. Finally, the frequencies of these variants were compared to the Korean Reference Genome ($N=397$). Using all these information, pathogenic or likely pathogenic variants defined by the ACMG guideline were investigated. **Results** For 172 BRCAX cases, 295K variants (on average, 1.7K variants per individual) were called and the mean depth was 42. After filtering common or benign variants, 896 markers with high impact remained and 121 of them were damaging according to multiple prediction tools. We classified 32 variants as likely pathogenic, as they were absent or extremely rare in public databases (PM2) or in control data (PS4), predicted to be deleterious by multiple tools (PP3), and located in mutational hotspot (PM1). Several identified genes were reported from the GWAS or suggested for panel sequencing for breast cancer. Two novel truncating variants located on *PANK4* and *ERCC5* were identified as heterozygous mutations from four and two cases, respectively. **Discussion** We identified novel pathogenic variants, by examining the WES data of BRCAX cases. We suggest two candidate genes associated with high-risk breast cancer, which could be used for screening after validation with larger samples and functional assays.

709T

Identifying candidate genes associated with breast cancer predisposition in *BRCA1/2* mutation-negative individuals. A.S.G. Lee^{1,2,3}, P. Munusamy¹, G.L. Koh¹, C.H.T. Chan¹, E.S.Y. Wong¹, S.Y. Loke¹, W.K. Lim⁴, J.X. Teo¹, M.H. Tan^{5,6,7}, Y.S. Yap⁸, P. Ang^{5,8}. 1) Division of Cellular and Molecular Research, National Cancer Centre Singapore, Singapore; 2) Department of Physiology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore; 3) Office of Clinical & Academic Faculty Affairs, Duke-NUS Medical School, Singapore; 4) SingHealth Duke-NUS Institute of Precision Medicine, Singapore; 5) Department of Medical Oncology, National Cancer Centre Singapore, Singapore; 6) Institute of Bioengineering and Nanotechnology, Singapore; 7) Lucence Diagnostics Pte Ltd, Singapore; 8) Oncocare Cancer Centre, Gleneagles Medical Centre, Singapore.

Purpose: Hereditary breast and/or ovarian cancer (HBOC) is a cancer predisposition syndrome, associated with mutations in *BRCA1/2* or other genes at a lower frequency. However, mutations are not detected in many high-risk individuals who have undergone multigene panel testing, suggesting that additional genes remain to be discovered. This study aimed to identify novel genes associated with breast cancer (BC) predisposition. **Methods:** Targeted next-generation sequencing for 283 cancer-associated genes, including 25 known or proposed BC susceptibility genes (25-genes) was performed on 239 individuals eligible for *BRCA1/2* genetic testing recruited at the National Cancer Centre, Singapore. Using extensive bioinformatics analysis, we applied multiple variant prioritization approaches to identify likely candidate genes. Case-control association analysis of candidate genes was done using East-Asian controls from the Exome Aggregation Consortium (ExAC). **Results:** Of the 239 subjects, 149 tested negative for 25-genes. From our in-depth analysis of these subjects, likely pathogenic variants in *DDR2* (2.68%), *EMSY* (4.7%), *APC* (6.71%), *AKT3* (2.01%) and *ARAF* (2.01%) were identified. Importantly, subjects who tested positive for the 25-genes did not harbor variants in these five genes. Association analysis revealed significant associations of these five genes with BC susceptibility at $P < 0.05$, with odds ratios ranging from 2.37 to 7.92. **Conclusion:** We report pathogenic or likely pathogenic variants in *DDR2*, *EMSY*, *APC*, *AKT3* and *ARAF* in *BRCA1/2* mutation-negative cases. Further studies on additional populations and larger cohorts, as well as segregation and functional studies are warranted.

710F

Cervical cancer in Guatemala: Somatic mutations and molecular analyses of the HPV16 D2 and D3 sub-lineages. H. Lou¹, J. Boland¹, E. Gonzalez², A. Albanez³, W. Zhou¹, M. Steinberg¹, L. Diaw¹, D. Roberson¹, M. Cullen¹, L. Garland¹, F. Pinto³, S. Bass¹, R. Burk⁴, M. Yeager¹, N. Wentzensen⁴, M. Schiffman⁴, L. Mirabello⁴, E. Gharzouzi⁵, M. Dean². 1) NCI-Leidos Biomedical Research Inc., Gaithersburg, MD; 2) Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, Leidos Biomedical Research, Inc., National Laboratory for Cancer Research, Gaithersburg, MD; 3) Instituto de Cancerologia, Guatemala City, Guatemala; 4) Laboratory of Cancer Genetics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Gaithersburg, MD USA; 5) Albert Einstein College of Medicine, Bronx, NY.

Cervical cancer (CC) is the most frequent cause of cancer mortality for women living in poverty, and is almost exclusively caused by human papilloma virus (HPV). We have surveyed 665 cases of cervical cancer from Guatemala. The average age is 52 (mean 51), average number of children is 6.0 (mean 5.0), 5.6% report tobacco use; 56% have stage 2 or 3 cancer and 82% of tumors are squamous cell carcinoma. In total 20% of subjects report a family history of any cancer and 6% for CC. Patients with a family history of CC have a younger age of onset (47 years), and were pre-menopausal cancer (57% versus 42% in the entire sample). While tobacco use is low, a high percentage (73%) cook at home with wood, and therefore are exposed to airborne carcinogens. To further understand the somatic mutations, we sequenced 24 tumor DNA samples for 245 cancer-related genes. In total, 11/24 tumors had a mutation in the *PIK3CA* gene (46%) and 58% of tumors had at least one mutation in the PI3K pathway. Other known driver mutations included *TP53*, *RB1*, and *CAP88*. Mutations in chromatin remodeling genes (*PBRM1*, *EP300*, *SMARCA4*, *KMT2C*, *KMT2D*, *HIST1H3B*, *HIST1H4E*, *HIST1H1E*) were also elevated with 46% of tumors having at least one mutation. Tumors had an average of 18 mutations (median 14), however 4 tumors had high (32-38 mutations) or very high (63-64) mutation load. All hypermutation subjects have *PIK3CA* mutations and are post-menopausal. Mutation signature analysis showed that the highest signature is APOBEC-related mutations (46%) and HPV is known to activate APOBEC. Pre-menopausal patients had a lower mutation load (13 mutations) and a lower fraction of APOBEC-related mutations (40%). Overall, HPV16 accounts for 58% of HPV infections in Guatemala. We have sequenced and variant classified 96 HPV16 + CC and found that the D2 and D3 sub-lineages represented 26% and 29% of the samples, respectively. A total of 65% (62/96) of the samples had integrated HPV16 sequences as determined by HPV DNA Capture on NimbleGen Array and A1 and D2 sub-lineages showed a higher frequency of integration compared to D3. Subjects with HPV16 integration have a significantly younger age, and D2 was observed in younger patients, as compared to A1. In conclusion, Guatemalan cervical tumors have a similar profile of somatic mutations to those in the US, with a high frequency of *PIK3CA* mutations, and a high frequency of the very high-risk HPV16 D2, D3 sub-lineages.

711W

Evaluate a new panel of SNPs to predisposition of development of thyroid cancer in admixture population. A.D. Luchessi¹, I.C.C. Santos¹, J. Genre², D.C.S. Marques³, A.M.G. Silva³, J.C. Santos³, J.N.G. Araujo³, V.H.R. Duarte³, A.M.A. Carracedo³, M. Torres³, G.M. Bastos³, C.C.F. Ramos³, V.N. Silbier⁴. 1) Federal University of Rio Grande do Norte, Natal, Rio Grande do Norte / Natal, Brazil; 2) Department of Clinical and Toxicological Analysis, Federal University of Rio Grande do Norte, Natal, Rio Grande do Norte, Brazil; 3) Centro Nacional de Genotipagem, Santiago de Compostela, La Coruña, Spain; 4) Department of Clinical Analysis and Toxicology of São Paulo University, São Paulo, SP, Brazil; 4) Liga Norte Riograndense Contra o Câncer, Rio Grande do Norte, RN, Brazil.

Thyroid cancer (TC) is the most common endocrine tumor. Genetic predisposition to this cancer might be associated with multiple penetrance genes that interact with each other and with the environment. Genetic variation are becoming useful biomarkers for diagnosis and risk stratification, and as potential treatment targets for aggressive forms of tumor. However, there is limit study of genetic predisposition to TC performed in the admixture population. In this way, the present project aims to evaluate the genetic predisposition of admixture population in Brazil to develop the TC. To this purpose, 580 sample were select, 279 paraffin blocks of the thyroid gland adjacent to the tumor from bank of the LNRCC Pathology Laboratory to compose the study group, and venous blood from 301 health donors from Hemovida Blood Bank as a control group. Genomic DNA from the paraffin blocks and blood samples were extracted and the Genotyping of a panel of 84 SNPs was performed using the MassaiRA[®] iPLEX Gold Mass Spectrometer MALDI-TOF Sequenom[®]. Thirty-six SNPs showed significant differences in their allelic frequencies when compared to cases and controls. Of these, 23 SNPs were associated with the risk of developing CT, mainly variants rs2997312 (OR = 6.33, IC95%: 1.80-22.29, P= 0.00085), rs116909374 (OR = 3.40, IC95%: 1.57-7.40, P= 0.00093), rs7024345 (OR = 2.97, IC95%: 1.05-8.46, P= 0.03032), rs965513 (OR = 2.91, IC95%: 1.62-5.24, P= 0.00020), and rs10788123 (OR = 2.78, IC95%: 1.60-4.82, P= 0.00016), since they confer greater risk to patients. In addition, 13 SNPs were presented as a protective factor for the non-development of this neoplasia, between them rs2651339 (OR= 0.48, IC95%: 0.34-0.69, P= 0.00005), rs34566348 (OR= 0.51, IC95%: 0.34-0.78, P= 0.00148), rs664677 (OR= 0.56, IC95%: 0.40-0.80, P= 0.00135), rs1801516 (OR= 0.57, IC95%: 0.37-0.87, P= 0.00802) and rs13447450 (OR= 0.57, IC95%: 0.40-0.82, P= 0.00186). Thus, we suggest a panel of SNPs that may represent biomarkers of predisposition or protection in the development of CT in brazilian admixture populations.

712T

Characteristics of potentially actionable germline findings identified on tumor-only sequencing: Implications for cancer risk assessment. K.N. Maxwell, D. Farengo-Clark, J. Powers, D.B. Lieberman, S.M. Domchek, J.J.D. Morrisette, K.L. Nathanson. University of Pennsylvania, Philadelphia, PA.

Background: Next-generation sequencing for tumor DNA profiling (tNGS) is increasingly advocated for all cancer patients given potential therapeutic indications. It is critical to understand possible incidental germline findings that may be revealed by tNGS, given the clinical implications for the patient and family. **Methods:** tNGS clinical reports with potentially actionable pathogenic germline variants (PV) were identified by a novel pipeline implemented at the University of Pennsylvania in 2016, with details described separately (Clark et al). Characteristics associated with the tumors and PVs were analyzed. **Results:** Forty-eight patients identified by the tNGS pipeline from 2016-2018 underwent confirmatory germline testing, and 28/48 (58%) had mutations that confirmed to be germline. Patients were most commonly referred for PVs in *BRCA2* (n=14), followed by *TP53* (n=6), Lynch genes (n=6), *BRCA1* (n=4), *CHEK2* (n=4), *PALB2* (n=4), and *CDH1* (n=3), with germline confirmation rates ranging from 33-79%. Patients with confirmed germline mutations (GPV) were similar in age to patients with somatic-only mutations (SPV) (average 56±13 vs 54±15, p=0.72). The tumor allele frequency (AF) of GPV was modestly higher than SPV (54±16% vs 40±25%, p=0.04); however, the range of AF overlapped in both groups: GPV, 11-86% versus SPV, 2-79%. 46% of GPV and 80% of SPV were found in tumors known to be associated with mutations in that gene. In tumors with estimated tumor purity over 25% (16 of 28 tumors) probable loss of heterozygosity (LOH) was inferred in patients with GPV. LOH was suggested in six of nine (67%) of tumors known to be part of the phenotypic spectrum of the germline finding, but in only one of seven (14%) of tumors outside this group (p=0.06). For example, in confirmed germline *BRCA2* carriers, LOH was suggested by SPV AF>65% in two of three pancreatic cancers, one breast, and one of two prostate cancers. However, two glioblastomas, one cholangiocarcinoma and one colon cancer with confirmed GPV in *BRCA1* or *BRCA2* had the corresponding SPV AF<55%. **Conclusions:** Tumor variant AF, probable LOH, patient age, and personal cancer history are not sufficient to identify all potentially actionable germline findings from tNGS. Highly integrated pathways between genetic testing laboratories, oncologists, and cancer risk genetics professionals are needed to adequately triage potentially actionable germline findings that may be revealed by tumor-only genomic profiling.

713F

eQTL and co-expression network analyses uncover genetic and regulatory mechanisms underlying multiple myeloma. *H.M. Natrj^{1,2}, M. Wilson Sayres¹, K. Buetow¹, J. Keats², N. Banovich².* 1) Center for Evolution and Medicine, Arizona State University, Tempe, AZ; 2) Translational Genomics Research Institute, Phoenix, AZ.

Multiple myeloma (MM) is the second most common hematological cancer, accounting for 2% of all cancer deaths. MM remains incurable and nearly all patients experience relapse. To this end, The Multiple Myeloma Research Foundation developed the CoMMpass study. CoMMpass is a longitudinal clinical trial aimed at accelerating the discovery of more targeted treatments for MM. Each of the 1147 newly-diagnosed participants receives a proteasome inhibitor and/or immunomodulatory agent treatment regimen. Clinical parameters and tumor specimens are collected at baseline and through the eight-year observation period. To identify genetic determinants of clinical outcomes, each tumor specimen is characterized by Whole Genome, Exome and Transcriptome sequencing. This work has uncovered a number of somatic mutations and copy number alterations associated with tumor progression and response to therapy as well as a novel classification of MM into 12 distinct subtypes based on gene expression. However, the contribution of germline genetic variation to gene expression changes and clinical phenotypes remains poorly understood. Genome-wide association studies have identified germline variants associated with MM risk, indicating inherited genetic susceptibility. Men are in a higher risk to develop MM than women, and previous studies have reported loss of X or Y chromosomes in up to 40% of MM cases. To better understand the genetic and biological basis of MM predisposition, we utilize a systems genomics approach to examine inherited genetic effects on gene expression and MM subtypes. Data from 616 CoMMpass participants with available low coverage WGS of normal blood samples and RNAseq of baseline tumor specimens were used to map *cis*- and *trans*-acting eQTLs. Modules of co-expressed genes were constructed using a weighted correlation network analysis. Associations between eVariants and MM tumor subtypes were explored to detect genetic effects on gene expression and MM etiology. We have constructed a global transcriptional network connecting coding and regulatory germline variants to clinical phenotypes through regulation of gene expression, providing a comprehensive picture of genetic and regulatory mechanisms underlying MM susceptibility. These and further analyses integrating germline and somatic variants, gene expression, and clinical outcomes will support the development of personalized medicine approaches for better treatment of MM.

714W

Evaluation of a 94-gene inherited cancer panel across 248 families referred for testing to a reference molecular diagnostic laboratory. *H. Onay¹, A.E. Solmaz¹, L. Yeniyay², E. Gokmen³, T. Atik⁴, F. Ozkinay^{1,4}.* 1) Medical Genetics Dept, Ege University, Izmir, Turkey; 2) General Surgery Dept, Ege University, Izmir, Turkey; 3) Internal Medicine Dept, Ege University, Izmir, Turkey; 4) Pediatrics Dept, Ege University, Izmir, Turkey.

Cancer is one of the major health problems worldwide with a high rate of morbidity and mortality. Hereditary cancer syndromes, which show extensive clinical and genetic heterogeneity, account for approximately 5–10% of all cancer. Recently, next-generation sequencing genetic testing panels have become clinically available for hereditary cancer, helping to identify genetic etiology more efficiently. This study reports the experience of a reference molecular genetic center which used 94-gene inherited cancer panel on a selected group of 248 families with cancer or cancer susceptible diseases. **METHODS** Two hundred and forty eight families tested for cancer susceptibility genes were selected by experienced clinicians who were able to interpret the genetic counseling data in accordance with the latest literature. Illumina Trusight Cancer panel (including 94 cancer genes) was used as Multi-gene sequencing panel on the Illumina MiSeq platform. Genetic variations were identified using Illumina Variant Studio software and interpreted using the *in silico* prediction tools, such as: SIFT, PolyPhen Mutation Taster. **RESULTS** In 99 (39.9%) of 248 patients tested a mutation were detected in one of the cancer genes included in the panel. Of these, 106 were *BRCA1/2* mutation negative breast cancer patients. Among these 106 patients, 30 (28.3%) carried 31 different mutations in 14 different panel genes. Of the 28 colon cancer patients 22(75%) and of the 6 stomach cancer patients 3 (50%) found to have a mutation in 1 one of 95 genes investigated. In the study twenty six patients were referred for Fanconi molecular analysis and responsible mutations were defined in 16 (61.5%) of them. In 21 patients, the clinical diagnosis was neurofibromatosis and 16 (76.1%) of them had mutations in *NF1* gene. Four (57.1%) of 7 patients analyzed for tuberous sclerosis had a mutation in either *TSC1* or *TSC2* gene. Molecular defects were also detected in two patients with Gorlin syndrome, 1 over cancer patient, 1 GIST patient, 1 retinoblastoma patient, 1 Werner syndrome patient, 1 Bloom syndrome patient and 1 ataxi telangiectasia patient. **CONCLUSION** As a conclusion using 94 cancer gene panel (Trusight Cancer panel by Illumina), molecular diagnosis can be achieved in considerable number of families with inherited cancer or cancer susceptible diseases.

715T

Putative non-coding cis-regulatory drivers in chronic lymphocytic leukaemia and skin cancer. H. Ongen^{1,2,3}, O. Delaneau^{1,2,3}, N.M.R. Lykoskoufis^{1,2,3}, A. Real^{1,2,3}, C. Borel^{1,2,3}, M.W. Stevens^{1,2,3}, C. Howald^{1,2,3}, E.T. Dermizakis^{1,2,3}.

1) Department of Genetic Medicine and Development, University of Geneva Medical School, 1211 Geneva, Switzerland; 2) Institute for Genetics and Genomics in Geneva (iGE3), University of Geneva, 1211 Geneva, Switzerland; 3) Swiss Institute of Bioinformatics, 1211 Geneva, Switzerland.

Perturbations of the coding genome and their role in cancer development have been studied extensively. However, the non-coding genome's contribution in cancer is poorly understood, because it is difficult to define the non-coding regulatory regions and the genes they regulate, and because there is limited power owing to the regulatory regions' small size. We try to resolve this by defining modules of coordinated non-coding regulatory regions of genes (Cis Regulatory Domains or CRDs) by using the correlation between histone modifications, assayed by ChIP-seq, in population samples of immortalized B-cells and skin fibroblasts. We screen for CRDs that accumulate an excess of somatic mutations in chronic lymphocytic leukaemia (CLL) and skin cancer, after accounting for somatic mutational patterns and biases. At 5% FDR, we find 90 CRDs with significant excess somatic mutations in CLL, 60 of which regulate 126 genes, and in skin cancer 59 significant CRDs, 25 of which regulate 37 genes. The genes identified include ones already implicated in tumorigenesis, and are enriched in pathways involved in the respective cancers, like the BCR signalling pathway in CLL and the TGF β signalling pathway in skin cancer. We discover that the somatic mutations in the significant CRDs of CLL are hitting bases more likely to be functional than the mutations in non-significant CRDs. Moreover, in both cancers, mutational signatures observed in the regulatory regions of significant CRDs deviate significantly from their null sequences. Both results indicate selection acting on CRDs during tumorigenesis. We find that the transcription factor binding sites that are disturbed by the somatic mutations in significant CRDs are enriched for factors known to be involved in cancer development. Previously we detected genes with putative somatic regulatory drivers in colorectal cancer using perturbations in allele specific expression (Ongen et. al. Nature 2014). We are therefore currently testing the effect of mutations in the CRDs on the expression and the allelic expression of the genes implicated to further confirm their driver potential. Finally, we are conducting cell culture experiments to show the driver potential of the genes identified. This study describes a new powerful approach to discover non-coding regions involved in tumorigenesis in CLL and skin cancer and this approach could be generalized to other cancers.

716F

Mutational patterns and novel mutations in a large cohort of parathyroid carcinomas. C. Pandya^{1,2}, A.V. Uzilov^{1,2}, J. Bellizzi³, P. Taik^{1,2}, A.S. Moe¹, M. Strahl⁴, K.V.D. Tuin⁴, M. Stevenson⁵, B.M. Cavaco⁶, F. Cetani⁷, C. Marcocci⁸, A. Karhu⁹, J. Arola⁹, C. Schalin-Jäntti¹⁰, H. Morreau⁴, R.V. Thakker⁸, E.E. Schadt^{1,2}, R. Sebra^{1,2}, S.D. Li^{1,2}, A. Arnold⁸, R. Chen^{1,2}.

1) Sema4 Genomics (a Mount Sinai venture), Stamford, CT; 2) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 3) Center for Molecular Medicine, University of Connecticut School of Medicine, Farmington, CT; 4) Department of Pathology, Leiden University Medical Center, Leiden, The Netherlands; 5) Academic Endocrine Unit, Radcliffe Department of Medicine, Oxford Centre for Diabetes, Endocrinology and Metabolism (OCDEM), University of Oxford, Oxford UK; 6) Molecular Endocrinology Group, Molecular Pathobiology Research Centre Unit (UIPM) of the Portuguese Institute of Oncology from Lisbon Francisco Gentil (IPOLFG), Lisbon, Portugal; 7) University Hospital of Pisa, Endocrine Unit, Pisa, Italy; 8) University of Helsinki, Genome-Scale Biology Research Program, Research Programs Unit, Helsinki, Finland; 9) HUSLAB, Helsinki Biobank, Department of Pathology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland; 10) Endocrinology, Abdominal Center, University of Helsinki and Helsinki University Hospital, Helsinki, Finland.

Lack of effective targeted therapies make parathyroid carcinoma (PC), a rare and diagnostically challenging malignancy, difficult to combat. Typically, if the cancer is not cured by early surgery, elevated parathyroid hormone levels can lead to severe hypercalcemia and ultimately fatal end-organ damage. As the molecular mechanisms of oncogenesis in PC are poorly understood, we established an international collaboration to gain further insights that could lead to improved clinical outcomes. Initial findings on genomic profiles of 24 PC cases from these collective efforts confirmed the importance of driver mutations in *CDC73* (the only established tumor suppressor in PC) and identified several putative drivers and aberrant pathways (Pandya et al., JCI Insight 2017; 2:e92061). To pursue these findings with additional power, we have now expanded our cohort of collected tumor (primary, local recurrence, or metastatic) and normal DNA samples to 73 patients with sporadic PC fulfilling stringent selection by WHO criteria. We used a PC-informed, custom Ion AmpliSeq 16-gene targeted panel to generate DNA sequence data for identifying somatic and germline genetic alterations. We observed 47 samples harboring ≥ 1 variant and 24 samples with no variant, corroborating previous findings of a moderate tumor mutation burden in PC. Consistent with earlier studies, we found 41% (30/73) of the patients to harbor loss-of-function (LOF) somatic *CDC73* mutations, with germline inactivating variants in 4 patients. Notably, the PI3K/AKT/mTOR pathway was altered in 12.3% of the 73 cases with known oncogenic hotspots in *PIK3CA* (p. K111E, p.E545K/A & p. H1047R), primarily in tumors without *CDC73* mutations, thereby reinforcing our published novel finding and consideration of this as a clinically actionable pathway in PC. Correlating the genomic and clinical features revealed that patients harboring a *CDC73* LOF mutation tend to have early-onset PC (P-value=0.0002; odds-ratio=7.4; adjusted P-value=0.003) and are likely to suffer from life threatening recurrent/metastatic PC (P-value=0.02; odds-ratio=4.2; adjusted P-value=0.29). Our study, which is the largest cohort of PC samples analyzed at a molecular level to date, facilitates improved understanding of the genomic underpinnings of this rare malignancy and helps to provide a rational basis for individualized treatments.

717W

Segmental odonto-maxillary dysplasia: A mosaic disorder due to missense mutations in *ACTB*. S. Polubothu^{1,2}, A. Thomas¹, R. Evans³, A. Calder⁴, V.A. Kinsler^{1,2}. 1) Genetics and Genomic Medicine, University College London Institute of Child Health, London, London, United Kingdom; 2) Paediatric Dermatology, Great Ormond St Hospital for Children, London; 3) Maxillofacial Surgery, Great Ormond St Hospital for Children, London; 4) Paediatric Radiology, Great Ormond St Hospital for Children, London.

Segmental odontomaxillary dysplasia (SOD) is a rare sporadic disorder of unknown aetiology presenting at birth or in childhood with unilateral abnormal maxillary development. This condition is well-known to maxillo-facial and dental specialists, but less well known in other specialties. Affected individuals usually present with colocalised abnormalities such as facial asymmetry, progressive hypertrichosis, increased or decreased pigmentation or erythema, dental anomalies and commissural lip fissures. Here we present detailed phenotypic, histological and radiological data from a cohort of eight patients, and investigate the likely mosaic basis of the disorder in three. Sequencing of DNA extracted directly from paired samples of affected skin and blood demonstrated missense variants in the gene encoding beta-actin (*ACTB*) at mosaic levels in affected tissue but undetectable in blood in all three. Germline GOF and LOF *ACTB* variants cause two distinct inherited disorders, and mosaic *ACTB* variants have previously been identified as the cause of a proportion of Becker's naevi. The exact mechanism of beta-actin action however are not clear. These findings uncover the cause of SOD, and suggest that beta-actin may be intrinsically involved in cross-talk between different colocalised embryonic tissues, as a single cell mutation during embryogenesis could not otherwise explain bony, soft tissue, and skin pathology. In addition, this finding highlights this clinical diagnosis to Clinical Geneticists and Paediatric Dermatologists, where we estimate the disorder is poorly recognised given the frequency in our department.

718T

Whole-exome sequencing identified mutational profiles of squamous cell carcinomas of anus. S. Shin^{1,2}, H-C. Park^{1,2}, M.S. Kim^{3,4}, S.H. Jung^{2,3}, Y-J. Chung^{1,2,3}, S.H. Lee^{3,4}. 1) Department of Microbiology, The Catholic University of Korea, Seoul, Republic of Korea; 2) Integrated Research Center for Genome Polymorphism, The Catholic University of Korea, Seoul, Republic of Korea; 3) Cancer Evolution Research Center, The Catholic University of Korea, Seoul, Republic of Korea; 4) Department of Pathology, The Catholic University of Korea, Seoul, Republic of Korea.

Anal squamous cell carcinoma (ASCC), either with human papillomavirus (HPV) (+) or (-), is a neoplastic disease with frequent recurrence and metastasis. To characterize ASCC genomes, we attempted to disclose novel alterations of ASCC genomes as well as other genetic features including mutation signatures. We performed whole-exome sequencing and copy number alteration (CNA) profiling for 8 ASCC samples from 6 patients (2 cases with primary and recurrent/metastatic tumors). We found known ASCC mutations (*TP53*, *CDKN2A* and *PIK3CA*) and CNAs (gains on 3q and 19q and losses on 11q and 13q). In addition, we discovered novel mutations in *HRAS* and *ARID1A* and CNAs (gain on 8q and losses 5q, 9p, 10q and 19p) that had not been reported in ASCCs. We identified 4 signature patterns of the mutations (signatures 1 and 2 with deamination of 5-methyl-cytosine, signature 3 with APOBEC and signature 4 with mismatch repair) in the ASCCs. While the signatures 1-3 have been detected in other SCCs, the signature 4 was first identified in ASCCs. In addition, we first found that ASCCs harbored chromothripsis, copy-neutral losses of heterozygosity and focal amplification of *KLF5* super-enhancer. Analyses of primary and recurrent/metastatic pair genomes revealed that driver events in development and progression of ASCC might not be uniform. Our data indicate that ASCCs may have similar mutation and CNA profiles to other SCCs, but that there are unique genomic features of ASCCs as well. Our data may provide useful information for ASCC pathogenesis as well as for developing clinical strategies for ASCC.

719F

Whole exome sequencing to identify genetic markers for trastuzumab-induced cardiotoxicity. C. Udagawa^{1,2,3}, H. Nakamura⁴, K. Tamura⁵, T. Shimoi⁶, M. Yoshida⁶, T. Yoshida⁷, H. Okamura³, Y. Totoki⁴, T. Shibata⁴, H. Zembutsu^{1,2}.

1) Liquid Biopsy Development Group, Cancer Precision Medicine Center, Japanese Foundation for Cancer Research, Tokyo, Japan; 2) Division of Genetics, National Cancer Center Research Institute, Tokyo, Japan; 3) Toyo Kohan Co., Ltd., Tokyo, Japan; 4) Division of Cancer Genomics, National Cancer Center Research Institute, Tokyo, Japan; 5) Department of Breast and Medical Oncology, National Cancer Center Hospital, Tokyo, Japan; 6) Department of Pathology and Clinical Laboratories, National Cancer Center Hospital, Tokyo, Japan; 7) Fundamental Innovative Oncology Core, National Cancer Center Research Institute, Tokyo, Japan.

Trastuzumab is used for the treatment in human epidermal growth factor receptor 2 (HER2) positive breast or gastric cancer patients. However, it is reported in many studies that approximately 5% of the patients treated with trastuzumab alone cause cardiac dysfunction. To identify a genetic marker(s) determining the risk of trastuzumab-induced cardiotoxicity, we performed whole exome sequencing of germline DNA samples from 9 patients with trastuzumab-induced cardiotoxicity, and conducted a case-control association study of 2,258 genetic variants between 9 cases (with trastuzumab-induced cardiotoxicity) and general Japanese population controls registered in Human Genetic Variation Database (HGVD). To further validate the result of the screening study, we carried out a replication study of 10 variants showing $P_{\text{min}} < 0.001$ in the screening study using 10 cases and 224 controls (without trastuzumab-induced cardiotoxicity). In the replication study, we observed that three variants had effect in the same direction as in the screening study (SNV1 in exon 2 of *HERADR1*, SNV2 in exon 2 of *HERADR2* and SNV3 in exon 44 of *HERLC*). A combined result of the screening and the replication studies suggested an association of a locus on chromosome 6q12 with trastuzumab-induced cardiotoxicity (SNV3 in *HERLC*, combined- $P_{\text{min}} = 0.00056$, OR = 13.73). *HERLC* protein have signal peptide, predicted cleavage site and multiple EGF-like domains. Moreover, *HERLC* is highly expressed in retina and pineal body according to the gene expression database. These lines of evidence suggest that *HERLC* is a secreted protein and might affect the efficiency of HER2 signal transduction as a ligand for HER2. Although further investigation is required to clarify the effect of *HERLC* on trastuzumab-induced cardiotoxicity, these findings provide new insights into personalized trastuzumab therapy.

720W

Whole-exome case-control association to characterize the contribution of rare coding variants to sarcoma susceptibility. Y. Yu, F. Hu, J. Chen, P. Scheet, C. Huff. UT MD Anderson Cancer Center, Houston, TX.

Genetic susceptibility has a well-established role in the etiology of sarcoma, but the contribution and effect size of rare genetic variants to sarcoma risk is underexplored. To evaluate the contribution of rare, protein-coding variation to sarcoma risk, we conducted a whole-exome case-control study of individuals with European ancestry involving 219 cases from the Cancer Genome Atlas (TCGA) and a population control set of 3,507 individuals from NDAR. We used XPAT to control for technological stratification biases from heterogeneous platforms. We then used VAAST 2.1 to conduct gene-based association tests for all samples and sub-groups of six soft-tissue sarcoma subtypes: leiomyosarcoma (LMS), dedifferentiated liposarcoma (DLPS), myxofibrosarcoma (MFS), undifferentiated pleomorphic sarcoma (UPS), malignant peripheral nerve sheath tumors (MPNST), and synovial sarcoma (SS). We observed nominally significant association signals for three well-established sarcoma susceptibility genes, *NF1* ($p=1 \times 10^{-5}$), *TP53* ($p=0.0025$), and *RB1* ($p=0.028$). *MSH2*, an HNPCC gene, was also nominally significant ($p=0.0085$); previous studies have suggested that sarcomas may be within the spectrum of HNPCC tumors. We also estimated the effect size estimates for variants in these. The likely-gene disrupting (LGD) variants in these genes exhibit large effect size to sarcoma relative to other cancer types. The ORs of *NF1* and *MSH2* LGD variants were 39.4 (95% CI: 7.1 – 219.3) and 33.0 (95% CI: 2.4 – 462.5), respectively. We observed predicted damaging missense variants with large effect sizes in *RB1* and *TP53* genes, which ORs were 12.0 (95% CI: 2.4 – 59.1) and ∞ (95% CI: 6.7 – ∞), respectively. In subtype-specific analyses, we found *NF1* was significantly associated with MPNST and LMS, *RB1* and *MSH2* were significantly associated with LMS, and *TP53* were significantly associated with UPS. Our results indicate that rare coding variants can confer substantial increases in risk for multiple soft-tissue sarcoma subtypes. In conclusion, our results confirm previously established sarcoma susceptibility loci and provide a foundation for future rare-variant association studies in sarcoma.

721T

Integrating of multi-omics data for the prognostic assessment of breast cancer. C. Yu^{1,2}, Y. Jiang^{1,2}, J. Dai^{1,2}, H. Ma^{1,2}. 1) Department of Epidemiology and Biostatistics, Nanjing Medical University, Nanjing, Jiangsu, China; 2) Jiangsu Key Lab of Cancer Biomarkers, Prevention and Treatment, Collaborative Innovation Center for Cancer Medicine, Nanjing Medical University, Nanjing, Jiangsu, China.

Background: Breast cancer is the most diagnosed cancer and the leading cause of cancer death among females, and both genetic variants and genes expression have been confirmed to be effective in the prediction of the prognosis of breast cancers. However, only partially breast cancer survival statuses can be explained, suggesting that integrating these factors may contribute the prediction of the prognosis of breast cancer. **Methods:** 810 samples from TCGA data sets are randomly divided into training set (540 subjects) and validation set (270 subjects). We first selected single nucleotide polymorphism (SNPs) and genes associated with breast cancer prognosis in the training set to construct the prognostic prediction model, and then replicated the prediction efficiency in the validation set. Cox proportional hazards models were performed to evaluate the prognostic effect of SNPs and genes. Prognostic index was calculated using HR (Hazard Ratio) weighted method. **Results:** We identified 4 SNPs and 3 genes that associated with the prognosis of breast cancers in the training set, and were incorporated into the prognostic model. The high-risk patients defined by the 4 SNPs and 3 genes signature-based prognostic index, had a statistically significantly worse overall survival (HR=9.43, 95% CI= 3.81 to 23.33, $P < 0.001$), and the C statistics of ROC of the signature-based prognostic index was 0.79. In the validation set, patients in the high-risk group had a significantly worse overall survival (HR= 4.55, 95% CI= 1.50 to 13.88, $P < 0.001$), and the C statistics of ROC of the signature-based prognostic index was 0.76. **Conclusions:** Our findings suggested that integrating multiple omics data could greatly increase the predictive performance of the prognosis of breast cancer survival. This strategy may be useful in identify poor prognosis risk populations for targeted cancer treatment.

722F

Cancer-associated mutations in colorectal adenomatous polyps. C. Zeng¹, M.J. Shrubsole¹, H.J. Murff^{2,3}, Q. Cai¹, Q. Liu³, Q. Sheng³, A. Madan¹, R.J.C. Coffey^{1,2}, W. Zheng¹. 1) Department of Medicine, Division of Epidemiology, Vanderbilt Epidemiology, Vanderbilt University Medical Center, Nashville, TN; 2) Department of Veterans Affairs, Tennessee, Valley Healthcare System, Nashville, TN; 3) Center for Quantitative Science, Vanderbilt University, Nashville, TN; 4) Covance Genomics Lab, Redmond, WA.

Background Few studies examined the associations of somatic mutation profiles with histologic characteristics of sporadic colorectal adenomatous polyps and risk of adenoma recurrence. We evaluated these associations in a study including 200 patients with sporadic adenomas. **Methods** Study participants were recruited as part of the Tennessee Colorectal Polyp Study from patients who received a colonoscopy at the Vanderbilt University Medical Center and Tennessee Valley Healthcare System Nashville Campus. The 200 participants were selected from among those for whom formalin-fixed paraffin-embedded (FFPE) blocks were available for DNA extraction and had at least one subsequent surveillance colonoscopy. Among these participants, 100 of them had synchronous small (<1 cm) tubular adenomas and the remaining had at least one advanced adenoma (≥ 1 cm or tubulovillous/villous or high-grade dysplasia) with or without the presence of synchronous adenoma. In addition, within each of these subgroups, 50 developed ≥ 1 metachronous adenoma and 50 had no future adenomas. For those with synchronous adenomas, the single adenoma that was sequenced was selected in the following order: most dysplastic, most villous, and largest in size. Targeted next generation sequencing for 728 genes was performed to detect somatic mutations in adenoma tissues obtained at the first diagnosis from these patients. We limited our analysis to potential colorectal cancer (CRC) driver genes ($n=19$) and potential driver mutations ($n=65$) proposed by previous studies in CRC. **Results** There were considerable inter-person variations in mutational profiles. No single mutation was statistically significantly associated with advanced features of adenomas or risk of recurrence ($P > 0.01$). Patients with APC gene mutations were more likely to have advanced adenomas than those without APC mutations (odds ratio (OR) = 4.3, $P=0.004$). Patients who carried potential driver mutations were more likely to have advanced adenomas (OR for carrying each potential driver mutation = 2.4, $P=0.0002$) and had a higher risk of adenoma recurrence (OR for carrying each potential driver mutation = 1.7, $P=0.02$), comparing with those who did not. **Conclusions** These data suggested that somatic mutations in cancer-associated genes could help identify adenoma patients at a higher risk of recurrence and facilitate personalized surveillance and prevention of CRC.

723W

Clinical presentation of individuals with mosaic APC variants. M.E.

Roberts, E.M. Webb, M.M. Marshall, L.R. Susswein, Y. Wang, R.T. Klein, K.S. Hruska. GeneDx, Gaithersburg, MD.

Background: Familial adenomatous polyposis (FAP) is characterized by the presence of many adenomatous polyps and a high risk for colorectal cancer (CRC). Individuals with an attenuated phenotype (AFAP) tend to have fewer polyps, later ages of onset, and/or lower risks for CRC. Pathogenic/likely pathogenic variants (PVs) in *APC* are known to cause FAP/AFAP, yet many cases of polyposis remain unexplained. Mosaic *APC* PVs have been reported to cause a subset of unexplained FAP cases, with mosaicism confirmed in patient leukocytes, lymphocytes, and adenomas. In a large published polyposis cohort, 4% of individuals with a PV in *APC* on germline testing were identified to be mosaic. In a subset of these cases, mosaicism was detected in adenomatous tissue rather than peripheral blood specimens. **Purpose:** We aim to describe the clinical presentation of a series of individuals with mosaic *APC* PVs identified at our clinical diagnostic laboratory. **Methods:** We performed a retrospective review of clinical data from test requisition forms for individuals found to carry a mosaic *APC* PV via multi-gene hereditary cancer panel testing. Testing was performed in peripheral blood and/or fibroblasts, but not in adenomatous polyps/CRCs/other affected tissues. **Results:** In total, 5/289 (1.7%) probands with *APC* PVs were found to be mosaic in peripheral blood. None of these five individuals carried a second PV in *APC* or in another gene and the average age of testing was 48.7yrs (38-55). All individuals identified to have a mosaic PV in *APC* reported phenotypes consistent with AFAP/FAP, with 38 to >100 polyps or CRC reported. Of the five individuals, one reported CRC only, two reported CRC and polyps, and one reported only polyps. Only two individuals reported a family history of CRC or polyps. Two of five individuals had subsequent testing in fibroblasts or targeted variant testing in a child, in which the *APC* PV was also identified, suggesting constitutional mosaicism. **Conclusion:** These data support that mosaic *APC* PVs may explain a subset of unexplained polyposis cases. The presence of phenotypes consistent with AFAP/FAP and supportive fibroblast and/or familial testing results in our cohort suggest that the mosaic PVs may have contributed to polyposis and/or CRC development.

724T

Using low coverage and off target sequencing for polygenic risk scores and ancestry. J.R. Homburger, C.L. Neben, C. Lai, W. Stedden, A. Zimmer, A.Y. Zhou, G. Mishne. Color Genomics, Burlingame, CA., USA.

Next generation sequencing (NGS) is an integral part of clinical care and management with applications such as targeted multi-gene panels for monogenic diseases. There is an emerging consensus that polygenic risk scores (PRS) also have validity and utility in stratifying disease risk, however, several barriers exist to implementing PRS into clinical practice. For example, PRS has traditionally been performed on genotyping arrays, and therefore assessment for polygenic risk in addition to monogenic risk currently adds additional tests, time, and cost. Recent advancements in NGS have made it possible to provide cost effective assessment of common genetic variation using low coverage sequencing while still maintaining the computational power of a genotyping array. This enables us to develop a single cost-effective assay that combines clinical grade targeted sequencing with low-coverage whole genome sequencing (lc-WGS) to determine monogenic and polygenic risk at the same time. Here, we assess the feasibility of using lc-WGS for genotype imputation and downstream applications including PRS and ancestry. First, similar to previously published studies, we used known control samples to show that 0.5x coverage achieves average imputation r^2 of > 0.9 of imputed dosages to true genotypes at common genetic variants ($MAF > 0.05$), demonstrating non-inferiority to array-based imputation. Second, we created a method for low coverage PRS, which uses imputed genotype dosages from lc-WGS data as inputs to polygenic risk algorithms developed from summary statistics using the LDpred approach. We found that our low coverage PRS, which combines both a high-coverage targeted panel and off target reads with a mean coverage of 0.3x, is a significant predictor of self-reported breast cancer (OR 1.43 per score standard deviation, $p < 10^{-16}$) after adjusting for age, gender, family history, and age at menarche. Third, we found that lc-WGS can be used for accurate continental and subcontinental ancestry determination, an important covariate in current polygenic risk approaches. In conclusion, we demonstrate that at certain coverages, lc-WGS is equivalent to genotyping arrays and may be an attractive option for PRS and ancestry assessment. Importantly, the accuracy of lc-WGS-based methods is highly dependent upon coverage and trait genetic architecture, indicating that strict quality metrics will need to be developed before clinical implementation.

725F

Characterizing known risk loci for colorectal cancer in African Americans. *H. Wang¹, C. Haiman², T. Burnett¹, I. Kato³, T. Keku⁴, W. Blot⁵, M. Pande⁶, K. Zanetti⁷, C. Harris⁸, J. Palmer⁹, S. Berndt¹⁰, P. Newcomb¹¹, D. West¹², L. Wilkens¹, R. Haile¹³, D. Stram², L. Le Marchand¹.* 1) University of Hawaii Cancer Center, Honolulu, HI; 2) Department of Preventive Medicine, Keck School of Medicine and Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA; 3) Wayne State University Karmanos Cancer Institute, Detroit, MI; 4) Center for Gastrointestinal Biology and Disease, University of North Carolina, Chapel Hill, NC; 5) Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center and Vanderbilt-Ingram Cancer Center, Vanderbilt University, Nashville, TN; 6) Department of Gastroenterology, Hepatology & Nutrition, the University of Texas M. D. Anderson Cancer Center, Houston, TX; 7) Division of Cancer Control and Population Sciences, National Cancer Institute, National Institutes of Health, Bethesda, MD; 8) Laboratory of Human Carcinogenesis, National Cancer Institute, National Institutes of Health, Bethesda, MD; 9) Slone Epidemiology Center at Boston University, Boston, MA; 10) Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD; 11) Fred Hutchinson Cancer Research Center, Seattle, WA; 12) Cancer Prevention Institute of California, Fremont, CA; 13) Cedars-Sinai Medical Center, Los Angeles, CA.

Genome-wide association studies of colorectal cancer, mostly conducted in Europeans and East Asians, have identified over 70 risk SNPs in 50 susceptibility regions. Re-assessment of these susceptibility regions in African Americans with a more diverse haplotype background can help identifying the best risk-defining variants and understanding the risk predicting ability of these known variants in this population. We previously performed a fine-mapping study for 21 of the risk regions (250kb on both sides of the index SNPs) using 6,597 African Americans (1,894 cases and 4,703 controls) imputed into the 1000 Genomes Project (phase 1). Here, we are updating the results with an additional 1,609 sample from 4 participating studies (652 cases, 957 controls) genotyped on OncoArray BeadChip and imputed into the Consortium on Asthma among African-ancestry Populations in the Americas (CAAPA) haplotype reference panel with 31 million markers. In addition, we will quantify the risk increment associated with a polygenic risk score, summarizing the effects of all known variants, and estimate the proportion of the familial colorectal cancer risk attributable to these SNPs. These data represent the largest genome-wide association study on colorectal cancer in African Americans and the findings would provide valuable insight into the genetic architecture of this complex disease.

726W

Co-localization of chromatin QTLs and expression QTLs to establish functional mechanisms at ovarian cancer risk loci. *M.K. Freund¹, J. Plummer², F. Abbas³, N. Mancuso⁴, C. Giambartolomei⁴, M. Roytman⁵, B. Davis², P. Pharoah⁶, A. Gusev⁷, B. Pasaniuc^{1,4,5}, S.A. Gayther², K. Lawrenson^{2,3}.* 1) Department of Human Genetics, David Geffen School of Medicine, UCLA, Los Angeles, CA; 2) Center for Bioinformatics and Functional Genomics, Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA; 3) Program at the Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA; 4) Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA; 5) Bioinformatics Interdepartmental Program, University of California Los Angeles, Los Angeles, CA; 6) Department of Public Health and Primary Care, Department of Oncology, University of Cambridge, UK; 7) Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA.

Epithelial ovarian cancer is a rare but highly lethal malignancy, expected to cause ~14,000 deaths in the US in 2018. Though recent studies have provided insight into the mechanisms of ovarian cancer risk, the specific regulatory pathways in ovarian surface epithelial cells, an ovarian cancer precursor cell type, are still largely unknown. Here, we examine the largest ovarian cancer precursor cell dataset to date, comprising H3K27ac ChIP-sequencing, RNA-sequencing, and GWAS genotyping data in primary normal ovarian surface epithelial cells from 54 patients. We first identify novel tissue-relevant H3K27ac chromatin QTLs (cQTLs) and expression QTLs (eQTLs). In combination with transcriptome wide association scans to identify candidate target genes associated with ovarian cancer risk loci, we then identify ovarian cancer risk regions with eQTL associations, cQTL associations, and regions where both eQTL and cQTL associations are co-localized. Preliminary results identify 29 ovarian cancer risk regions with significant cQTL associations. In parallel, HiChIP analysis in ovarian normal and cancer cell lines to characterize the 'interactome' in these tissues is ongoing, to validate the putative eQTL-cQTL interactions and discover additional novel transcript-histone modification interactions. In addition to providing insight into the regulatory landscape of healthy ovarian surface epithelial cells and identifying putative tissue-specific ovarian cancer pathways, these results can be contrasted to similar QTL analyses in ovarian tumors to pinpoint the major changes in gene regulation and pathway activation that occur during neoplastic transformation.

727T

Rare disruptive coding variants and prostate cancer risk in men of African ancestry. M. Matejčić¹, Y. Patel¹, L. Lilyquist², F.J. Couch^{2,3}, D.V. Conti¹, C.A. Haiman¹. 1) Department of Preventive Medicine, University of Southern California, Los Angeles, CA; 2) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota; 3) Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota.

We explored the contribution of rare coding variation to risk of PCa in men of African ancestry by sequencing a panel of 19 DNA repair and known cancer susceptibility genes in 2,098 cases and 1,481 controls of African ancestry (2,442 African American men and 1,137 Ugandan men). Variants with minor allele frequency <1% were annotated based on their functional consequence and pathogenic effect according to CAVA, ClinVar and five prediction algorithms (Polyphen2-HumDiv, PolyPhen2-HumVar, LRT, Mutation Taster, and SIFT). Analyses were performed at single variant and gene burden level to test individual and aggregate effects of rare coding variation on overall PCa risk as well as by disease aggressiveness. We identified 85 loss-of-function (Tier 1: frameshift insertion/deletions, stop gain and essential splice site changes) and 600 missense (Tier 2: non-synonymous and exon start/end codon changes) variants meeting the criteria for being deleterious. Of these, 153 (22.3%) were not previously reported in either gnomAD or ClinVar. The strongest single variant association with PCa risk was observed for three missense (Tier 2) variants in *MSH2* (Gln170Glu, $p = 0.00308$), *MRE11A* (Leu7Phe, $p = 0.00363$) and *ATM* (Glu1991Gly, $p = 0.0078$), which were replicated in Ugandans ($p = 0.01279$ for Gln170Glu, 0.02082 for Leu7Phe and 0.0156 for Glu1991Gly). In gene-burden testing with Tier 1 variants, significant associations were observed for *BRCA2* ($p = 0.00196$), *ATM* ($p = 0.0282$) and *PALB2* ($p = 0.03395$) in the overall study population and for *BRCA2* ($p = 0.00364$), *ATM* ($p = 0.02495$) and *MSH6* ($p = 0.03027$) in Ugandans. These associations were stronger in high-aggressive cases ($n = 988$) compared to controls ($p = 0.00017$ for *BRCA2*, 0.01749 for *PALB2*, 0.03224 for *ATM*). In Tier 2 gene-burden testing, *PTEN* was significantly associated with PCa risk in the overall ($p = 0.01326$) and African American ($p = 0.01522$) populations. The aggregation of Tier 1 variants by DNA-repair pathways in the overall study population showed a significantly increased risk associated with genes involved in the Fanconi anemia ($p = 0.00168$) and homologous recombination ($p = 0.00438$) signalling networks, with a stronger effect size in Ugandans ($p = 0.00221$ and 0.0005 , respectively). Our findings suggest that rare coding variation may have appreciable effects on disease risk in men of African ancestry, especially in Ugandans, which highlights the importance of sequencing in diverse populations for a more efficient screening of high-risk patients. .

728F

GWAS identifies the first risk loci for gastroesophageal reflux disease, with most influencing risk of Barrett's esophagus and/or esophageal adenocarcinoma. J. An¹, P. Gharahkhani¹, M.H. Law¹, J.S. Ong¹, X. Han¹, I. Tomlinson², T.L. Vaughan³, M.F. Buas^{3,4}, R.C. Fitzgerald⁵, J. Schumacher⁶, C. Palles⁷, D.C. Whiteman⁷, S. MacGregor¹. 1) Statistical Genetics, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia; 2) Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham, UK; 3) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA; 4) Department of Cancer Prevention and Control, Roswell Park Cancer Institute, Buffalo, New York; 5) Medical Research Council (MRC) Cancer Unit, Hutchison-MRC Research Centre and University of Cambridge, Cambridge, UK; 6) Institute of Human Genetics, and Department of Genomics, Life & Brain Center, University of Bonn, Bonn, Germany; 7) Cancer Control, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia.

Gastroesophageal reflux disease (GERD) is a very common chronic disease where stomach acid flows back up into the esophagus. GERD causes mucosal damage and complications in the upper digestive system and is associated with Barrett's esophagus (BE) and esophageal adenocarcinoma (EA). We conducted a case-control GWAS in UK Biobank with cases defined in terms of a) GERD status or b) GERD medication history. Results were replicated in an independent case-control cohort (QSKIN - Australia). Concordance in terms of GWAS effect sizes between GERD status and GERD medication history was very high, justifying merging to maximize power. Taken together, 72,334 cases and 263,352 controls had genome-wide data. We identified 28 independent loci for GERD, the first genome-wide significant loci to be identified for this condition. We then investigated the effect of these GERD loci in GWAS data on Barrett's esophagus (BE; 6,167 cases, 17,159 controls) and esophageal carcinoma (EA; 4,112 cases, 13,663 controls). The vast majority of the trait increasing alleles at the GERD loci also increased risk of BE and/or EA. Three of the 28 GERD loci overlapped with those previously implicated in BE and/or EA at $P < 5e-8$. The majority of the loci implicated in our GERD analysis and showed $P < 0.05$ for BE and/or EA, including SNPs in our HIST1H3J, ZSCAN12 and SLC9A3. Finally, using LD-Score regression, we identified strong genetic correlations between GERD and BMI, and GERD and depression.

729W

Genome-wide interaction analysis of menopausal hormone therapy and breast cancer risk. X. Wang^{1,2}, P. Middha³, P. Auer⁴, R. Milne^{5,6}, P. Kraft^{7,8}, J. Chang-Claude³, S. Lindström^{1,2}, Breast Cancer Association Consortium (BCAC).

1) Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA; 2) Department of Epidemiology, University of Washington, Seattle, WA; 3) Division of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany; 4) Department of Biostatistics, School of Public Health, University of Wisconsin Milwaukee, Milwaukee, WI; 5) Cancer Epidemiology and Intelligence Division, Cancer Council Victoria, Melbourne, Vic, Australia; 6) Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, the University of Melbourne, Melbourne, Vic, Australia; 7) Department of Epidemiology, Harvard Chan T. School of Public Health, Boston, MA; 8) Department of Biostatistics, Harvard Chan T. School of Public Health, Boston, MA.

Background: Use of menopausal hormone therapy (MHT) has been associated with higher risk for breast cancer. However, the relevant mechanisms and its interaction with genetic variants are not fully understood. **Methods:** We conducted a genome-wide interaction analysis between current MHT use and genetic variants for breast cancer risk among 20,872 cases and 26,550 controls from 21 observational studies. All women were post-menopausal and of European ancestry. Multivariate logistic regression models were used to test for multiplicative interactions between SNP and current MHT, adjusting for main effects of SNPs and current MHT use, age, study, former use of MHT and principal components. We considered interaction p-values $< 5 \times 10^{-8}$ as genome-wide significant. **Results:** Among the total of ~10 million SNPs tested for SNP-MHT interactions, none reached genome-wide significance level. 33 SNPs had p-values $< 1 \times 10^{-5}$. The most significant SNP rs190966839 (p-value: 1.13×10^{-5}) was close to the tumor suppressor candidate 1 (*TUSC1*) gene, on chromosome 9. **Conclusions:** In this genome-wide SNP-MHT interaction study of breast cancer, we identified suggestive variants that potentially interact with MHT use; however, no SNP reached genome-wide significance. Our next step is to replicate these results in similar-sized independent datasets that are currently under preparation.

730T

Whole genome sequencing identifies candidate germline variants leading to familial non-medullary thyroid cancer. Y. Wang¹, S. Liyanarachchi¹, K.E. Miller², T.T. Nieminen¹, D.F. Comiskey¹, W. Li¹, P. Brock³, H. He¹, D.C. Koboldt², A. de la Chapelle¹.

1) Comprehensive Cancer Center, The Ohio State University, Columbus, OH; 2) Institute for Genomic Medicine, Nationwide Children's Hospital, Columbus, OH; 3) Department of Internal Medicine, The Ohio State University Comprehensive Cancer Center, The Ohio State University, Columbus, OH.

Thyroid cancer is the most common malignancy of the endocrine system. Based on the data from National Cancer Institute for 2018, the estimated number of new thyroid cancer patients is 53,990 in the United States. It has been reported that familial non-medullary thyroid cancer (FNMTc) demonstrates more aggressive features than sporadic non-medullary thyroid cancer. FNMTc constitutes only 3-9% of all thyroid cancers. Its predisposition pattern is complex and most likely due to multiple mutations. To identify rare pathogenic variants in non-medullary thyroid cancer (NMTC) families, here we performed whole genome sequencing (WGS) on 57 samples from 17 NMTC families. Blood DNA samples from at least 3 patients within each family were sequenced by Novogene using Illumina HiSeq platform. We also performed whole genome linkage analysis using Illumina HumanCytoSNP-12 BeadChip in a total of 106 samples including unaffected family members. Variants were analyzed using MendelScan v1.2.3. Candidate variants were selected based on the maximum allele frequency in gnomAD population data, functional prediction score using Varsome, and the evidence of constraint in ExAC database. Alternatively and additionally, to avoid missing important truncating variants caused by algorithm bias, variants from the same VCF file were further analyzed by using Baseplayer (1.0.0). A list of rare variants possibly causing FNMTc was generated including the variants identified by both methods. All these variants were shared by every WGS tested individuals within each family. We identified a total of 86 candidate variants, comprising 71 single nucleotide variants (49 missense, 12 nonsense, and 10 splice site variants) and 15 frameshift variants. Among all the variants, 51% (44/86) were novel and have not yet been reported in gnomAD database. Furthermore, Sanger sequencing validation within all family members with DNA available has confirmed complete cosegregation with NMTC in 11 of 26 variants tested so far. Functional validation experiments will be performed to prove the roles of these rare variants in FNMTc predisposition. Our study identified a group of rare germline variants that may play a putative role in creating familial thyroid cancer risk.

731F

Characterizing the spectrum of low-penetrance variants predisposing to papillary thyroid cancer. D.F. Comiskey, H. He, S. Liyanarachchi, L.K. Genu-tis, P. Brock, A. de la Chapelle. The Ohio State University, Columbus, OH.

Papillary thyroid cancer (PTC) arises from thyroid follicular epithelial cells and accounts for over 90% of all thyroid cancer diagnoses. PTC demonstrates high heritability and a low somatic mutation burden relative to other cancers. To understand the genetic basis of PTC many have taken to using both case-control studies and genome-wide association studies (GWAS) in order to identify low-penetrance germline variants. Recently, a meta-GWAS in PTC patients and controls of European descent yielded five new loci at 1q42, 3q26, 5q22, 10q24, and 15q22 genome-wide significantly associated with PTC. The underlying mechanisms of many of the PTC risk loci is unknown, however the sentinel single nucleotide polymorphisms (SNPs) are located adjacent to promoter and enhancer elements. Overall, our data support a hypothesis that SNPs in these loci alter the transcriptional activity of genes relevant to thyroid cancer pathogenesis. Here we report the characterization of three loci. SNP rs6793295 at 3q26 is a missense variant that lies within exon 7 of *LRRC34* and shows allele-specific expression in normal thyroid tissue. Mass spectroscopy of *LRRC34* indicates that this protein has functions related to protein folding, chromatin regulation, and cellular metabolism. We have also detected another region in the promoter of *LRRC34* that also shows association with PTC risk. SNP rs12129938 at 1q42 is located in a 16 kilobase haplotype in the first intron of *PCNXL2*. This block is a promoter-flanking region in the vicinity of several expression quantitative trait loci (cis-eQTL) for thyroid tissue. *PCNXL2* is modestly expressed in thyroid and its expression is two-fold higher in PTC tumor as compared to normal tissue. Finally, SNP rs73227482 at 5q22 lies within an intron of a non-coding transcript of *EPB41L4A*. *EPB41L4A* is highly expressed in thyroid and this SNP shows allele-specific expression in normal thyroid tissue. However, it does not show differential expression between normal and PTC tumor tissue. To understand the functions of these genes we have performed a combination of luciferase assays, mass spectroscopy, overexpression, and knockdown followed by transcriptome profiling in PTC cell lines. In the era of GWAS it is often difficult to extrapolate functional genetic variants from statistical association. The strength of our studies lies in the functional validation of these loci, which will allow us to better understand genetic risk and network of genes involved in PTC.

732W

Genome-wide association study to identify novel biomarkers for trastuzumab-induced cardiotoxicity. M. Hara^{1,2}, C. Udagawa¹, A. Shimo³, Y. Kojima², R. Yoshie², H. Zaha³, N. Abe³, T. Motonari³, M. Unesoko³, K. Tamura⁴, T. Shimo⁵, M. Yoshida⁶, T. Yoshida⁶, T. Mushiroda⁷, K. Tsugawa², H. Zembutsu^{1,8}. 1) Cancer Precision Med. Ctr. JFCR, Tokyo, Japan; 2) Dept. Breast and Endocrine Surg., St. Marianna Univ., Sch. of Med. Kanagawa, Japan; 3) Dept. Breast Surg., Nakagami Hosp. Okinawa, Japan; 4) Dept. Breast and Med. Oncology, Natl. Cancer Ctr. Hosp. Tokyo, Japan; 5) Dept. Pathol, Natl. Cancer Ctr. Hosp. Tokyo, Japan; 6) Fund. Innov. Oncol. Core, Natl. Cancer Ctr. Res. Inst. Tokyo, Japan; 7) RIKEN Ctr. Integrative Med. Sci. Kanagawa, Japan; 8) Div. Genetics, Natl. Cancer Ctr. Res. Inst. Tokyo, Japan.

To identify a novel genetic marker(s) determining the risk of trastuzumab-induced cardiotoxicity, we carried out a genome-wide association study (GWAS) using 11 cases (with trastuzumab-induced cardiotoxicity) and 257 controls (showing no sign of trastuzumab-induced cardiotoxicity). Top 100 single nucleotide polymorphisms (SNPs) which revealed smallest *P* value in GWAS ($P = 1.61 \times 10^{-7} - 2.30 \times 10^{-3}$) were further examined using replication samples consisted of 14 cases and 199 controls. The combined analysis of the screening and replication study indicated 5 loci showing stronger association with trastuzumab-induced cardiotoxicity than GWAS phase (SNP A on chromosome 13; $P_{\text{combined}} = 6.00 \times 10^{-6}$, SNP B on chromosome 4; $P_{\text{combined}} = 8.60 \times 10^{-5}$, SNP C on chromosome 15; $P_{\text{combined}} = 1.01 \times 10^{-4}$, SNP D on chromosome 17; $P_{\text{combined}} = 1.07 \times 10^{-4}$ and SNP E on chromosome 15; $P_{\text{combined}} = 1.60 \times 10^{-4}$). Of the above 5 loci, non-coding RNAs were included on Chr 13 and 15, and a secreted protein was also included on Chr 15. These findings provide new insights into personalized trastuzumab therapy for patients with human epidermal growth factor receptor 2 (HER2)-positive cancer.

733T

The role of genetic variants in methotrexate-induced toxicities in patients with osteosarcoma. M.J.H. Coenen¹, M.J. Klumpers², E.G.E. Hurkmans¹, H.I. Vos², S.H. Vermeulen^{1,3}, M.M. Haglertner², U. Flucke⁴, H.W.B. Schreuder⁵, F.N. van Leeuwen¹¹, H. Gelderblom⁶, E.S.J.M. de Bont⁷, L.C.M. Kremer⁸, J. Bras⁹, H.N. Caron⁸, H.-J. Guchelaar¹⁰, D.M.W.M. te Loo². 1) Dept. of Human Genetics, Radboud Institute for Health Sciences, Radboud university medical center, Nijmegen, The Netherlands, Netherlands; 2) Dept. of Pediatric Hematology and Oncology, Amalia Children's Hospital, Radboud university medical center, Nijmegen, The Netherlands; 3) Dept. for Health Evidence, Radboud Institute for Health Sciences, Radboud university medical center, Nijmegen, The Netherlands; 4) Dept. of Pathology, Radboud university medical center, Nijmegen, The Netherlands; 5) Dept. of Orthopedic Surgery, Radboud university medical center, Nijmegen, The Netherlands; 6) Dept. of Medical Oncology, Leiden University Medical Center, Leiden, The Netherlands; 7) Dept. of Pediatric Hematology and Oncology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands; 8) Dept. of Pediatric Oncology, Emma Children's Hospital/Academic Medical Center, Amsterdam, The Netherlands; 9) Dept. of Pathology, Academic Medical Center, Amsterdam, The Netherlands; 10) Dept. of Clinical Pharmacy & Toxicology, Leiden University Medical Center, Leiden, The Netherlands; 11) Laboratory of Pediatric Oncology, Amalia Children's Hospital, Radboud university medical center, Nijmegen, The Netherlands.

Background: High-dose methotrexate (HD-MTX) is a cornerstone agent in the chemotherapeutic treatment of patients with osteosarcoma. Due to inter-individual variation in MTX metabolism, patients often develop HD-MTX-induced toxicities which cannot always be prevented by MTX plasma level monitoring. We aim to identify determinants of HD-MTX-induced toxicities in osteosarcoma patients, by investigating the relation between MTX plasma levels and toxicities, and the contribution of genetic variants in genes related to drug absorption, distribution, metabolism and elimination (ADME). **Methods:** A cohort of 114 osteosarcoma patients was genotyped for 1,931 variants in 231 genes using the Drug Metabolism Enzymes and Transporters (DMET) Plus array. Laboratory results during treatment concerning renal function, liver damage and myelopoiesis were obtained to evaluate toxicity. Associations were tested between toxicity data, 48-hour MTX plasma levels, and ADME genetic variants by performing generalized estimating equation analyses to take repeated measurements into account. **Results:** Analyses of data concerning 1,238 HD-MTX courses identified no association between 48-hour MTX plasma levels and related toxicities. Three of the tested genetic variants were significantly associated with HD-MTX induced toxicities. We found an association between a variant in *CYP2F1* (rs58285195) and decreased neutrophil count (coef -0.600 [95% CI -0.753–0.448]; $p=1.40 \times 10^{-4}$). Also, a variant in *CYP4F2* (rs3093153) was significantly associated with increased alanine transaminase levels (coef 0.813 [95% CI 0.469–1.157]; $p=3.67 \times 10^{-6}$) and a variant in *CYP4F8* (rs4808326) was associated with a protective effect on platelet count (coef 0.198 [95% CI 0.105–0.291]; $p=2.91 \times 10^{-5}$) in our cohort. **Conclusions:** Our study identified three variants in genes involved in drug processing to be strongly associated with HD-MTX-induced toxicities. To date, the variants in these genes were not known to play a role in MTX metabolizing pathways and were also not previously found to have an influence of the development and severity of bone marrow suppression or liver damage. Further functional investigation of variants in the identified genes is ongoing to study if and how they affect the development of HD-MTX-induced toxicities.

734F

Optimizing cisplatin chemotherapy through *SNAIL* pathway expression regulation implicates *S100P* as potential target. C. Del Greco, L. Mortensen, A.L. Stark. University of Notre Dame, Notre Dame, IN.

Cisplatin is a common platinum chemotherapy treatment given to breast cancer patients, but it is plagued with non-response and resistance. Much research has been focused on trying to understand and improve therapy hurdles. One promising avenue is the *SNAIL* pathway, which has complex roles throughout the cell. Given the knowledge of the cisplatin mechanism, we hypothesized that gene expression changes seen after *SNAIL* upregulation would be reversed after cisplatin treatment. In order to investigate this prediction, we analyzed gene expression in MCF-7 cells in response to upregulated *SNAIL* as found in the Expression Atlas. The fourteen genes with the lowest p-values for expression changes in response to *SNAIL* overexpression ($p=1.88 \times 10^{-8}$) were then researched to prioritize the most promising targets for cisplatin therapy modification. Genes selected for initial study were *THBS1*, *S100P*, *LDHB*, *DECRL1*, *TSPYL5*, and *TMEM47*. qPCR was performed to assess the expression changes in each gene when cells are exposed to 5 μ M cisplatin treatment for 24 and 48 hours. Most of the genes showed very modest expression changes after both time points of cisplatin exposure, suggesting they may not be affected by cisplatin therapy. However, *S100P* expression increased by a log fold of 4.47 ($p=5.70359 \times 10^{-6}$) after 24 hours of exposure to cisplatin treatment, and increased by a log fold of 14.34 ($p=4.98091 \times 10^{-5}$) after 48 hours of exposure to cisplatin treatment. This is especially relevant given that *S100P* has not only been observed to decrease cell adhesion while increasing invasion and metastasis, but has also been known to interact with, and in some cases desensitize cells to, therapies such as cisplatin. Future research will be therefore aimed towards assessing the effects of inhibiting *S100P* in cisplatin-treated cells on cellular growth with the hope of optimizing cisplatin chemotherapy outcomes.

735W

Association between imputed pancreatic gene expression profiles and the development of L-asparaginase-induced pancreatitis. *B.I. Drogemoller^{1,2}, G.E.B. Wright^{2,3}, S. Ito⁴, B.C. Carleton^{2,5,6}, C.J.D. Ross^{1,2}*, *The Canadian Pharmacogenomics Network for Drug Safety Consortium*. 1) Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver, BC, Canada; 2) BC Children's Hospital Research Institute, University of British Columbia, Vancouver, BC, Canada; 3) Department of Medical Genetics, Centre for Molecular Medicine and Therapeutics, University of British Columbia, Vancouver, BC, Canada; 4) Clinical Pharmacology and Toxicology, The Hospital for Sick Children, University of Toronto, Toronto, ON, Canada; 5) Division of Translational Therapeutics, Department of Pediatrics, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada; 6) Pharmaceutical Outcomes Programme, BC Children's Hospital, Vancouver, BC, Canada.

L-asparaginase is highly effective in the treatment of pediatric acute lymphoblastic leukemia. Unfortunately, the use of this treatment is limited by the occurrence of pancreatitis, a severe and potentially lethal adverse drug reaction, which occurs in 2-18% of patients. As previous studies have been unable to identify strong associations between clinical variables and susceptibility to L-asparaginase-induced pancreatitis, genetic factors are expected to play an important role in the development of this adverse drug reaction. To uncover these genetic factors, a total of 284 patients treated with L-asparaginase were recruited from 13 pediatric oncology units across Canada. Extensive demographic and clinical data were collected for all patients and genotyping was performed using the Illumina HumanOmniExpress array. Pancreatic gene expression profiles were imputed using GTEx v7 and S-PrediXcan. Genes that were predicted to be differentially expressed in cases and controls ($P < 0.01$) were included in overrepresentation enrichment analyses using the WEB-based GENE SeT AnaLysis Toolkit. S-PrediXcan analyses identified one gene, *SLC39A13*, which exhibited significantly higher imputed expression levels in individuals with L-asparaginase-induced pancreatitis ($P = 1.2 \times 10^{-5}$). WEB-based GENE SeT AnaLysis Toolkit analyses revealed an enrichment of genes (*HLA-DQB1*; *HLA-DRB5*) that have previously been associated with drug-induced toxicity (adjP=0.02). In conclusion, this study has uncovered associations with L-asparaginase-induced pancreatitis and predicted gene expression levels of (i) *SLC39A13* - a gene involved in Ehlers-Danlos syndrome, a condition in which pancreatitis has been reported to occur, and (ii) *HLA* genes - mirroring previous associations with the *HLA* region and pancreatitis caused by other drugs. These findings have provided additional insight into the genetics factors that are involved in the development of L-asparaginase-induced pancreatitis, the results of which will facilitate future replication and validation studies.

736T

Elucidating the relationship between ER stress and cisplatin on gene regulation in colorectal cancer cells. *A.P. Hubert, R. Georgiadis, A. Stark*. College of Science, University of Notre Dame, Notre Dame, IN.

The chemotherapeutic agent cisplatin has been shown to increase survival rates in colorectal cancer patients. It produces a cytotoxic environment inside cells, initiating several death pathways while producing ER stress. Impacts of cisplatin treatment on gene expression could help characterize cellular responses to the drug and identify important genes to therapy response. Utilizing the Expression Atlas, expression profiles of HCT116 colorectal cells with induced ER stress by Brefeldin A identified genes with large expression changes. Over 800 genes were identified with expression changes $p < 0.00011$, which was unsurprisingly due to the importance of the ER stress pathway in the cell. Based on their large fold change, three genes were initially selected for examination with cisplatin treatment: *MALAT1* ($p = 5.624 \times 10^{-5}$), *SLC7A11* ($p = 1.04 \times 10^{-4}$) and *SPX* ($p = 2.827 \times 10^{-5}$). Due to the expression fold changes observed under ER stress (³⁶), our hypothesis is that during cisplatin treatment, cells would upregulate the expression of all three genes. HCT116 colorectal cancer cells were treated with 5 uM cisplatin for 24 hours and 48 hours. Quantitative PCR was run on cDNA to determine the effect on expression of the genes in question. All three genes showed evidence of significant down-regulation. *MALAT1* is a gene normally upregulated in cancer cells and the observed downregulation suggests that this gene may play a role in cisplatin response. High expression of *SLC7A11* has been observed in the literature in cisplatin resistant cell lines in several cancers including lung and bladder cancer. However, the observed downregulation suggests tissue specificity. *SPX* is a gene which is involved maintaining homeostasis and further work needs to be done to understand its role in cisplatin treatment. Overall, these genes were expected to show increased expression due to cisplatin induced ER stress, however, because they did not, further work needs to be done to understand the role of ER stress in colorectal cancer cells cisplatin response. Ultimately, understanding the mechanistic similarities between ER stress and cisplatin treatment in colorectal cells will pinpoint future therapy strategies for more effective treatment.

737F

Genome-wide association study of anthracycline-induced cardiotoxicity in adult cancer patients. E.N. Scott^{1,2}, B.I. Drögemöller^{2,3}, G. Wright^{1,2}, K. Gelmon⁴, D. Villa⁴, J.N. Trueman⁵, B.C. Carleton^{5,6}, C.J.D. Ross^{1,2,3}. 1) Department of Medical Genetics, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada; 2) BC Children's Hospital Research Institute, Vancouver, BC, Canada; 3) Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver, BC, Canada; 4) Division of Medical Oncology, BC Cancer, Vancouver, BC, Canada; 5) Pharmaceutical Outcomes Programme, BC Children's Hospital, Vancouver, BC, Canada; 6) Division of Translational Therapeutics, Department of Pediatrics, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada.

Background Anthracycline agents, such as doxorubicin and epirubicin, are commonly prescribed for a variety of cancers and contribute largely to the survival of patients. Nevertheless, up to 18% of adult patients develop severe life-threatening anthracycline-induced cardiotoxicity (ACT). Clinical risk factors include cumulative anthracycline dose, age, prior thoracic radiotherapy, and pre-existing cardiac disease. Germline genetic variants associated with a higher ACT risk in adults have not yet been identified. The objective of this study is to replicate previous, and discover new, ACT-associated genetic variants in adult cancer patients in order to improve our understanding of the genetic causes of ACT. **Methods** Genotyping of 1.7 million variants has been performed on 175 adult cancer patients (59 cases, 116 controls) treated with anthracyclines, recruited from academic health centres in Vancouver, and rigorously assessed for ACT. Targeted analyses will examine genetic variants in 25 genes that have been previously implicated in ACT and identified through an extensive literature review. Genome-wide analyses will identify novel variants associated with ACT. Targeted and genome-wide analyses will be performed using logistic regression, including significantly-associated clinical variables as covariates. Functional and biological relevance of candidate variants will be examined using various computational annotation tools, such as VEP, GTEX, and CADD, and associated genes will be investigated for enrichment in specific pathways. **Results** An extensive review of the literature has identified 25 genes that have been previously associated with ACT, the role of which will be validated in our cohort through a targeted analysis. We will present the top candidate variants that reach statistical significance in both targeted and genome-wide analyses. **Conclusions** The results from this study will enable the development of genetic tests that can be used to determine which patients are at high risk of experiencing ACT. In addition, the identification of genetic variants associated with ACT will play an important role in improving the safety of anthracyclines and, ultimately, in reducing the burden on the healthcare system as well as affected patients.

738W

The alkylating drug Irofulven shows potent tumor inhibition in an ERCC3 mutant background. S. Topka^{1,2}, S. Khalil³, Z. Steinsnyder^{1,2}, V. Ravichandran^{1,2}, M. Winkel Madsen⁵, E. De Stanchina⁴, J. Vijai^{1,2,7}, K. Offit^{1,2,6,7}. 1) Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY; 2) Cancer Biology and Genetics Program, Memorial Sloan Kettering Cancer Center, New York, NY; 3) Gerstner Sloan Kettering Graduate School of Biomedical Sciences, New York, NY; 4) Antitumor Assessment Core Facility, Molecular Pharmacology Program, Memorial Sloan Kettering Cancer Center, New York, NY; 5) Oncology Venture ApS, Horsholm, Denmark; 6) Clinical Genetics Service, Department of Medicine, Memorial Sloan Kettering, New York, N.Y.; 7) Department of Medicine, Weill Cornell Medical College, New York, N.Y.

The ATP dependent helicase ERCC3 plays a crucial role in normal basal transcription and nucleotide excision repair. A recurrent Ashkenazi Jewish founder mutation in ERCC3 (p.R109X) has been shown to confer risk for multiple cancer types. Modeling of this mutation in mammary epithelial cell lines via CRISPR/Cas9 showed a consistent pattern of reduced expression of the protein and concomitant hypomorphic functionality when treated with the DNA alkylating agent IlludinS. A derivative of IlludinS- Irofulven- has previously been evaluated as an antitumor agent in multiple clinical trials and has shown high antitumor activity. Higher effective doses of this drug have been shown to give rise to side effects, including retinal toxicity. Based on the observation that ERCC3 mutant cell lines are sensitive to lower doses of IlludinS we aimed to explore the IlludinS-derivative Irofulven as a potential therapeutic agent to treat patients with hypomorphic ERCC3 mutations. We tested the efficacy of Irofulven in isogenic WT and ERCC3 mutant mammary epithelial cell lines *in vitro* and in xenograft models. *In vitro*, cells were treated with various doses of Irofulven and cell viability was assessed at 72h following treatment. For generation of xenograft models hRAS-V12 expressing genetically engineered ERCC3 mutant as well as isogenic ERCC3 wild-type cell lines were injected into flanks of nude mice and treated with different doses of Irofulven. Tumor volumes were measured over the course of treatment and tumors excised after termination of treatment for final weighing and for histological and subsequent molecular-genetic analysis. *In vitro* studies showed an increased sensitivity of the ERCC3 mutant cells towards the drug Irofulven, while no difference was observed for other commonly used drugs such as PARP inhibitors or platinum-based agents. Xenograft experiments showed enhanced *in vivo* sensitivity in the ERCC3 mutant background. Additional experiments are ongoing including RNA sequencing of tumor tissue and exploration of combinatorial therapies. Similar to tumors arising in the background of homologous repair defects, mutations in nucleotide excision repair genes such as ERCC3 could constitute potential therapeutic targets in a subset of cancers. The present study provides a potential pre-clinical approach to a subset of cancer cases with an inherited mutation of a nucleotide excision repair pathway gene, demonstrating sensitivity to the DNA alkylating agent Irofulven.

739T

Application of human genetics to launch and guide the discovery of a novel cancer immunotherapy. J. Hutz, C. Zhang, Y. Huang, D. Chandra, A. Endo, K. Huang, X. Bao. Eisai, Andover, MA.

Though it has been shown that human genetics is useful for the identification of drug targets, less has been shared to show how human genetics can continuously impact drug discovery and development after target identification. We present an example that highlights the many contributions of human genetics to the discovery of E7766, a small molecule agonist of STING for the treatment of cancer. First, STING (encoded by *TMEM173*) was chosen as a promising target based on both GWAS and Mendelian support. Linked SNPs in the *TMEM173* locus had been associated with higher esophageal cancer risk and lower interferon alpha production. Additionally, gain-of-function mutations in *TMEM173* had been shown to cause a rare type 1 interferonopathy (SAVI). The combination of these findings suggested that STING-induced interferon production could be beneficial for the treatment of cancer. Based in part on this information, chemistry and pharmacology efforts were initiated and resulted in the discovery of E7766, a potent STING agonist. Second, human genetics played a critical role in outlining a pharmacogenomic basis for differentiation of E7766 from other STING agonists. Five common *TMEM173* haplotypes exist worldwide. In collaboration with a consumer genetics company, we involved sequenced research participants in a study and demonstrated that PBMCs carrying all common genotypes responded to E7766 *in vitro* but that other STING agonists may benefit from pharmacogenomic testing. Understanding the geographic frequencies of these genotypes supported selection of clinical trial sites that can test clinical efficacy in all genotypes. Third, human genetics provided a clear set of biomarkers to be tested in clinical trials. SAVI patients demonstrate increased expression of a type 1 interferon gene signature. We confirmed that E7766 induces similar but temporary gene expression changes in mice and human primary cells and may test expression of these genes to confirm activity of E7766 in treated patients. Finally, reports have emerged of successful treatment of SAVI patients using JAK inhibitors, which may be useful in considering how to address unexpected adverse events in clinical trials of STING agonists. Though it is unknown whether E7766 will eventually become an approved therapy for patients with cancer, its discovery and future development provide useful insight on how human genetics can guide therapeutic development.

740F

Exploring function of genetic variants in lung adenocarcinoma by integrated association analysis of multi-omics data. J.B. Chen, S.S. Dong, Y. Rong, X.F. Chen, H. Wu, Y.J. Zhang, Y. Guo, T.L. Yang. Key Laboratory of Biomedical Information Engineering of Ministry of Education, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, Shannxi, China.

Previous genome-wide association studies (GWASs) have identified many single nucleotide polymorphisms (SNPs) associated with human complex traits. Expression quantitative trait loci (eQTL) analysis were then used to explore how these genomic variants influence gene expression and associated traits. However, most of SNPs are located in genome's noncoding regions such as intergenic or intron regions, which may regulate gene expression rather than change gene expression levels directly, therefore, could not be identified by eQTL analysis. Here, for interpreting how these genomic variants associated with human complex traits, we performed association analyses of genotype data with matched transcriptomic (gene expression, alternative splicing and back-splicing), epigenomic (DNA methylation) and proteomic (RPPA) data in lung adenocarcinoma (LUAD) samples from The Cancer Genome Atlas (TCGA). After quality control and analyzing covariates, 12,336 cis-eQTLs, 61,926 cis-splicing quantitative trait loci (sQTLs), 81,552 DNA methylation quantitative trait loci (mQTLs) and 1,215 protein quantitative trait loci (pQTLs) were identified with the significance threshold of false discovery rates (FDR) <0.05. Importantly, we introduced circRNA quantitative trait loci (cQTLs) as the association of genotype with circRNA expression levels which focused on gene expression regulatory network by genomic variants via back-splicing events. Finally, we identified 33,442 cQTLs in LUAD samples. A large portion of known lung adenocarcinoma genes such as *AKT1/2*, *BRAF* and *EGFR* were included in these QTLs genes, especially for cQTL, mQTL and sQTL genes that overlapped with more than half of known genes respectively. Among the top QTL genes with large numbers of SNPs out of known cancer genes, *ANXA2*, *COL1A1*, *PIN4* and *YWHAZ* were further validated by survival analysis, indicating their role of candidate lung cancer genes. For sQTLs, skipped exons (SE) event harbored the most sQTLs and some known lung cancer genes such as *AKT1/2*, *ERBB2* and *TPM3* had complex splicing events. A large number of skipped exons were matched with circularized exons that also validated the exon-skipping model of circRNAs formation. Totally, these QTLs overlapped with 22% of GWAS SNPs. Our research that integrating various dimension of data resources will be helpful to obtain a more comprehensive map of the cancer genomic landscape and investigate the regulating mechanisms of these QTLs in lung adenocarcinoma pathology.

741W

A meta-analysis of more than 237,380 men of diverse ancestries identifies 40 new risk loci for prostate cancer. D.V. Conti¹, L.C. Moss¹, X. Sheng¹, S.K. Van Den Eeden², H. Nakagawa³, J. Witte^{4,5}, R.A. Eeles⁶, Z. Kote-Jarai⁶, C.A. Haiman¹, PRACTICAL Consortium. 1) Preventive Medicine, University of Southern California, Los Angeles, CA., United States of America; 2) Division of Research, Kaiser Permanente Northern California, Oakland, California, United States of America; 3) Laboratory for Genome Sequencing Analysis, RIKEN Center for Integrative Medical Sciences, Tokyo, Japan; 4) Department of Epidemiology and Biostatistics, University of California, San Francisco, California, United States of America; 5) Institute for Human Genetics, University of California, San Francisco, California, United States of America; 6) The Institute of Cancer Research, London, United Kingdom.

Although genome-wide association studies (GWAS) have identified over 150 independent loci associated with prostate cancer risk, these novel discoveries were primarily in populations of European ancestry. In order to identify additional prostate cancer risk loci, we conducted a meta-analysis of ~30 million genotyped and imputed SNPs in 110,406 prostate cancer cases and 126,974 controls consisting of individuals of African, European, Asian, and Latino ancestry. After excluding known prostate cancer risk loci ($\pm$ 800kb of index marker), our meta-analysis identified over 40 novel loci associated with overall prostate cancer risk (P value $< 5 \times 10^{-8}$). We examine the polygenic risk score, heritability and proportion of familial relative risk within and between populations captured by our novel loci, as well as previously reported regions. These findings demonstrate the value of combining populations with individuals of diverse ancestry for both novel discovery and improving risk prediction. Additionally, these results will further guide future research into understanding the biological mechanisms underlying prostate cancer susceptibility.

742T

Increased penetrance of acute lymphoblastic leukemia susceptibility loci in children with Down syndrome. A.J. de Smith¹, A.L. Brown², V.U. Gant³, M.E. Scheurer⁴, K.M. Walsh⁵, N.J. Winick⁶, N.A. Heerema⁵, A.J. Carroll⁶, M.J. Borowitz⁷, B.L. Wood⁸, W.L. Carroll⁹, E.A. Raetz¹⁰, E. Feingold¹¹, W. Yang¹², M. Devidas¹³, D. Sinnett¹⁴, C.G. Mullighan¹⁵, S.P. Hunger¹⁵, C.H. Pui¹², M. Loh¹⁶, M.E. Zwick¹⁷, C. Metayer¹⁸, X. Ma¹⁹, B.A. Mueller²⁰, S.L. Sherman¹⁷, J.L. Wiemels^{1,16}, J.J. Yang¹², M.V. Relling¹², P.J. Lupo³, K.R. Rabin². 1) Department of Preventive Medicine, University of Southern California, Los Angeles, CA; 2) Baylor College of Medicine, Houston, TX; 3) Duke University, Durham, NC; 4) University of Texas Southwestern Medical Center, Dallas, TX; 5) Ohio State University, Columbus, OH; 6) University of Alabama Birmingham, Birmingham, AL; 7) Johns Hopkins Hospital, Baltimore, MD; 8) University of Washington Medical Center, Seattle, WA; 9) New York University, New York, NY; 10) University of Utah Huntsman Cancer Institute, Salt Lake City, UT; 11) University of Pittsburgh, Pittsburgh, PA; 12) St. Jude Children's Research Hospital, Memphis, TN; 13) University of Florida, Gainesville, FL; 14) University of Montreal, Montreal, QC, Canada; 15) Children's Hospital of Philadelphia, Philadelphia, PA; 16) University of California San Francisco, San Francisco, CA; 17) Emory University, Atlanta, GA; 18) University of California Berkeley, Berkeley, CA; 19) Yale School of Public Health, New Haven, CT; 20) Fred Hutchinson Cancer Research Center, Seattle, WA.

Down syndrome (DS) is one of the strongest risk factors for acute lymphoblastic leukemia (ALL), the most common pediatric malignancy, as children born with trisomy 21 have an approximately 20-fold increased risk of disease. ALL in children with DS (DS-ALL) is associated with inferior outcomes and unique somatic characteristics, including a high frequency of CRLF2 overexpression. However, the role of inherited genetic variation in DS-ALL is unknown. To this end, we performed the first genome-wide association study (GWAS) of DS-ALL, including four independent case-control studies comprising 542 DS-ALL cases and 1192 DS controls. We carried out genome-wide imputation of autosomal and disomic single nucleotide polymorphisms (SNPs), and combined results from all studies using fixed effects meta-analysis. Genome-wide significant SNPs were identified at four loci previously associated with non-DS ALL, including: rs58923657 near *IKZF1* (odds ratio [OR] = 2.02, P_{meta} = 5.32×10^{-16}), missense SNP rs3731249 in *CDKN2A* (OR = 3.63, P_{meta} = 3.91×10^{-10}), rs7090445 in *ARID5B* (OR = 1.60, P_{meta} = 8.44×10^{-8}), and rs3781093 in *GATA3* (OR = 1.73, P_{meta} = 2.89×10^{-8}). Case-case analyses comparing the frequency of ALL risk alleles (at *IKZF1*, *CDKN2A*, *ARID5B*, *GATA3*, *CEBPE*, *BMI1*, and *PIP4K2A*) in DS-ALL versus non-DS ALL ($n = 3082$ cases) revealed significantly higher risk allele frequencies in DS-ALL cases for SNPs in *CDKN2A* (OR = 1.58, P_{meta} = 4.08×10^{-4}), *GATA3* (OR = 1.34, P_{meta} = 4.38×10^{-5}), and *IKZF1* (OR = 1.18, P_{meta} = 0.015). Genetic risk scores (GRS) calculated from the seven susceptibility loci were also significantly higher in DS-ALL cases than in non-DS ALL cases (weighted GRS, $P = 3.33 \times 10^{-6}$), indicating that DS-ALL cases harbor significantly more risk alleles than non-DS cases. Furthermore, among controls without ALL, neither the GRS nor any individual ALL risk alleles were associated with DS status, suggesting that the higher risk allele frequencies in DS-ALL cases represents a larger magnitude of effect for known ALL risk loci in the genetic background of trisomy 21. Finally, shRNA knockdown of *IKZF1* in lymphoblastoid cell lines (LCLs) revealed significantly higher proliferation rates in DS LCLs versus non-DS LCLs, as measured by serial counts ($P < 0.05$) and proportion of cells in S-phase ($P < 0.05$). Our results support a higher penetrance of ALL susceptibility loci in the DS population, and may inform surveillance strategies for children with DS who are at the greatest risk of developing ALL.

743F

Multi-omics analysis of Ashkenazi Jewish melanoma patients reveals novel genetic markers of melanoma recurrence. A. Archambault¹, D. Simpson¹, T. Hekal¹, R. Ferguson¹, E. Kazlow¹, U. Moran², Y. Lee², A. Izsak², M. Wilson², R. Shapiro², A. Pavlick³, I. Osman², T. Kirchhoff⁴. 1) Perlmutter Cancer Center, NYU Langone Health, New York, NY; 2) Department of Dermatology, NYU Langone Health, New York, NY; 3) Department of Surgery, NYU Langone Health, New York, NY; 4) Department of Medicine, NYU Langone Health, New York, NY.

A challenge in clinical management of cutaneous melanoma (CM) is the limited predictability of recurrence and hence the progression to advanced stages that associate with less favorable outcomes. Aggressive follow-up care and novel adjuvant therapy strategies in early-stage patients with high-risk of recurrence would likely improve CM-specific survival and reduce mortality. We developed a genome-wide approach to identify germline genetic determinants of melanoma prognosis as putative personalized biomarkers for patients at risk of CM recurrence. Exploring both the coding and non-coding transcribed genome we performed germline whole genome sequencing (WGS) and tumor RNA-seq on 96 CM patients with tumor/blood matched specimens. All patients were of primary stages I-IIb and of Ashkenazi Jewish ancestry to reduce genetic heterogeneity. We compared 48 patients that recurred in <4 years versus 48 patients with recurrence in > 6 years. Univariate and multivariate logistic regression, gene-burden analysis (SKAT), and differential expression analyses of both mRNA and long non-coding RNA (lncRNAs) were used to identify germline regions associated with melanoma recurrence. Several gene regions were associated with melanoma recurrence, with NEGR1 and MGST3 both passing SKAT levels of significance ($p < 1e-05$), and logistic regression analysis on common WGS variants found over 100 variants with significance $p < 1e-03$. In addition, we found 200 differentially expressed putative lncRNAs ($p < 0.05$). The analysis of germline WGS found rs199818927, a 1bp insertion previously associated with both NEGR1 and lncRNA LINC01360, among our top 5 most significant regression results (OR=9.107 $p=2.31e-05$). This initial phase of our large scale whole genome scan has uncovered germline variants in several coding loci and putative lncRNAs that associate with melanoma recurrence. Most notably, we identified germline variation in a lncRNA near NEGR1, a putative tumor suppressor that has been shown to be under-expressed in advanced cancers, indicating its putative role in cancer progression to metastatic stages. We are currently expanding the patient cohorts and validating these results.

744W

Association of non-homologous recombination genes with ovarian cancer using clinical exomes as controls. K. Arvai, R. Torene, L. Susswein, M. Roberts, M. Marshall, Z. Zhang, K. Retterer, K. Hruska. GeneDx, Gaithersburg, MD.

The decreasing cost of next generation sequencing (NGS) has allowed clinical laboratories to expand hereditary cancer panels to include newer genes in addition to well-established cancer predisposition genes like *BRCA1* and *BRCA2*. The associations of newer genes with ovarian cancer have been inconsistently reported in the literature. While recent case-control studies have successfully leveraged publicly available genetic data to estimate various cancer associations these data are limited by lack of phenotype information. We aimed to improve ovarian cancer association estimates by using unaffected females as controls and including age as a covariate in a case-control study to estimate ovarian cancer association with pathogenic variants (PV) in *ATM*, *BARD1*, *BRCA1*, *BRCA2*, *BRIP1*, *CHEK2*, *NBN*, *PALB2*, *RAD51C*, and *RAD51D*. Cases were women with ovarian cancer referred for hereditary cancer multi-gene panel testing. Controls were mothers with no reported family history of cancer selected from a genotyped cohort of 10,138 complete trios who were referred for exome sequencing. Both cases and controls were limited to those of European ancestry. Classification of sequence variants in cases and controls was based on the AMP/ACMG guidelines. Individuals with multiple PVs were excluded. We performed simple and multiple logistic regression to generate crude and adjusted odds ratios (OR). The sample population consisted of 6,220 unique cases and 5,629 controls. The mean age of diagnosis in cases (58.7y) and age of testing in controls (41.8y) were significantly different (t-test, $p < 0.05$). The PV carrier frequency in the controls ranged from 0.04% in *BARD1*, *RAD51C*, and *RAD51D* to 1.0% in *CHEK2*. The PV carrier frequency in the cases ranged from 0.08% in *BARD1* to 3.7% in *BRCA1*. Associations in *BRIP1* and *PALB2* have previously been reported with ovarian cancer. Both genes had 95% confidence intervals (CI) >1.0 by crude OR ($p=0.01$, $p=0.04$), but adjusted CI were <1.0 ($p=0.99$, $p=0.20$), indicating no association with ovarian cancer. We show that case-control studies using internal controls from clinical labs provide valuable information for genotype-phenotype associations. Using a statistical model, we were able to reproduce moderate to strong associations in genes that have been reported with ovarian cancer such as *BRCA1*, *BRCA2*, *RAD51C*, and *RAD51D*. Results from our novel approach provide reliable evidence suggesting that *BRIP1* and *PALB2* are not associated with ovarian cancer.

745T

Validation of a prostate cancer genetic risk score for clinical use. *M.H. Black¹, S. Li¹, H. LaDuca¹, J. Chen¹, R. Hoiness¹, S. Gutierrez¹, B. Tippin Davis¹, H.M. Lu¹, M. Gielzak², K. Wiley², Z. Shi², H. Yu², B.T. Helfand², W. Isaacs³, J. Xu².* 1) Ambry Genetics, Aliso Viejo, CA; 2) Northshore University Health System, Evanston, IL; 3) Johns Hopkins University, Baltimore, MD.

Background: Genome-wide association studies (GWAs) have identified over 100 common single nucleotide polymorphisms (SNPs) associated with prostate cancer (PrCa) risk. While the risk associated with each SNP is modest, a genetic risk score (GRS) based on their combined genotypes may have substantial predictive value for risk stratification. The overarching goal of this study is to provide evidence for the analytical and clinical validity, as well as clinical utility, of a PrCa GRS in the patient care setting. **Methods:** We genotyped 73 PrCa-associated SNPs using next-generation sequencing in order to examine whether a risk score based on these SNPs was predictive of PrCa in 3,148 Caucasian men (1,731 cases who had previously undergone radical prostatectomy and 1,417 controls unaffected with PrCa) from a large, multi-site collaborative study. We constructed a population-standardized GRS, with each SNP weighted by per-allele relative risks in Caucasians from large GWAs and population-specific allele frequencies, and tested GRS association with PrCa using multivariate logistic regression models adjusted for age alone, as well as age and family history of PrCa. **Results:** Mean±SD age of PrCa cases at diagnosis and age of controls at the time of testing or last clinic visit was 59.0±6.7 and 55.7±11.1 years, respectively. Among cases, 58.5% had Gleason score ≤6, while 41.5% had Gleason ≥8. In addition, 38.1% of cases and 16.5% of controls had ≥1 first- or second-degree relative with PrCa, and 6.4% of cases and 3.9% of controls tested positive for mutations in PrCa predisposition genes. The GRS was significantly higher in cases than controls (mean±SD 1.79±2.97 vs. 1.01±0.46, $p<0.0001$). Compared to men in the 1st quartile of age-adjusted GRS, those in the 2nd, 3rd and 4th quartile were 1.66 (95% CI: 1.35-2.04), 2.34 (95% CI: 1.90-2.88) and 3.77 (95% CI: 3.04-4.69) times as likely to have PrCa (all $p<0.0001$). Additional adjustment for family history among the subset of patients with this information yielded similar results. GRS predictive performance was consistent with prior literature (AUROC=0.64, 95% CI: 0.62-0.66). **Conclusions:** These data suggest that a 73-SNP GRS is predictive of PrCa. Our ongoing analyses will include GRS association with clinical features of PrCa, validation in non-Caucasian populations, and interaction with high- and moderate-penetrance PrCa predisposition genes, as well as clinical utility for personalized PrCa screening.

746F

Evaluation of the impact of rare-variants on adrenocortical carcinoma susceptibility. *R.J. Bohlender, Y. Yu, C. Huff.* Epidemiology, University of Texas MD Anderson Cancer Center, Houston, TX.

Evaluation of the Impact of Rare-Variants on Adrenocortical Carcinoma Susceptibility Authors: Ryan Bohlender, Yao Yu, Chad Huff Adrenocortical Carcinoma (ACC) is a rare form of cancer, with an incidence estimate as low as 1 to 2 per million per year. As a rare disease, epidemiological data is lacking and the underlying etiology of the disease, both environmental and genetic, is poorly understood. We conducted a whole-exome sequencing case-control study involving 75 ACC cases of European ancestry from The Cancer Genome Atlas (TCGA) and 3507 controls of European ancestry from the National Database for Autism Research (NDAR) to investigate sources of risk. We used the Variant Annotation, Analysis, and Search Tool (VAAST) to conduct a case-control association analysis. VAAST is a gene-based test, which leverages variant level information about phylogenetic conservation and amino acid substitution to prioritize variants likely to influence protein function. Due to differences in sequencing platforms, laboratory preparation, and capture technologies, we initially observed high levels of Type I error inflation. We used the Cross Platform Association Toolkit (XPAT), which is designed to control for cross-platform artifacts in association studies with heterogeneous sequencing datasets. Using XPAT, we observed consistency between empirical and observed p-values to alpha levels of 0.005. Application of Phevor2, which leverages phenotype ontology databases to reorder gene rankings establishes *KCNJ10* and *NF1* as the first and second ranked genes. We find *NF1* to be nominally significant in VAAST ($p=0.00122$), and its associated syndrome Neurofibromatosis Type 1, has been previously associated with ACC susceptibility. To estimate the additional risk of ACC contributed by mutations in *NF1*, we calculated gene-level odds ratios. Rare coding variants in *NF1* result in an estimated 3.4-fold (1.2, 8.0) increase in relative risk. The gene *KCNJ10* has not been previously associated with ACC, but is nominally significant in the VAAST analysis ($p=0.00041$), and results in an estimated 6.1-fold (1.5, 17.9) increase in relative risk. *KCNJ10* is involved in SeSAME Syndrome, which causes, e.g., epilepsy and deafness. .

747W

A novel statistical model for GWAS to identify somatic mutation effects using matched data. Y. Bu, M. Chen. The University of Texas at Dallas, Richardson, TX.

Genome-wide association studies (GWAS) has emerged as a useful tool to identify common genetic variants that are linked to complex diseases. Conventional GWAS are based on the case-control design where the individuals in cases and controls are independent. In cancer research, matched pair designs, which compare tumor tissues with normal ones from the same subjects, are becoming increasingly popular. Such designs succeed in identifying somatic mutations in tumors while controlling both germline and environmental factors. Somatic variation is one of the most important cancer risk factors that contribute to continuous monitoring and early detection of various cancers. However, most GWAS analysis methods, developed for unrelated samples in case-control studies, cannot be employed in the matched pair designs. A novel framework is proposed in this manuscript to accommodate for the particularity of matched-data in association studies of somatic mutation effects. In addition, we develop a Bayesian model to combine multiple markers to further improve the power of mapping genome regions to cancer risks.

748T

Variants in the PSCA gene associated with risk of cancer and non-neoplastic diseases: Systematic research synopsis, meta-analysis and epidemiological evidence. H. Cui¹, M. Zhang¹, M. Tang¹, S. Chen¹, Y. Fang¹, S. Liu¹, Z. Zeng¹, Z. Shen¹, J. Lu², D. Gu¹, B. Zhang^{1,2,3}. 1) Department of Epidemiology and Biostatistics, First Affiliated Hospital and Southwest School of Medicine, Army Medical University, Chongqing, 400038, China; 2) Department of Epidemiology, School of Public Health, Institute for Chemical Carcinogenesis, Guangzhou Medical University, Guangzhou, 510182, China; 3) Department of Epidemiology, School of Public Health, Shanghai Jiaotong University School of Medicine, Shanghai, 200025, China.

Variants in the PSCA gene have been linked with risk of multiple cancers and other diseases, yet with generally inconsistent results. We did a comprehensive research synopsis and meta-analysis to evaluate associations between variants in the PSCA gene and cancer and non-neoplastic diseases risk based on 54 articles including 81 657 cases and 442 693 controls. Venice criteria and false-positive report probability tests were used to grade the cumulative evidence of significant associations. Six variants showed a significant association with assessed diseases ($P < 0.05$). Cumulative epidemiological evidence of an association was graded as strong for rs2294008 [odds ratio (OR) = 1.32, $P = 5.10 \times 10^{-33}$], rs2976392 (OR = 1.30, $P = 7.39 \times 10^{-12}$), rs9297976 (OR = 0.75, $P = 1.36 \times 10^{-7}$), and rs138377917 (OR = 0.53, $P = 0.008$) with gastric cancer, rs2294008 with bladder cancer (OR = 1.15, $P = 1.50 \times 10^{-16}$), duodenal ulcer (OR = 1.49, $P = 2.23 \times 10^{-61}$), gastric ulcer (OR = 1.13, $P = 6.36 \times 10^{-7}$), and gastritis (OR = 1.35, $P = 1.19 \times 10^{-5}$). Moderate evidence was graded for rs12155758 (OR = 1.19, $P = 0.002$) and rs2976391 (OR = 1.38, $P = 6.07 \times 10^{-5}$) with gastric cancer. Functional annotation suggested these variants might fall in a region with strong enhancer activity. Our study provides summary evidence that variants in PSCA gene are associated with risk of gastric, bladder, and non-neoplastic disease and highlights its significant role in the pathogenesis of these diseases.

749F

Using germline variants to predict glioma risk and identify glioma subtype pre-operatively. J. Eckel-Passow¹, P. Decker¹, M. Kosel¹, T. Kollmeyer¹, A. Molinaro², T. Rice², A. Caron¹, K. Drucker¹, C. Praska¹, M. Pekmezci², H. Hansen², L. McCoy², P. Bracci², B. Erickson¹, J. Wiemels², J. Wiencke², M. Bondy², B. Melin⁴, T. Burns¹, C. Giannini¹, D. Lachance¹, M. Wrensch², R. Jenkins¹. 1) Mayo Clinic, Rochester, MN; 2) University of California, San Francisco, CA; 3) Baylor College of Medicine, Houston, TX; 4) Umea University, Umea, Sweden.

Background: Twenty-five single nucleotide polymorphisms (SNPs) are associated with overall glioma risk or with risk of specific subtypes of glioma. We hypothesized that the inclusion of these 25 SNPs with age at diagnosis and sex could predict risk of glioma as well as identify glioma subtypes. **Methods:** Case-control design and multinomial logistic regression were used to develop models to estimate the risk of glioma development while accounting for histologic and molecular subtypes. Case-case design and logistic regression were used to develop models to predict IDH mutation status. Each model included all 25 glioma risk SNPs, patient age at diagnosis and sex. A total of 1273 glioma cases and 443 controls from Mayo Clinic were used in the discovery set, and 852 glioma cases and 231 controls from UCSF were used in the validation set. All samples were genotyped using a custom Illumina OncoArray. **Results:** Patients in the highest 5% of the risk score had more than a 14-fold increase in relative risk of developing an IDH-mutant glioma. Large differences in lifetime absolute risk were observed at the extremes of the risk score percentile. For both IDH-mutant 1p/19q non-codeleted glioma and IDH-mutant 1p/19q-codeleted glioma, the lifetime risk increased from almost null to 2.3% and almost null to 1.7%, respectively. The SNP-based model that predicted IDH mutation status had a validation c-index of 0.85. **Conclusions:** These results suggest that germline genotyping has the potential to provide a new tool for clinicians for the initial management of newly-discovered brain lesions. Specifically, given the low lifetime risk of glioma, SNP-based risk scores should not be useful for population screening. However, they may be useful in certain clinically-defined high-risk groups.

750W

Immunomodulatory pathways impact genetic susceptibility to the development of multiple primary melanomas. R. Ferguson¹, A. Archambault¹, D. Simpson¹, E. Kazlow¹, R. Lax¹, U. Moran², M. Wilson³, R. Shapiro⁴, A. Pavlick², I. Osman², D. Polsky², T. Kirchhoff¹. 1) Perlmutter Cancer Center, NYU Langone Health, New York, NY; 2) Department of Dermatology, NYU Langone Health, New York, NY; 3) Department of Medicine, NYU Langone Health, New York, NY; 4) Department of Surgery, NYU Langone Health, New York, NY.

During their lifetime about 8% of patients with single primary cutaneous melanoma (SPM) will develop multiple primary melanomas (MPM), which are associated with significantly higher mortality compared to patients with SPM. Based on the evidence that the immune system plays a role in regulating melanoma progression we explored whether germline genetic variants controlling the expression of immunomodulatory genes (immunomodulatory quantitative trait loci, eQTLs) discern risk of MPM compared to patients with SPM or healthy controls. Previously, we identified 50 eQTLs significantly associated with the expression of 265 immunomodulatory genes using the MuTHER twin cohort. These 50 SNPs were genotyped in 837 SPM and 104 MPM individuals using MassARRAY system. 1047 healthy controls were obtained from a publicly available GWAS on CM ascertained at MD Anderson (phs000187.v1.p1). We employed multivariate logistic regression to test the association of SNPs with MPM vs cancer-free controls and MPM vs SPM. When comparing MPM vs SPM, rs2071304, previously linked to expression of *SPI1* in MuTHER data, showed a strong association with reduction of MPM risk (OR= 0.60; 95% CI=0.45-0.81; p=0.0007). Intriguingly, this variant also trended toward significance when comparing MPM vs controls (OR=0.61; 95% CI: 0.44-0.85; p=0.003). Finally, our most significant association when comparing MPM to controls was for rs2276645 (OR= 0.60; 95% CI=0.45-0.81; p= 0.0008), an eQTL associated with *Zap-70* expression. Our data, for the first time, indicate that the inherited host immunity impacts risk of MPM in individuals with SPM, highlighting an importance of immune involvement in melanoma progression. The MPM risk-predicting genetic variants identified here or in expanded efforts, currently underway, may eventually lead to a diagnostic tool allowing for enhanced screening and clinical management of patients at risk of MPM, hence reducing elevated MPM-associated mortality. Additionally, our results further support that MPM and SPM may have different genetics underpinnings and should be treated as separate clinical entities.

751T

PSCA in multiple cancer types: Association with cancer risk and survival. O. Florez-Vargas¹, C. Tanikawa², W. Yan¹, K. Matsuda², L. Prokunina-Ols-son¹. 1) Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, US National Institutes of Health, Bethesda, Maryland, USA; 2) Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo, Japan.

A common single nucleotide polymorphism (SNP) rs2294008 in the Prostate Stem Cell Antigen (*PSCA*) gene has been associated with susceptibility to bladder and gastric cancer in genome-wide association studies. Here, we performed a comprehensive analysis of association of rs2294008 in multiple cancer types. A logistic regression analysis was performed for this SNP to calculate odds ratios and 95% confidence intervals controlling for age and sex in Europeans (20K cases representing 11 cancer types and 13K controls) and Japanese (34K cases representing 14 cancer types and 138K controls). We also evaluated *PSCA* expression in relation to this SNP in 9877 samples representing 46 normal tissue types in the Genotype-Tissue Expression (GTEx) project, and in 8510 tumor samples in 33 cancer types in The Cancer Genome Atlas (TCGA). Of cancers explored both in Europeans and Japanese, rs2294008 was significantly ($p < 0.05$) associated with risk of bladder cancer and suggestively, with ovarian cancer. Additionally, rs2294008 was associated with risk of kidney, pancreatic and prostate cancer in Europeans and gastric and lung cancer in Japanese. In GTEx, *PSCA* expression was increased with the risk rs2294008-T allele in 35 of 46 normal tissue types, whereas in TCGA, increased *PSCA* expression was significantly ($p < 0.05$) associated with rs1045574-A allele (a proxy of the risk rs2294008-T allele; $R^2 = 0.99$ and $D' = 0.98$) in 23 out of 33 cancer types based on 5000 permutations and controlling for relevant covariates. A multivariate Cox regression survival analysis was carried out in the TCGA dataset for gene expression categorized in terciles and genotype according to dominant model. We showed that patients with the highest compared to the lowest *PSCA* gene expression (third vs. first tercile) had significantly lower survival of breast, kidney and pancreatic cancer, acute myeloid leukemia, mesothelioma, sarcoma, and thymoma, and better survival of bladder cancer. The risk rs2294008-T allele was associated with decreased survival of kidney cancer, mesothelioma, adrenocortical carcinoma and increased survival of rectum adenocarcinoma. Taken together, these results suggest that the rs2294008-T allele in *PSCA* is a multi-cancer susceptibility variant, which significantly increases the mRNA expression of *PSCA* in multiple normal and cancer tissue types. In addition, both *PSCA* mRNA expression and SNP contribute to survival of patients with multiple cancer types.

752F

Association of race, socioeconomic status and stress with Decipher® genomic classifier risk scores and gene expression: Implications on prostate cancer personalized treatments. A. Grover, S. Greenfield, S. Kaplan, R. Stevenson, E. Uchio. University of California, Irvine, Irvine, CA.

Introduction: The Decipher® genomic classifier risk prediction model was developed to predict early metastasis in prostate cancer patients. It was recently incorporated into the National Comprehensive Cancer Network clinical practice guidelines for prostate cancer prognostication and treatment decisions in patients on active surveillance. Decipher scores range from 0 to 1 and are categorized as: 0-0.45, 0.45-0.6, and 0.6-1 indicating low, intermediate, and high-risk of metastasis, respectively. Various factors (race, socioeconomic status (SES) and stress) might influence these predictions. Additionally, gene expression markers like *EZH2*, *GSTP1* and *TOP2A* (included in Decipher gene expression panel) are associated with DNA methylation changes and elevated risk of metastasis in prostate cancer. Decipher scores and gene expression markers were not previously studied for their association with race, SES and stress levels prospectively and therefore, we intended to study these associations in a prospective cohort study. **Methods:** We analyzed the mean Decipher scores and gene expression markers in different races, SES levels (defined by years of education) and stress (composite measure of self-reported long term stress) using one way analysis of variance and linear regression respectively, with adjustment for confounders (age, race and marital status).

Results: We found that out of 108 prostate cancer patients from five different hospitals in Southern California, there were 56 Whites, 35 Blacks, 5 Hispanics, 6 Asians, 3 mixed race and 3 other races who had mean Decipher scores of 0.39 ± 0.22 , 0.44 ± 0.22 , 0.55 ± 0.10 , 0.59 ± 0.13 , 0.41 ± 0.31 , 0.45 ± 0.32 respectively ($p = 0.05$). There was a 0.5% increase in 5-year predicted metastasis risk with every additional year of education. African Americans and higher SES individuals had higher levels of stress as compared with Whites ($p = 0.006$) and lower SES individuals ($p = 0.005$). Patients with high stress had higher Decipher scores ($p = 0.04$) independent of race and SES. There was a differential gene expression of *EZH2*, *GSTP1* and *TOP2A* in African Americans and patients with high SES and stress. **Conclusions:** These results could provide an epigenetic molecular framework to understand potential influences of race, SES and stress levels on Decipher scores and risk of metastasis in prostate cancer patients. These associations could help formulate predictive models for prostate cancer outcomes and personalized treatments.

753W

COMT modifies vitamin E effects in cancer prevention: Gene-supplement interactions in two randomized clinical trials. K. Hall^{1,2}, J. Buring^{1,2}, K. Mukamal^{2,3}, V. Moorthy¹, P. Wayne^{1,2}, T. Kaptchuk^{2,3}, E. Battinelli^{1,2}, P. Ridker^{1,2}, H. Sesso^{1,2}, S. Weinstein⁴, D. Albanes⁴, N. Cook^{1,2}, D. Chasman^{1,2}. 1) Brigham and Women's Hospital, Boston, MA; 2) Harvard Medical School, Boston, MA; 3) Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA; 4) Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD.

Background: Vitamins are among the most frequently used supplements (48% of US adults). However, little is known about contributions of genetic variation to their efficacy and safety. Multiple pathways link catechol-O-methyltransferase (COMT) to vitamin E (alpha-tocopherol) and cancer. **Methods:** Here we determined if COMT exerted pharmacogenetic effects on cancer prevention in two randomized trials of vitamin E supplementation. Pharmacogenetic effects of common COMT rs4680 (val158met), which encodes a non-synonymous valine-to-methionine substitution, were examined in trial and 10-year post-trial follow-up periods of The Women's Genome Health Study (WGHS, N=23,294), a 10-year vitamin E and aspirin trial with 10 years post-trial follow-up. Results were validated in a case (N=2396)/control (N=2235) subset of the Alpha-Tocopherol Beta-Carotene Cancer Prevention Study (ATBC, N=29,133). The primary outcome was total cancers in the overall period (trial plus 10-year posttrial). Rates of cancer types prevalent in women (colorectal, breast, lung, uterine, and lymphoma/leukemia) were also examined. **Results:** Random-effects meta-analysis of rs4680 genotype strata, in WGHS and ATBC overall periods, revealed differential vitamin E effects compared to placebo: met/met (HR, 0.88; 95% CI, 0.80-0.97; P=0.01); val/met (HR, 0.99; 95% CI, 0.92-1.06; P=0.74) and val/val (HR, 1.18; 95% CI, 1.06-1.31; P=0.002) with a significant COMT-by-vitamin E interaction (Pinteraction=6E-05). Timing of effects differed, with stronger effects in WGHS trial and ATBC post-trial periods. **Conclusion:** Pharmacogenetic analysis of COMT and cancer prevention in two large randomized trials revealed significant COMT-by-vitamin E interactions, such that vitamin E was beneficial among rs4680 met-allele (28%), but not val-allele (22%) homozygotes. These results suggest genetic testing may be helpful in guiding use of over-the-counter supplements like vitamin E for cancer prevention.

754T

Deep targeted tumor sequencing of colorectal cancer cases to study associations of molecular subtypes with clinical, genetic, and lifestyle risk factors. T.A. Harrison¹, S. Zaidi², A.I. Phipps¹, R. Steinfeld¹, Q.M. Trinh², B.L. Banbury¹, P. Hofer³, S. Brezina³, C. Grasso⁴, M. Giannakis⁵, M. Quist¹, C.S. Fuchs⁶, L.A. Garraway⁶, L. Hsu¹, L. Stein², A. Gsur⁸, S. Ogino⁷, S. Gallinger⁸, P.A. Newcomb¹, P.T. Campbell⁹, W. Sun¹, T.J. Hudson², U. Peters¹. *Deep targeted tumor sequencing in a sizeable set of colorectal cancer cases and associations of tumor.* 1) Fred Hutchinson Cancer Research Center, Seattle, WA; 2) Ontario Institute for Cancer Research, Toronto, ON, Canada; 3) Medical University of Vienna, Vienna, Austria; 4) University of California, Los Angeles, CA; 5) Dana Farber Cancer Institute, Boston, MA; 6) Yale Cancer Center, New Haven, CT; 7) Harvard Medical School, Boston, MA; 8) Mount Sinai Hospital, Toronto, ON, Canada; 9) American Cancer Society, Atlanta, GA.

Colorectal cancer (CRC) has considerable intertumoral heterogeneity, and while sequencing has enabled improved characterization and definition of subtypes, the relationship of CRC subtypes to patient survival, genetics, and epidemiologic risk factors has not been comprehensively studied in a sizeable population. We sequenced nearly 2,400 CRC tumors and matched normal tissue from the Colon Cancer Family Registry (CCFR) and Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO) using a targeted AmpliSeq panel of CRC related genes and genomic regions, as well as homopolymer repeats to evaluate microsatellite stability. Mean coverage was 740x for DNA from FFPE tumor and 240x for matched normal samples. We used Strelka and MuTect to call somatic single nucleotide variants, VarScan2, VarDict and Strelka to call indels, and GATK to identify copy number alterations. We identified non-silent mutations in genes belonging to the WNT (77%), p53 (44%), IGF2/PI3K (22%), RTK-RAS (47%), and TGF-beta (26%) signaling pathways. Among the 15% of hypermutated tumors, 39% exhibited non-silent mutations in *MLH1*, *MLH3*, *MSH2*, *MSH6*, and *PMS2* and 41% exhibited non-silent mutations in *POLE* and *POLD1*. In a subset of studies with clinical outcomes data, we used Cox proportional hazards regression to estimate adjusted hazard ratios (HRs) and 95% CIs for the association of mutated genes and pathways with CRC-specific survival (CSS) and overall survival (OS). As expected, OS and DSS were significantly more favorable among individuals with hypermutated vs. non-hypermutated tumors, regardless of *POLE* mutations (HR=0.80, 95% CI: 0.63-1.01, p=0.04 and HR=0.36, 95% CI: 0.22-0.59, p=2x10⁻⁶, respectively). There was no significant association of mutations in WNT, p53, IGF2/PI3K, RTK-RAS, or TGF-beta pathways with survival outcomes. We are further analyzing tumor features defined from our targeted sequencing data in relation to germline genetics and epidemiologic risk factors. Comprehensive molecular characterization of this sizeable set of CRC cases, and analyses relating tumor subtypes to CRC survival and risk factors will help to better elucidate the underlying carcinogenic mechanisms that drive CRC survival and risk.

755F

Application of a joint longitudinal mixed model and survival model to genetic analysis of advanced neoplasia in a colonoscopy screening cohort. E.R. Hauser^{1,2}, X. Qin^{1,2}, T. Hyslop^{1,2}, Y. Li², K.J. Sims³, A. Stone³, M.C. O'Leary¹, T.S. Redding IV¹, Z.F. Gellad^{1,2}, A. Madison¹, B. Sullivan^{1,2}, D. Lieberman^{4,5}, D. Provenzale^{1,2}. 1) Durham VA Health Care System, Durham, NC, USA; 2) Duke University, Durham, NC, USA; 3) Central Arkansas Veterans Health System, Little Rock, AR, USA; 4) Portland VA Medical Center, Portland, OR, USA; 5) Oregon Health and Science University, Portland, OR, USA.

Repeated measures of a phenotype may provide additional information about the presence of germline susceptibility variants for complex traits. Cohort studies following participants over time and making multiple phenotypic measurements provide such data, however, such studies also suffer from loss to follow-up and death of participants. Our goal in this work is to evaluate a Bayesian joint model that fits a mixed random effects model to the longitudinal binary outcomes along with a survival model in a colorectal cancer (CRC) screening study. The Department of Veteran Affairs (VA) Cooperative Study 380 is a longitudinal study of 3121 asymptomatic and otherwise healthy participants aged 50-75 from 13 VA Medical Centers who had a screening colonoscopy between 1994-1997. Participants were followed over 10 years capturing the histopathological outcomes for all colonoscopies or other colon cancer screening tests along with ascertainment of vital status from the National Death Index and the VA Electronic Medical Record (EMR). Advanced neoplasia (AN) at each colonoscopy was defined using pathology reports indicating at least one polyp with tubular adenoma ≥ 10 mm, adenoma with villous histology, adenoma with high-grade dysplasia or invasive cancer. Time to death from all causes or time to last visit was recorded. A total of 612 individuals were genotyped on the Illumina Infinium Omni2.5-8 v1.3 Beadchip with 2,369,593 (99.8%) high-quality markers for analysis. In this study we analyzed the European-American participants and selected SNPs with MAF > 0.05 . We concentrated on 135 candidate SNPs associated with CRC or AN in other studies. We fit a mixed random effects model to all colonoscopies including SNP genotype along with age and two genetic principal components (PCs). We also fit a Cox proportional survival model to time to death including SNP genotype along with age, comorbidity count and 2 PCs. We then fit the Bayesian joint model (R package JMBayes) that joins the mixed and survival models. We compared results from the input longitudinal model to the result for AN risk from the joint model. The effect size estimates were in excellent agreement ($\rho=0.99$); however, the standard errors from the joint model were much smaller potentially resulting in anti-conservative detection p-values. We successfully implemented the joint model in this genetic analysis; however, more work is required to correctly calibrate the results for gene discovery.

756W

Genetic interaction analysis among oncogenesis-related genes revealed novel genes in lung cancer development. Y. Li¹, X. Xiao², I. Gorlov¹, Y. Han¹, R. Hung³, C. Amos² OncoArray Consortium. 1) Biomedical Data Science Department, Dartmouth Geisel School of Medicine, Lebanon, NH 03756; 2) Baylor College of Medicine, Institute of Clinical and Translational Research, One Baylor Plaza, Houston, TX 77030; 3) Prosserman Centre for Population Health Research, Lunenfeld-Tanaka Research Institute, Sinai Health System, Toronto, ON M5T 3L9.

Cancer is driven by the accumulation of many genetic variants that may act epistatically. We hypothesize that there are substantial interactions among oncogenesis-related genes in the early development of cancer. To explore epistasis among oncogenesis-related genes in lung cancer development, we conducted a filtered pairwise genetic interaction analysis among 35,031 SNPs from 2027 oncogenesis-related genes selected from cancer gene catalogs. To provide sufficient power for the proposed analysis, we collected the genotype data from 44,438 individuals, including 20,401 controls and 24,037 cases, from three independent cohorts. We first conducted an efficient epistasis analysis using 'fast-epistasis' option in PLINK and then validated the signals using standard logistic regression interaction analysis in a discovery cohort comprising of 18,401 cases and 14,260 controls from OncoArray lung cancer consortium. Candidate SNP pairs were further submitted to replication study including 5,636 cases and 6,141 controls from two independent genome-wide association studies. SNP pairs with interaction p value $< 5 \times 10^{-10}$ in joint analysis were reported as significant signals. Genetic interactions in *RGL1:RAD51B* (OR=0.44, p value= 3.27×10^{-11} in overall lung cancer and OR=0.41, p value= 9.71×10^{-11} in non-small cell lung cancer), *SYNE1:RNF43* (OR=0.73, p value= 1.01×10^{-12} in adenocarcinoma) and *FHIT:TSPAN8* (OR=1.82, p value= 7.62×10^{-11} in squamous cell carcinoma) gene pairs were identified in our analysis. None of these genes have been identified from previous main effect association studies in lung cancer. The candidate gene pairs with suggestive evidence (p value < 0.05) in replication study were further submitted to gene set enrichment analysis using the Ingenuity Pathway Analysis program. Potential pathways and gene networks underlying the identified genetic interactions were revealed in lung cancer development. The identified adenocarcinoma- and squamous cell carcinoma- specific genetic interactions also suggested differences in epistasis and signaling pathways involved in lung cancer subtype tumorigenesis and enhance our understanding of the molecular mechanisms underlying different subtype of lung cancer diseases. Our results provide evidence that genetic epistasis is an important mechanism involved in lung tumorigenesis and several latent genes may contribute to lung cancer development by interacting with other modifier genes.

757T

A pan-cancer GWAS of metastasis identified a novel metastasis susceptibility loci. M. Machado¹, M. Dean¹, M. Yeager^{1,2}, E. Karlins², N. Chatterjee², N. Freedman¹, W. Huang¹, S. Chanock¹, N. Caporaso¹, L. Mirabello¹. 1) Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, MD, USA; 2) Cancer Genomics Research Laboratory, Frederick National Laboratory for Cancer Research, Leidos Biomedical Research, Inc., Frederick, MD, USA; 3) Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA.

Metastasis is responsible for about 90% of cancer deaths. Discovering new genes and biological pathways in metastasis is critical to understanding its process and developing new strategies for treatment. Although the clinical course for each type of cancer occurs with distinct kinetics and favors unique organ sites, it is possible that shared inherited variation in genes that accelerate or retard cellular activities critical to the complex process of metastasis may impact the clinical onset of metastatic disease for solid tumors in general. The contribution of germline genetic variation to metastasis susceptibility is unknown. Therefore, to determine if germline genetic variation contributes to risk of metastatic disease across solid cancers, we performed the first pan-cancer metastasis GWAS of 13,153 cancer cases, including nine cancer types (lung, breast, ovarian, kidney, pancreas, stomach, testis, endometrial, and prostate). Genotype data was combined for a common set of SNPs from ten distinct GWAS studies conducted at the Division of Cancer Epidemiology and Genetics (DCEG), National Cancer Institute. The clinical phenotype data was carefully harmonized across all studies based on information on the TNM staging system and disease stage to define metastasis. We compared 2,725 cancer cases with advanced stage IV, metastatic disease, to 10,428 cases with early stage, localized disease, using Plink software adjusting for relevant covariates. Case data was pooled for a combined association analysis and also each cancer type was analyzed separately and combined using the statistical tool ASSET, which is able to test association to a subset of the studies and allows opposing effects on different phenotypes. Our preliminary combined results identified a novel locus in a gene desert at 13q33 significantly associated with metastatic disease across cancers (OR=1.72, 95% CI, Padj=2.2x10⁻⁷). Several other compelling loci were significantly associated in subsets of cancers with risk of metastasis and with site of metastatic disease (e.g., metastasis to the liver). We have assembled a replication set of approximately 30k additional solid cancer cases with information on metastatic disease to validate our findings. Our data suggest that novel germline loci are associated with shared metastasis susceptibility across multiple solid cancer types, and if our findings replicate, this could have importance as a prognostic marker with potential to improve patient management.

758F

IFNL4-DG allele is associated with an interferon signature in tumors and survival of African-American men with prostate cancer. L. Prokunina-Olson¹, W. Tang², T. Wallace², M. Yi², C. Magi-Galluzzi⁴, T. Dorsey², O. Onabajo¹, A. Obajemu¹, S. Jordan², S. Loffredo², C. Stephens², R. Silverman¹, G. Stark⁴, E. Klein⁴, S. Ambros². 1) Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics (DCEG), NCI, NIH, Bethesda, MD, USA; 2) Laboratory of Human Carcinogenesis, Center for Cancer Research (CCR), NCI, NIH, Bethesda, MD, USA; 3) Advanced Biomedical Computing Center, Leidos Biomedical Research, NCI, Frederick, MD, USA; 4) Pathology, Lerner Research Institute, and Glickman Urological and Kidney Institute, Cleveland Clinic, Cleveland, OH, USA; 5) Cancer Prevention and Control Program, Lombardi Comprehensive Cancer Center, Georgetown University Medical Center, Washington, DC, USA.

Men of African ancestry experience an excessive prostate cancer mortality which could be related to an aggressive tumor biology. We previously described an immune-inflammation signature in prostate tumors of African-American patients. Here, we further deconstructed this signature and investigated its relationships with tumor biology, survival, and a common germline variant in the interferon lambda 4 (*IFNL4*) gene. We analyzed gene expression in prostate tissue datasets and performed genotype and survival analyses. We also overexpressed *IFNL4* in human prostate cancer cells. We found that a distinct interferon signature that is analogous to the previously described "Interferon-related DNA Damage Resistance Signature" (IRDS) occurs in prostate tumors. Evaluation of two independent patient cohorts revealed that IRDS occurs about twice as often in prostate tumors of African-American than European-American men. Furthermore, analysis in The Cancer Genome Atlas (TCGA) showed an association of IRDS with early recurrence of prostate cancer. To explain these observations, we assessed whether IRDS is associated with a germline variant within the *IFNL4* gene (rs368234815-ΔG) that controls production of IFNL4, a type-III interferon, and is most common in subjects of African ancestry. We show that the *IFNL4* rs368234815-ΔG allele was significantly associated with increased IRDS in prostate tumors and overall survival of African-American patients. Moreover, *IFNL4* expression in 22Rv1 and PC-3 human prostate cancer cells induced IRDS. Our study links a germline variant that controls production of IFN-λ4 to the occurrence of a clinically relevant interferon signature in prostate tumors that may predominantly affect men of African ancestry. Taken together, these observations indicate that IRDS and *IFNL4*-dG allele may have an important function in the tumor biology and survival of African-American patients, and may influence immune therapy outcomes, which should be examined in future studies.

759W

Height and body mass index as modifiers of ovarian cancer risk in 22,588 carriers of *BRCA1* or *BRCA2* mutations: A Mendelian randomization study. F. Qian¹, M.A. Rookus², G. Leslie³, C. Engel⁴, K. Kast⁵, O.I. Olopade⁶, G. Chenevix-Trench⁶, T. Rebbeck⁷, D. Huo⁸ on behalf of The Consortium of Investigators of Modifiers of *BRCA1/2* (CIMBA). 1) Department of Medicine, University of Chicago, Chicago, IL; 2) Department of Epidemiology, Netherlands Cancer Institute, Amsterdam, The Netherlands; 3) Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; 4) Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, Germany; 5) Department of Gynecology and Obstetrics, Technical University of Dresden, Dresden, Germany; 6) Department of Genetics, QIMR Berghofer Medical Research Institute, Brisbane, Australia; 7) Department of Epidemiology, Dana Farber Cancer Institute, Boston, MA, USA; 8) Department of Public Health Sciences, University of Chicago, Chicago, IL, USA.

Background Lifetime risk of ovarian cancer in *BRCA1/2* mutation carriers is substantial, though this relationship is likely modified by lifestyle, reproductive, and genetic factors. Anthropometric characteristics such as height and body mass index (BMI) were shown to increase ovarian cancer risk in the general population, but it remains unclear whether these relationships are causal.

Methods We applied a Mendelian randomization approach to assess the association between height/BMI and ovarian cancer risk in the Consortium of Investigators for the Modifiers of *BRCA1/2* (CIMBA), consisting of 14,676 *BRCA1* and 7,912 *BRCA2* mutation carriers, with 2,923 cases of ovarian cancer. We created a height genetic score (height-GS) using 586 height-associated variants and a BMI genetic score (BMI-GS) using 93 BMI-associated variants. Associations were assessed using Cox models, with adjustments for principal components, birth cohort, mutation status, menopausal status, and country of enrollment. **Results** Height-GS explained 13.4% of the variation in height and the BMI-GS explained 2.6% of variation in BMI. We found that observed height was not associated with ovarian cancer risk (hazard ratio [HR]: 1.06 per 10 cm increase in height, 95% confidence interval [CI]: 0.93 – 1.22). Height-GS was not associated with ovarian cancer risk, either (HR: 1.03 per 10 cm increase in genetically determined height, 95% CI: 0.85 – 1.24). Observed BMI and BMI-GS were not statistically associated with ovarian cancer risk – HR: 1.04, 95% CI: 0.95 – 1.14 and HR: 1.10 per 5 kg/m² increase in genetically determined BMI, 95% CI: 0.86 – 1.42, respectively. However, we found a significant interaction between BMI or BMI-GS and menopausal status ($P_{\text{interaction}} = 0.036$ and 0.002 for observed BMI and BMI-GS, respectively). Higher BMI conferred elevated ovarian cancer risk in premenopausal women (HR for observed BMI 1.26, 95% CI: 1.04 – 1.53; HR for BMI-GS 1.99, 95% CI: 1.24 – 3.20), but not in postmenopausal women (HR for observed BMI 1.00, 95% CI: 0.90 – 1.11; HR for BMI-GS 0.84, 95% CI: 0.62 – 1.12). **Conclusion** We did not observe statistically significant relationships between observed or genetically predicted height or BMI with ovarian cancer risk, although the relative risk estimates were consistent with effect sizes observed in the general population. The significant relationship between BMI and ovarian cancer risk in premenopausal women warrants further research.

760T

Genetic association analysis of longitudinal colonoscopies in a colonoscopy screening cohort in the VA Medical System. X. Qin^{1,2}, T. Hyslop^{1,2}, K.J. Sims¹, A. Stone³, W.C. Carter³, D. Humphries⁴, M.C. O'Leary⁵, T.S. Redding IV⁶, Z.F. Gellad^{1,2}, A. Madison¹, B. Sullivan^{1,2}, D. Lieberman^{5,6}, D. Provenzale^{1,2}, E.R. Hauser^{1,2}. 1) Durham VA Health Care System, Durham, NC, USA; 2) Duke University, Durham, NC, USA; 3) Central Arkansas Veterans Health System, Little Rock, AR, USA; 4) VA Boston Healthcare System, Boston, MA, USA; 5) Portland VA Medical Center, Portland, OR, USA; 6) Oregon Health and Science University, Portland, OR, USA.

Repeated measures of disease activity or recurrences over time may identify a wider spectrum of phenotypic severity across individuals and should allow better characterization of the association between complex disease phenotypes and genetic variants. We hypothesized that multiple outcomes across colonoscopies over 10 years would provide additional power to detect genetic association with the colorectal cancer (CRC) precursor phenotype of advanced neoplasia (AN) when compared to a cumulative, binary outcome assessment of 'ever' vs 'never' having been diagnosed with AN. The Veteran Affairs (VA) Cooperative Study 380 is a longitudinal study of 3121 asymptomatic and otherwise healthy participants aged 50-75 from 13 VA Medical Centers who had a screening colonoscopy between 1994-1997. The biorepository includes 815 individuals with blood, serum, DNA or normal rectal tissue samples, with additional frozen and FFPE samples from removed polyps. Using the VA EMR we identified all colonoscopies performed from the start of the study to 10 years after baseline. AN at each colonoscopy was defined using pathology reports indicating at least one polyp classified as either an adenoma ≥ 10 mm, adenoma with villous histology, adenoma with high-grade dysplasia, or invasive cancer. A total of 612 individuals were genotyped on the Illumina Infinium Omni2.5-8 v1.3 Beadchip with 2,369,593 (99.8%) high-quality markers for analysis. In this study we analyzed the 500 European-American participants and selected SNPs with MAF > 0.05. We analyzed the data using a case-control GWAS as implemented in PLINK with case status defined as AN at any colonoscopy. We then applied a mixed random effects model to include AN at each of the 1220 colonoscopies for the 500 individuals. In both analyses we adjusted for age at baseline and two genetic principal components. We concentrated on 135 SNPs previously associated with CRC or AN. We compared effect size estimates and p-values for the results of the two models for each SNP. The Spearman rank correlation coefficient for the two sets of p-values from each model across all SNPs was 0.86 ($p < 0.01$), suggesting good agreement in the relative ranks of the SNPs. No candidate SNP met a corrected significance level of 0.0003. The top SNPs from the mixed model analysis were rs11886781, rs1493383 and rs6105057 ($p < .02$). Larger genetic analyses of repeated outcomes over time may enhance discovery of genes involved in the progression from AN to CRC.

761F

Interconnections between knock-down genes, cancer types, and landmark genes. *M.B. Rao, S. White, T. Guan, M. Medvedovic, J. Reichard.* Environmental Health, University of Cincinnati, Cincinnati, OH.

An experiment is conducted to examine how a gene knocked-down from a cell line of a cancer tissue affects expressions of a set of landmark genes. 1. Cancers considered (7): skin; lung; kidney; liver; intestine; breast; prostate. 2. Number of genes considered for the knock-down experiment: 2024 3. Number of landmark genes examined: 978 Experiment: Take cancer cell lines derived from a specific tumor type. Select a gene G from the knock-down list. Knock-down the gene G from a subset of the cell lines and the remaining cells serve as controls. From each cell line, determine expressions of the genes from the landmark genes list. Compare (Z-statistic or t-statistic) the expressions of each of the landmark genes using the expression data collected. Using the Z-statistic values, the top 50 genes (up-regulated) are identified by +1, bottom 50 genes (down-regulated) are identified by -1, and the remaining 878 genes are denoted by 0. Repeat the experiment with every cancer cell line in the list and every gene on the knock-down list. Data: Our work begins with the data so collected. The data look like a contingency table of dimensions 2024*7*978. There is a difference. The entries in a contingency table are counts. In our case, the entries are categorical with categories +1, 0, or -1. Finding patterns in such a data is challenging. Building a multinomial regression model for a gene in the landmark list with responses +1, 0, -1 with categorical predictors knock-downs and cancers is computationally challenging. We pursued a number of existing methods and developed new methods to elicit patterns in the three-dimensional data. Methods: 1. Bi-clustering methods: These methods are applicable when we have two-dimensional categorical data. For every cancer fixed, we will identify sub-matrices defined by a subset of genes in the knock-down list and a subset of genes in the landmark list in which every entry is +1, -1, or 0. 2. Tri-clustering methods: This is a new method, which will strive to identify a three-dimensional sub-matrix in which every entry is +1, -1, or 0. 3. Reduce the three-dimensional 2024*7*978 categorical data to two dimensions by summing or counting over the third dimension and then apply two-dimensional methods. In summary: A number of methods, known and new, will be tried to elicit hidden patterns in the data and our presentation will touch upon the trials and tribulations we underwent in the course of our sojourn.

762W

Pathogenic germline mutations in lung cancer in the Clinical Cancer Genomics Community Research Network. *K.L. Reckamp¹, T. Slavin², S.W. Gray², D. Castillo², J. Herzog², S. Sand², S. Mach¹, A. Abing¹, T. Tsou¹, R. Mejia², J. Wang³, S. Tao³, J. Weitzel².* 1) Medical Oncology, City of Hope, Duarte, CA; 2) Clinical Cancer Genomics, City of Hope, Duarte, CA; 3) Sequencing Core, City of Hope, Duarte, CA.

Background: Lung carcinoma (LC) is the leading cause of cancer death and 10-15% occurs in never smokers. Little is known about germline genetic predisposition to LC. **Methods:** Germline DNA from LC patients (N=177) enrolled in Clinical Cancer Genomics Community Research Network was sequenced and analyzed for coding mutations using a custom-designed Agilent SureSelect targeted gene capture kit. The bait design included full exon, 5' and 3' untranslated regions, and splicing region coverage for 789 genes. Genes suspected to be involved in LC susceptibility on the panel included: *EGFR*, *TP53*, *YAP1*, *CLPTM1L*, *FASLG*, *CASP8*, *DLEC1*, *ERBB2*, *ERCC6*, *RASSF1*, and *PPP2R1B*. The panel also included other candidate predisposition and known hereditary cancer genes. After library preparation, reaction products were evaluated using paired-end sequencing at an average of 100x coverage depth. Variants with an allele fraction $\geq 35\%$ and read depth ≥ 5 were annotated using Ingenuity Variant Analysis using American College of Medical Genetics Guidelines. **Results:** Participants characteristics: 62% Caucasian, 23% Asian, 16% Hispanic, and 3% African-American; 73% female; 90 (51%) never smokers. Of the 77 smokers, mean pack year was 24 (median 20 (0.5-50)). Thirty-two percent had between 1-3 first degree relatives with LC, 22% had between 1-4 second degree relatives with LC, 53% had no family history of LC. Somatic mutations included: 94 *EGFR*, 28 *TP53*, 7 *ALK*, 7 *MET*, 4 *ROS*, 3 *BRCA1*, 3 *BRCA2*, 2 *BRAF*, and 1 *KRAS*. Germline cancer predisposition pathogenic/likely pathogenic variants (PV) were seen in 15 individuals (8.5%) in: 1) known LC associated genes: 2 *EGFR* (p.T790M and p.R776H); 2 *TP53*; and, 2) known cancer associated genes: 2 *ATM*, 2 *BRCA2*, 2 *BRIP1*, 2 *MLH1*, *NBN*, *RAD51C*, *PMS2*. Three individuals had PVs in *DLEC1*, a strong candidate predisposition gene for LC. PVs were also identified in other candidate genes: 2 *BLM*, *FANCM*, *MAP3K6*, and 2 *RAD50*. Both *EGFR* germline mutation carriers had multiple relatives with LC. Neither *BRCA2* carriers had breast/ovarian cancer. Both *TP53* carriers had multiple primary cancers but did not meet Li-Fraumeni Syndrome criteria. **Conclusions:** Our cohort had a greater prevalence of hereditary predisposition gene mutations than expected in the general population. Mutations were identified in known LC association genes, established cancer risk genes, and strong candidate LC predisposition genes. Our findings support the need for further LC predisposition research.

763T

Clinical and molecular characteristics of somatic NF1 mutations identified on hereditary cancer multi-gene panels. A. Safonov¹, J. Dolinsky², V. Speare², Z. Powis², M. Pronold², A. Yussuf², C. Gau², H. Laduca², S. Li², K. Allen², A. Yussuf², K. Nathanson¹. 1) Abramson Cancer Center, Hospital of University of Penn, Philadelphia, PA; 2) Ambry Genetics, Aliso Viejo, CA.

Background: Evaluation of patients with clinical features of Neurofibromatosis type 1 (NF1) has led to the longstanding clinical observation of a high rate of somatic mosaicism in NF1. Somatic mutations of NF1 also have been shown to play a role in several tumor types. We explored the clinical and genetic features of patients with NF1 mutations identified through multi-gene panel cancer susceptibility testing (MGPT). **Methods:** Results from a cohort of 118,768 patients who underwent MGPT were queried to identify probands with a NF1 pathogenic/likely pathogenic variant (mutation). Next generation sequencing was performed on panels of 2-67 hereditary cancer genes. Review of test requisition s and available medical records/pedigree was performed. NF1 cases were defined by concordance with NIH criteria. Putative mosaic cases were identified as variant allele frequency (VAF) < 0.30; these cases were subsequently confirmed by Sanger sequencing. **Results:** 276 total patients were identified. 153 (55.1%) patients had a reported clinical diagnosis of NF1; 123 (44.9%) had a mutation in NF1 but were not reported as having NF1. Cancer history was generally similar between groups with a few exceptions. Patients with clinical NF1 were diagnosed with cancer earlier (44.2 vs 53.1 years, $p = 5.1 \times 10^{-8}$), were more likely to have sarcoma (OR 4.8, $p = 0.038$), and; GI malignancies (colon and gastric cancer) (OR 0.12, $p = 0.004$). Putative mosaicism was associated with bilateral breast cancer (OR 4.25, $p = 0.034$). Mutations in other cancer susceptibility genes were present in 7.1% of patients with clinical NF1 and 8.9% of NF1 mutation only cases. When considering patients with multiple mutations, 90% of patients with clinical NF1 and 100% of NF1 mutation only cases met NCCN guidelines for high-risk gene testing. **Conclusions:** Our data did not reveal significance evidence of a "milder" oncologic phenotype in putative mosaic NF1 patients, as compared to heterozygous cases. These results, along with the findings of more frequent bilateral breast cancer suggest that mosaic NF1 may be associated with breast cancer susceptibility. Potential explanations of low VAF of the NF1 gene include somatic mosaicism, circulating tumor, or clonal hematopoiesis of indeterminant potential (CHIP). Additionally, our data suggests MGPT in the NF1 patient population may be reserved for cases in which personal and family history is indicative of an additional hereditary cancer syndrome.

764F

Frequent basal cell cancer development is a clinical marker for inherited cancer susceptibility and DNA repair defects. K. Sarin¹, H. Cho¹, K. Kuo¹, S. Li¹, I. Bailey¹, S. Aasi¹, A. Chang¹, A. Oro¹, E. Epstein², J. Tang¹. 1) Dermatology Department, Stanford University, Stanford, CA; 2) Childrens Hospital Oakland Research Institute, Oakland, CA.

Innate DNA repair mechanisms play a critical role in protecting skin keratinocytes from ultraviolet mutagenesis and skin cancer development. We hypothesized that individuals who develop frequent skin cancers may harbor germline defects in DNA repair genes and have increased predisposition to internal malignancies. We enrolled 61 patients with unusually frequent BCC development seen at Stanford Hospital and Clinics from January 2005 until December 2015, for germline analysis of 29 DNA repair genes. In parallel, a case-control retrospective review was performed to interrogate the association of malignancies with frequent BCC development in a large U.S. medical insurance claims database (Truven), which included 13,264 individuals with 6 or more BCCs during 2007-2011. 19.7% of the frequent BCC cohort (N=12) harbored pathogenic mutations in DNA repair genes: APC, BARD1, BRCA1, BRCA2, CDH1, CHEK2, MLH1, MSH2, MSH6, MUTYH, NBN, and PALB2. Individuals with 6 or more BCCs had an increased risk of other malignancies, with a 3.5-fold increase in frequent BCC cohort and a 3.2-fold increase in the Truven database. Individuals who developed frequent BCCs have an increased prevalence of germline mutations in DNA repair genes and increased malignancy risk. Our data implicate frequent BCC development as an external marker of inherited cancer risk.

765W

Sex differences in gene coexpression across the cancer genome atlas studies.

A.D. Skol^{1,2,3}, R.L. Grossman², B.E. Stranger^{1,2}. 1) Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, IL; 2) Institute for Genomics and Systems Biology, University of Chicago, Chicago, IL; 3) Center for Data Intensive Science, University of Chicago, Chicago, IL.

Sex is an important clinical variable known to be associated with incidence, age of onset, and disease severity in many cancers. Thyroid cancer, for example, is 2.5x more common in women than men, whereas bladder cancer is 4x more common in men. Previous studies have shown sex-specific differences in levels of tumor mRNA and miRNA expression, methylation, and somatic mutations in a subset of The Cancer Genome Atlas (TCGA) cancers. Few studies have examined sex differences in gene coexpression, which can help reveal broader functional consequences than gene-level investigations. Using TCGA cancers, we employed three statistical methods to search for significant coexpression differences between male and female cancer patient tumors. The three methods employed, Weighted Gene Coexpression Network Analysis (WGNCA), Algorithm for the Reconstruction of Accurate Cellular Networks (ARACNE), and Bayesian Biclustering (BicMix), identify gene coexpression networks in distinct ways, and thus not surprising, reveal varied results. However, the genes identified by all methods can be revealing. For example, when studying the 340 liver hepatocellular carcinoma tumor samples (107 female, 233 male), BicMix identified 255 pairs of genes that are coexpressed in one sex, but not the other (112 unique to females patients and 143 unique to male patients). Of the genes identified in coexpression networks unique to females, the ARACNE results revealed that three were transcription factors (TFs) that had significantly fewer coexpressed target genes in males than females, including SIM2 (prostate cancer), MITF (renal cancer, melanoma), and THRB (renal cancer). Of the 143 genes pairs found in uniquely male coexpression networks, ARACNE found that five were TFs with significantly fewer coexpressed target genes in females. These include HLF (lung, renal, head and neck), HNF4A (liver, colon), NR1H4 (liver, colorectal), MLXIPL, and KLF15 (breast, renal). Three of these genes, HNF4A, NR1H4, and MLXIPL were also found in sex-distinct WGNCA defined gene modules. These results suggest that sex-specific differences in loss of regulatory control may play a role in clinical differences observed between men and women. Our hope is that by learning more about the disruption of gene regulation, and other cancer-specific genomic differences between men and women, we will gain an improved understanding of how sex-specific genetic differences leads to disparities in clinical measures and outcomes.

766T

Lipid traits variants and the risk of non-Hodgkin lymphoma subtypes: A Mendelian randomization study.

S.L. Slager¹, G. Kleinstern¹, N.J. Camp², S.I. Berndt³, B.M. Birmann⁴, A. Nieters⁵, P.M. Bracci⁶, J. McKay⁷, H. Ghesquière⁸, Q. Lan⁹, H. Hjalgrim^{9,10}, Y. Benavente^{11,12}, A. Monnereau¹³, S.W. Wang¹⁴, Y. Zhang¹⁵, M.P. Purdue³, A. Zeleniuch-Jacquotte¹⁶, G.G. Giles¹⁷, L.R. Teras¹⁸, A.R. Brooks-Wilson¹⁹, C.M. Vajdic²⁰, E. Kane²¹, N. Caporaso³, K.E. Smedby²², G. Salles²³, V. Joseph²⁴, S.J. Chanock³, C.F. Skibola²⁵, N. Rothman³, J.R. Cerhan³ on behalf of the NHL GWAS Consortium. 1) Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA; 2) Huntsman Cancer Institute and Division of Hematology and Hematological Malignancies, University of Utah, Salt Lake City, USA; 3) Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA; 4) Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; 5) Center for Chronic Immunodeficiency, Freiburg, Germany; 6) University of California San Francisco, Dept of Epidemiology & Biostatistics; 7) International Agency for Research on Cancer, Lyon, France; 8) Centre Léon Bérard, Lyon, France; 9) Department of Epidemiology Research, Division of Health Surveillance and Research, Statens Serum Institut, Copenhagen, Denmark; 10) Department of Hematology, Rigshospitalet, Copenhagen, Denmark; 11) CIBER de Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain; 12) Cancer Epidemiology Research Programme, Catalan Institute of Oncology-IDIBELL, L'Hospitalet de Llobregat, Barcelona, Spain; 13) Centre for Research in Epidemiology and Population Health (CESP), Villejuif, France; 14) City of Hope Beckman Research Institute, Duarte, CA; 15) Yale School of Public Health, New Haven, CT; 16) NYU School of Medicine, New York, NY; 17) Cancer Council Victoria, Melbourne, VIC; 18) American Cancer Society, Atlanta, GA; 19) BC Cancer Agency, Vancouver, British Columbia, Canada; 20) Centre for Big Data Research in Health, University of New South Wales, Sydney, Australia; 21) Epidemiology and Cancer Statistics Group, Department of Health Sciences, University of York, York, UK; 22) Department of Medicine, Solna, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden; 23) Hospices Civils de Lyon, Lyon, France; 24) Memorial Sloan-Kettering Cancer Center, New York, NY; 25) Department of Hematology and Medical Oncology, Emory University School of Medicine, Atlanta, GA, United States.

Background Lipid traits have been suggested as non-Hodgkin lymphoma (NHL) risk factors, but results are inconclusive. Currently, single nucleotide polymorphisms (SNPs) associated with lipid traits explain 8-11% of the total variation in lipid traits in populations of European ancestry. Here we apply a Mendelian randomization (MR) analysis to examine the possibility of causal relationship between lipid traits and the risk of diffuse large B-cell lymphoma (DLBCL), chronic lymphocytic leukemia (CLL), follicular lymphoma (FL), and marginal zone lymphoma (MZL). **Methods** GWAS data from the InterLymph Consortium were available for 2661 DLBCLs, 2179 CLLs, 2142 FLs, 824 MZLs, and 6221 controls. SNPs associated ($P < 5 \times 10^{-8}$) with high-density lipoprotein (HDL, $N=206$), low-density lipoprotein (LDL, $N=160$), total cholesterol (TC, $N=204$), and triglyceride (TG, $N=169$) were used as instrumental variables (IV). SNPs were not in strong linkage disequilibrium ($r^2 < 0.2$). MR estimates for the association between each lipid trait and NHL subtype were calculated using the inverse variance-weighted method, after testing for pleiotropy evidence using MR-Egger regression to examine for violation of the standard IV assumptions. Associations were reported as odds ratios (OR) per standard deviation increase in the genetic risk score along with 95% confidence intervals (CI). A conservative Bonferroni-corrected threshold of $P < 0.003$ was used. **Results** Except for a nominally significant association of TC with DLBCL ($P=0.02$), there was no evidence of violation of the assumptions for any of the associations tested using MR-Egger regression. TG was inversely associated with risk of CLL (OR=0.86, CI: 0.74-0.98) with nominal significance ($P=0.03$). A positive trend was observed for TC with risk of FL (OR=1.07, CI: 0.99-1.17; $P=0.1$), and an inverse trend for LDL with risk of MZL (OR=0.89, CI: 0.79-1.01; $P=0.07$). None of the associations were noteworthy after adjusting for multiple testing. **Conclusions** Our relatively large study did not find evidence of a causal association of these lipid traits and the most common NHL subtypes. The amount of variance accounted by these SNPs for the lipid traits is larger than in many MR studies, and these IVs have been previously linked to colorectal and prostate cancers. Our results are in agreement with most studies that assessed history of hyperlipidemia or statin use with risk of NHL, and suggests positive associations may be due to reverse causality or confounding.

767F

Pathogenic germline variants associated with young onset colorectal cancer.

T.P. Slavin¹, I. Solomon¹, B. Nehoray¹, D. Castillo¹, J. Herzog¹, K. Tsang¹, K. Blazer¹, S.L. Neuhausen¹, K. Yang¹, J. Wang¹, S. Tao¹, S. Manjarrez¹, P. Rice², K. Blankstein³, P. Mora⁴, M. Purdy⁵, C. Hake⁶, D. Margileth⁷, K. Bobolis⁸, J.N. Weitzel¹, *Clinical Cancer Genomics Community Research Network*. 1) Clinical Cancer Genomics, City of Hope National Medical Center, Duarte, CA; 2) Frederick Memorial Hospital, Frederick, MD; 3) Hunterdon Cancer Center, Flemington, NJ; 4) Instituto Nacional de Enfermedades Neoplásicas (INEN), Lima, Peru; 5) Lovelace Cancer Center, Albuquerque, NM; 6) ProHealth Care Research Institute, Waukesha, WI; 7) St. Joseph Hospital, Orange, CA; 8) Sutter Cancer Center, Roseville, CA.

BACKGROUND Colorectal cancers (CRC) generally occur with advanced age; however, diagnoses are rising in the young. Early onset CRC has been associated with genetic predisposition. Our goal was to discover new, and evaluate known, genetic predisposition variants in young-onset CRC.

METHODS Individuals diagnosed with colorectal adenocarcinoma ≤40 years (n=252) were selected from the Clinical Cancer Genomics Community Research Network registry. For those without a previously identified hereditary predisposition mutation from commercial testing next generation sequencing was completed using an Agilent SureSelect hybrid capture research panel of 767 known and candidate cancer predisposition genes. Variants with a call quality of ≥20, a read depth of ≥10, and allele fraction of between 30-70% were classified using the American College of Medical Genetics guidelines. Copy number variation was completed using in silico algorithms. **RESULTS** Average age of diagnosis was 33 years (range 17-40); 56% were female; 52% non-Hispanic Caucasian; 43% stage III/IV; only 11% met Amsterdam I or II criteria. Seventy-one (28%) were identified with an actionable colon cancer predisposition gene likely pathogenic or pathogenic variant (PV) by either commercial (n=55/153) or research testing (16/196) (45 Lynch syndrome, 16 APC, 3 TP53, 2 CHEK2, and 1 BRMP1A). An additional 25 candidate gene PV's were also identified, including 7 (2.7%) in Fanconi Anemia genes: BRCA2, FANCA, FANCD2 (3 mutations), FANCI, and FANCL. **CONCLUSION** In a young CRC cohort, mutations across multiple actionable genes appear common, supporting the use of diagnostic multigene panels. Fanconi Anemia gene PV's also appear enriched in this population. Tumor and case control analyses to support causality are ongoing.

768W

Colorectal cancer susceptibility loci and influence on survival.

N. Song¹, K. Kim², A. Shin^{1,3,4}, J.W. Park^{5,6}, H.J. Jang⁶, J. Shi⁷, Q. Cai⁷, D.Y. Kim⁸, W. Zheng⁷, J.H. Oh⁶. 1) Cancer Research Institute, Seoul National University College of Medicine, Seoul, Seoul, Korea; 2) Division of Epidemiology and Biostatistics, University of Illinois at Chicago School of Public Health, Chicago, IL, USA; 3) Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, Korea; 4) Molecular Epidemiology Branch, National Cancer Center, Goyang, Korea; 5) Department of Surgery, Seoul National University College of Medicine and Hospital, Seoul, Korea; 6) Center for Colorectal Cancer, National Cancer Center, Goyang, Korea; 7) Division of Epidemiology, Department of Medicine, Vanderbilt University School of Medicine, Nashville, TN, USA.

Genome-wide association studies (GWAS) have identified multiple single-nucleotide polymorphisms (SNPs) associated with colorectal cancer risk. To evaluate the potential influence of colorectal cancer susceptibility SNPs on disease prognosis, we investigated whether GWAS-identified colorectal cancer risk SNPs and polygenic risk scores (PRSs) might be associated with survival among colorectal cancer patients. A total of 1,374 colorectal cancer patients were recruited from the Korean National Cancer Center. For genotyping, 30 colorectal cancer-susceptibility SNPs previously identified by GWAS were selected. The Cox proportional hazard model was used to evaluate associations of these risk SNPs and PRSs with disease-free survival (DFS) and overall survival (OS). The prognostic values were compared between genetic and non-genetic models by using Harrell's c index. During the follow-up period (median: 88, 91 months for DFS, OS), 570 DFS (41.5%) and 487 OS (35.4%) events were observed. We found that 5 SNPs were significantly associated with DFS or OS among colorectal cancer patients at $P < 0.05$: rs10936599 at 3q26.2 (MYNN), rs704017 at 10q22.3 (ZMIZ1-AS1), rs11196172 at 10q25.2 (TCF7L2), rs3802842 at 11q23.1 (COLCA1-2), and rs9929218 at 16q22.1 (CDH1). The PRSs constructed using these 5 SNPs were associated with worse survival (DFS: $P_{\text{trend}} = 0.02$ un-weighted PRS, $P_{\text{trend}} = 0.01$ weighted PRS, OS: $P_{\text{trend}} = 3.7 \times 10^{-3}$ un-weighted, $P_{\text{trend}} = 0.02$ weighted PRS). Furthermore, the performance of the genetic prognostic models including survival SNPs were significantly improved compared with the non-genetic models ($P = 0.03$ for DFS, $P = 0.04$ for OS). Our results suggest that several colorectal cancer susceptibility SNPs could be also related to survival by influencing disease progression.

769T

Breast cancer and ovarian cancer risk genes: Meta-analysis of sequencing results of NGS multi-gene panels. M. Suszynska, K. Klonowska, P. Kozłowski. Department of Molecular Genetics, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland.

Background: Germline mutations occurring in highly penetrant genes *BRCA1* and *BRCA2* are responsible for only 16-40% of familial breast cancer (BC) and ovarian cancer (OC) cases, thus several other genes are considered as potential BC and/or OC risk factors. However, the predisposing role of the most of the candidate genes is still elusive, lacking statistically significant and convincing proof. **Methods:** To estimate a reliable BC and OC risk associated with pathogenic variants in the selected candidate BC/OC predisposition genes, a comprehensive meta-analysis of 48 NGS multi-gene panels based studies, analyzing BC and/or OC patients, was conducted. The role of 37 genes was evaluated, comparing in total, the mutation frequency of ~120,000 BC/OC cases and ~120,000 controls, summarized in GnomAD database, what guaranteed strong statistical support with high confidence to most analyzed genes. **Results:** 13 and 11 of analyzed genes were significantly associated with an increased BC and OC risk, respectively. Genes contributed to the (i) high, (ii) moderate, and (iii) low BC risk were as follow: (i) *BRCA1*, *BRCA2*, *CDKN2A*, *PALB2*, *PTEN*, *TP53*, (ii) *ATM*, *BARD1*, *CDH1*, *MSH6*, and (iii) *BRIP1*, *CHEK2**, *MSH2*. In relation to OC, (i) *BRCA1*, *BRCA2*, *BRIP1*, *MSH2*, *MSH6*, *PTEN*, *RAD51C*, *RAD51C*, *TP53*, and (ii) *ATM*, *NBN*, *PALB2* were classified as high-, and moderate-risk genes, respectively. **Conclusions:** Our study defines which of frequently analyzed NGS multi-panel genes, are associated with the low, moderate, and high BC and OC risk as well as excludes the role of others. Mutations in a few genes are attributed to high BC risk, at the level similar to *BRCA2* mutations, especially in *CDKN2A*, earlier not often mentioned gene in the context of BC predisposition. The profiles of genes contributing to either BC or OC risk differ, as several genes much more strongly predispose to OC than BC.

770F

Penetrance for carriers of a DNA mismatch repair gene specific variant.

A.K. Win^{1,2}, J.G. Dowty¹, D.D. Buchanan^{2,3}, J.C. Reece¹, G. Lee¹, G. Moslein⁴, R.W. Haile⁵, F.A. Macrae^{6,7}, I.M. Winship^{2,7}, M.A. Jenkins¹, *International Mismatch Repair Consortium (IMRC)*. 1) Centre for Epidemiology & Biostatistics, The University of Melbourne, Parkville, Victoria, Australia; 2) Genetic Medicine, Royal Melbourne Hospital, Parkville, Victoria, Australia; 3) Colorectal Oncogenomics Group, Genetic Epidemiology Laboratory, Department of Pathology, The University of Melbourne, Parkville, Victoria, Australia; 4) Helios St. Josefs Hospital Bochum-Linden, Linden, Germany; 5) Cedars-Sinai Medical Center, Los Angeles, CA, USA; 6) Colorectal Medicine and Genetics, Royal Melbourne Hospital, Parkville, Victoria, Australia; 7) Department of Medicine, The University of Melbourne, Parkville, Victoria, Australia.

Background: Lynch syndrome is a cancer predisposition caused by inherited mutations in mismatch repair genes *MLH1*, *MSH2*, *MSH6* and *PMS2*, or *EPCAM*. Previous estimates of colorectal cancer risk are averages over hundreds of different mutations in these genes. There is statistical evidence of a substantial heterogeneity in these average risks with some carriers at low risk of colorectal cancer and others at high risk. A possible reason for this risk heterogeneity is that risk varies by specific variants in each gene. **Methods:** We estimated colorectal cancer risk for a single specific variant in *MSH2* using 234 families of *MSH2* c.942+3A>T variant. Pedigree data was provided by clinicians and researchers to the International Mismatch Repair Consortium (IMRC) (<http://www.sphinx.org.au/imrc>). For carriers of this specific variant, we estimated the age-specific cumulative risks (penetrance) and 95% confidence intervals using a modified segregation analysis with appropriate conditioning depending on ascertainment and allowing for risk to vary between families by fitting a polygenic effect. **Results:** The estimated average cumulative risks to age 70 years (95% confidence intervals), were 56% (38%-78%) for male carriers and 45% (28%-67%) for female carriers of *MSH2* c.942+3A>T variant. However, the lifetime risks for different people were estimated to vary widely about these average risks ($p=0.001$). For carriers of this specific variant, 26% of males and 16% of females had colorectal cancer risk less than 20%; and 24% of males and 37% of females had risk greater than 70%. **Conclusions:** These findings suggest that even for a specific variant in a DNA mismatch repair gene, there is a wide range of colorectal cancer risks. This is consistent with the existence of strong modifiers of risk, that if known, could be used to provide personalized risk of colorectal cancer for Lynch syndrome.

771W

Functional informed genome-wide interaction analysis of body mass index on colorectal cancer risk. Z. Xia¹, Y. Su², D. Albanes⁴, H. Brenner^{5,6,7}, D.D. Buchanan^{8,9,10}, P.T. Campbell¹¹, G. Casey¹², A.T. Chan^{13,14,15}, J. Chang-Claude^{16,17}, D.V. Conti¹⁸, S.J. Gallinger¹⁹, W. Gauderman¹⁸, G.G. Giles^{20,21}, S.B. Gruber¹⁸, M. Hoffmeister¹⁸, L. Le Marchand²², L. Li²³, V. Moreno^{24,25}, P.A. Newcomb^{1,2}, P.D. Pharoah²⁶, M.L. Slattery²⁷, C.M. Ulrich²⁸, X. Wang², A. Wolk²⁹, M.O. Woods³⁰, L. Hsu^{2,3}, U. Peters^{1,2}. 1) Department of Epidemiology, University of Washington, Seattle, WA, USA; 2) Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA; 3) Department of Biostatistics, University of Washington, Seattle, Washington, USA; 4) Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA; 5) Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany; 6) Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany; 7) German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany; 8) Colorectal Oncogenomics Group, Department of Clinical Pathology, The University of Melbourne, Parkville, Victoria 3010 Australia; 9) University of Melbourne Centre for Cancer Research, Victorian Comprehensive Cancer Centre, Parkville, Victoria 3010 Australia; 10) Genomic Medicine and Family Cancer Clinic, Royal Melbourne Hospital, Parkville, Victoria 3010 Australia; 11) Behavioral and Epidemiology Research Group, American Cancer Society, Atlanta, Georgia, USA; 12) Center for Public Health Genomics, University of Virginia, Charlottesville, Virginia, USA; 13) Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA; 14) Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA; 15) Clinical and Translational Epidemiology Unit, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA; 16) Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany; 17) University Medical Centre Hamburg-Eppendorf, University Cancer Centre Hamburg (UCC), Hamburg, Germany; 18) Department of Preventive Medicine, USC Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, California, USA; 19) Lunenfeld Tanenbaum Research Institute, Mount Sinai Hospital, University of Toronto, Toronto, Ontario, Canada; 20) Cancer Epidemiology and Intelligence Division, Cancer Council Victoria, Melbourne, Victoria, Australia; 21) Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia; 22) University of Hawaii Cancer Center, Honolulu, Hawaii, USA; 23) Case Comprehensive Cancer Center, Case Western Reserve University, Cleveland, Ohio, USA; 24) Cancer Prevention and Control Program, Catalan Institute of Oncology-IDIBELL, L'Hospitalet de Llobregat, Barcelona, Spain; 25) CIBER Epidemiología y Salud Pública (CIBERESP), Madrid, Spain; 26) Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; 27) Department of Internal Medicine, University of Utah, Salt Lake City, Utah, USA; 28) Huntsman Cancer Institute and Department of Population Health Sciences, University of Utah, Salt Lake City, Utah, USA; 29) Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden; 30) Memorial University of Newfoundland, Discipline of Genetics, St. John's, Canada.

Background Gene-environment (GxE) interaction studies help identify novel loci and biological interactions that give insight to the pathogenesis of colorectal cancer (CRC), as single variants discovered by genome-wide association studies only explain a small fraction of total heritability. Body mass index (BMI), an indicator of obesity and well known risk factor for CRC, is a complex phenotype that might interact with genetic variants to act on CRC risk. As statistical power remains a major concern, incorporating functional genomic information may improve power and help to interpret the underlying biology. **Methods** Genetically determined gene expression levels were tested for interaction with BMI on CRC risk among 26,017 incident cases and 20,692 controls from Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO), Colon Cancer Family Registry (CCFR), and Colorectal Transdisciplinary (CORECT) Study. Using PrediXcan, we imputed the gene expression levels of colon tissues (transverse only) from the Genotype-Tissue Expression (GTEx) Project (n=169) for all genes with sufficient heritability ($\geq 1\%$). We used a mixed-effects model to assess the GxE effects by modeling the interaction between predicted gene expression and continuous measurement of BMI (per 5kg/m²) (fixed effects), and residual variant-specific GxE effects that are not accounted for (random effects). The study was stratified by sex as BMI-CRC associations differ by sex. False discovery rate (FDR) was used to account for multiple comparisons, with FDR<0.2 considered genome-wide significant. **Results** Among the 4839 genes tested, *FOXA1* ($p_{\text{interaction}}=1.94 \times 10^{-5}$), *PSMC5* ($p_{\text{interaction}}=9.96 \times 10^{-5}$) and *CD33* ($p=1.28 \times 10^{-4}$) significantly modified the effect of BMI on CRC risk among males, and *KIAA0753* ($p_{\text{interaction}}=2.80 \times 10^{-5}$) and *SCN1B* ($p_{\text{interaction}}=6.33 \times 10^{-5}$) had a significant interaction with BMI on CRC risk among females. The *FOXA1* gene encodes hepatocyte nuclear factors, which has been found to be highly correlated with some protein expression in breast cancer; *CD33*, a transmembrane receptor expressed on myeloid and lymphoid cells, is the target for the treatment of acute myeloid leukemia. **Conclusions** Incorporating functional information, we discovered some novel genes that may interact with BMI on CRC risk. Independent replication and functional follow up of these findings will add additional evidence.

772T

Estimation of the polygenetic architecture of ten cancers and its implications for future discoveries. Y. Zhang¹, A. Wilcox², Z. Kote-Jara³, R. Eeles^{3,4}, C. Haiman⁵, D. Easton^{6,7}, P. Kraft^{8,9}, J. Simard¹⁰, M. Landi², C. Amos¹¹, J. McKay¹², P. Pharoah^{6,7}, T. Sellers¹³, A. Berchuck¹⁴, M. Jenkins¹⁵, M. Hoffmeister¹⁶, P. Campbell¹⁷, T. O'Mara¹⁸, A. Spurdle¹⁹, D. Thompson⁷, M. Iles¹⁹, M. Bondy²⁰, M. Wrensch²¹, J. Wiencke²¹, M. Purdue², G. Scelo¹², P. Brennan¹², S. Chanock², N. Chatterjee^{1,22}, M. Garcia-Closas². 1) Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD; 2) Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, MD; 3) Division of Genetics and Epidemiology, The Institute of Cancer Research, Sutton, Surrey, SM2 5NG, UK; 4) Royal Marsden National Health Service (NHS) Foundation Trust, London and Sutton SM2 5PT, UK; 5) Keck School of Medicine, University of Southern California, Los Angeles, CA; 6) Department of Oncology, Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge, CB1 8RN, UK; 7) Department of Public Health and Primary Care, Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge, CB1 8RN, UK; 8) Program in Genetic Epidemiology and Statistical Genetics, Harvard T.H. Chan School of Public Health, Boston, MA; 9) Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA; 10) Genomics Center, Centre Hospitalier Universitaire de Québec Research Center, Laval University, Québec City, Québec, Canada; 11) Geisel College of Medicine, Community and Family Medicine, Dartmouth College, Lebanon, New Hampshire; 12) Genetic Epidemiology Group, Genetics Section, International Agency for Research on Cancer, Lyon, France; 13) Moffitt Cancer Center, Tampa, FL; 14) Duke Division of Gynecologic Oncology, Durham, NC; 15) Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Parkville, VIC, 3010, Australia; 16) Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), 69120, Heidelberg, Germany; 17) Epidemiology Research Program, American Cancer Society, Atlanta, GA; 18) Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia; 19) Section of Epidemiology and Biostatistics, Leeds Institute of Cancer and Pathology, University of Leeds, Leeds, UK; 20) Department of Medicine, Section of Epidemiology and Population Sciences, Dan L. Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, Texas; 21) Department of Neurological Surgery and Institute of Human Genetics, School of Medicine, University of California, San Francisco, California; 22) Department of Oncology, School of Medicine, Johns Hopkins University, Baltimore, MD.

Background: Common variants can explain a substantial fraction of familial aggregation of several cancers; however, the proportion of heritability explained by known discoveries in current GWAS varies widely across cancers due to differences in sample sizes. Estimation of underlying effect size distributions of variants can allow an assessment of the diversity of genetic architecture across cancers and a roadmap of future studies. **Methods:** We obtained summary-level data for variant relative risk estimates across GWAS for ten cancer sites with 6,000 to 44,000 cases, plus controls. We applied GENESIS, a method we recently developed, to characterize the effect size distribution of common variants in terms of the proportion of susceptibility variants (polygenicity) and using normal mixture models for their effects. We compared cancers within three groups with similar sample sizes as size can affect the estimation of polygenicity: 1. head and neck, endometrial, glioma, melanoma and renal (<15,000 cases); 2. colorectal and ovarian (15,000-30,000 cases); and 3. breast, prostate and lung (>30,000 cases). **Results:** According to our model, all cancers were highly polygenic with important differences in the degree and nature of polygenicity across sites. In group 1, head and neck cancer had a much higher polygenicity (~6,000 susceptibility variants) than others (~1,000-2,000), and melanoma had susceptibility variants with an average effect size 2-4 fold higher than others. Cancers in group 2 had similar genetic architecture with ~1,000-1,500 variants each. In group 3, prostate cancer had the smallest number of variants (~3,000) with the largest average effect sizes. For all cancers except head and neck, a two-component normal mixture model fitted much better than a single-component model, indicating the presence of clusters of variants with larger effects than the rest. We estimated that sample sizes required to explain 80% of GWAS heritability vary from 100,000 (melanoma) to 1,000,000 (lung) cases and comparable numbers of controls. The projected trajectories for future discoveries showed points of inflections that can inform optimal study designs. Analysis of additional cancers and absolute risk projections are ongoing. **Conclusion:** Increasing the size of GWAS has the potential to explain most of the genetic architecture of cancers. Optimal sample sizes could be achieved in near term for common cancers, while it will take longer for less common cancers.

773F

A mixed-model approach for powerful testing of genetic associations with cancer risk incorporating tumor characteristics. H. Zhang^{1,2}, N. Zhao¹, T. Ahearn², M. Garcia-Closas², N. Chatterjee^{1,3}. 1) Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; 2) National Cancer Institute, Division of Cancer Epidemiology and Genetics, Rockville, MD; 3) Department of Oncology, School of Medicine, Johns Hopkins University, Baltimore, MD.

Background: Cancers of many sites are composed of clinically distinct subtypes characterized by various factors, including histo-pathological characteristics and molecular markers, such as gene expression and somatic mutation signatures. While it has been shown that incorporation of subtype information can potentially lead to discovery of new genetic loci, it is not evident what is the optimal strategy for handling correlated markers, missing data and increased degrees-of-freedom/multiplicity in the underlying tests of associations.

Method: We propose a score test for genetic association using a mixed-effect two-stage polytomous model (MTOP). In the first stage, a standard polytomous model is used to specify a model for genetic association across all possible subtypes defined by the cross-classification of different markers. In the second stage, the subtype-specific case-control odds ratios are specified using a more parsimonious model based on the case-control odds ratio parameter for a baseline disease subtype, and the case-case parameters associated with the individual tumor markers. Further, to reduce the degrees-of-freedom of the model, we specify case-case parameters for exploratory markers using a random-effect model. We account for missing data on tumor markers using an EM algorithm. The score-test and its distribution theory are developed borrowing analogous techniques developed for group-based association tests for rare variants. We illustrated the use of the model using data from the Polish Breast Cancer Study (PBCS) with genome-wide genotyping data on 2,078 breast cancer cases and 2,219 controls; and information on estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2) and grade on the tumors. **Results:** Large-scale simulations across a wide range of realistic scenarios show that MTOP has the correct type one error and robust power. In the PBCS, both MTOP and standard logistic regression methods identified a well-known susceptibility variant in the *FGFR2* locus at $P < 5 \times 10^{-8}$. Further, MTOP identified a second known susceptibility locus in chromosome 11 (*CCND1*) at $P < 5 \times 10^{-8}$ that showed evidence for heterogeneity by ER ($P = 0.005$) and grade ($P = 0.032$). This locus was not genome-wide significant in standard analyses using logistic or polytomous models. **Conclusion:** MTOP is a robust method for testing genetic association in the presence of disease subtype heterogeneity defined by multiple disease characteristics.

774W

Transcriptome-wide association study for pancreatic cancer. J. Zhong¹, L. Wu², I. Collins¹, M. Zhang^{1,3}, S. Lei⁴, J.W. Hoskins¹, A. Jermusyk¹, C.C. Chung⁴, T. Zhang⁴, W. Xiao⁵, E. Abdolalizadeh¹, N. PanScan⁶, N. PanC4⁷, R. Stolzenberg-Solomon⁸, A.P. Klein⁹, B.M. Wolpin⁹, K. Brown¹, X. Shu², S.J. Chanock⁴, S. Olson¹⁰, N. Chatterjee¹¹, G.M. Petersen¹², J. Smith¹³, J. Shi¹⁴, P. Kraft¹⁴, W. Zheng², L.T. Amundadottir¹. 1) Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA; 2) Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, TN, USA; 3) US Food and Drug Administration, Silver Spring, MD, USA; 4) Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA; 5) National Center for Toxicological Research, US Food and Drug Administration, Jefferson, AR, USA; 6) Pancreatic Cancer Cohort Consortium; 7) Pancreatic Cancer Case-Control Consortium; 8) Department of Pathology, Sol Goldman Pancreatic Cancer Research Center, Johns Hopkins School of Medicine, Baltimore, MD, USA; 9) Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA; 10) Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, New York, USA; 11) Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA; 12) Department of Health Sciences Research, Mayo Clinic College of Medicine, Rochester, MN, USA; 13) Department of Medicine, Georgetown University, Washington, DC, USA; 14) Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA.

Pancreatic cancer is currently the third leading cause of cancer-related deaths in United States with a 5-year survival rate of only 8%. Inherited predisposition plays an important role in pancreatic cancer risk. Rare, moderately to highly penetrant mutations in hereditary cancer and pancreatitis genes, identified in families with a high incidence of disease, account for a small fraction of cases. At the other end of the spectrum, common risk variants of low penetrance have been discovered through genome-wide association studies (GWAS). Currently, over 20 common pancreatic cancer risk signals have been identified through GWAS in populations of European ancestry. However, these loci explain a small fraction of genetic heritability for pancreatic cancer, and the causal genes for the large majority of these loci are unknown. Most susceptibility alleles discovered through GWAS are located in noncoding regions of the genome, and are likely to function through allele specific regulation of gene expression. Transcriptome-wide association studies (TWAS) build on this premise by inferring genetically regulated gene expression levels into GWAS datasets to discover genes whose *cis*-regulated expression is associated with complex traits. To apply this approach to pancreatic cancer, we used two reference datasets generated from normal-derived pancreatic tissue samples from individuals of European ancestry, the DCEG/LTG ($n = 95$) and Genotype-Tissue Expression (GTEx v7, $n = 174$) datasets, to build robust predictors for gene expression, using 5 different models (bslmm, blup, lasso, enet and top1) in FUSION. We then applied these predictors to pancreatic cancer GWAS summary statistics from 9,040 pancreatic cancer cases and 12,496 controls (from PanScan I-III and PanC4), to identify associations between genes and the trait. We discovered 4 genes whose genetically predicted expression is significantly associated with pancreatic cancer risk ($FDR < 0.1$), including 3 novel genes. The most significant TWAS gene identified in this study was *ABO* ($P = 1.43 \times 10^{-13}$), whose expression is positively associated with pancreatic cancer risk. Results from the TWAS and ongoing functional experiments for some of the novel genes will be presented. Our findings showcase the utility of integrating multiple datasets to unravel regulatory impact of genetic variants on complex traits such as pancreatic cancer.

775T

Differences in germline *TP53* variant classifications affect the estimates of population prevalence of Li-Fraumeni syndrome: A gnomAD-based analysis. *KC. De Andrade^{1,2}, MN. Frone¹, T. Wegman-Ostrosky^{1,3}, PP. Khincha¹, J. Kim¹, A. Amadou¹, KM. Santiago², FP. Fortes², N. Lemmonnier⁴, L. Mirabello¹, DR. Stewart¹, P. Hainaut¹, LP. Kowalski⁵, SA. Savage¹, MI. Achatz^{1,6}.*
1) Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, MD, USA; 2) International Research Center, A.C. Camargo Cancer Center, São Paulo, Brazil; 3) Division of Research, National Cancer Institute, Mexico City, Mexico; 4) Institute for Advanced Biosciences, Inserm U 1209 CNRS UMR 5309 Université Grenoble Alpes, Site Santé, Allée des Alpes, 38700 La Tronche, France; 5) Department of Head and Neck Surgery and Otorhinolaryngology, AC Camargo Cancer Center, São Paulo, Brazil; 6) Institute for Research and Teaching, Sírio-Libanês Hospital, São Paulo, Brazil.

Recent suggestions of a variable Li-Fraumeni syndrome (LFS) phenotype have raised questions regarding the true disease prevalence. Our previous report on a higher-than-expected population prevalence of pathogenic and likely pathogenic germline *TP53* variants, the primary cause of LFS, was partly attributable to two specific controvertible variants, p.N235S and p.R290H. This study aims to further evaluate germline *TP53* variant prevalence estimates in the gnomAD database, which comprises germline variation from over 130,000 individuals. Variants were selected and classified based on our previously published algorithm. In addition, we provide alternative estimates based on three other classifications databases, ClinVar, HGMD, and the UMD_ *TP53* website. Conservative estimates of pathogenic and likely pathogenic germline *TP53* variant prevalence are within the range of one carrier in 3,500-5,700 individuals. Less stringent classification increased approximate estimated prevalence to one carrier in every 400-900 individuals mainly due to the inclusion of the p.N235S, p.V31I, and p.R290H variants. We observed that the population prevalence of pathogenic and likely pathogenic variants may be higher than previously described. However, it can vary widely due to differences in variant classification. Comprehensive approaches are needed to better understand the interplay of germline *TP53* variant classification, prevalence estimates, cancer penetrance and phenotype of LFS.

776F

Evaluating segregation evidence from multiple extended high-risk pedigrees. *R.G. Waller¹, T.M. Darlington², M.J. Madsen², N.J. Camp².* 1) Biomedical Informatics, University of Utah, Salt Lake City, UT; 2) Psychiatry, University of Utah, Salt Lake City, UT; 3) Hunstman Cancer Institute, University of Utah, Salt Lake City, UT.
High-risk pedigrees (HRPs) are a key design in mapping rare and highly-penetrant genes in Mendelian-like diseases. However, success with the HRP design in complex diseases has been modest, in part because standard methods do not adequately address genetic heterogeneity. Novel methods are needed to re-invigorate HRP designs for gene-discovery in complex diseases. Extended high-risk pedigrees can contain sufficient meioses to gain power for gene mapping as single pedigrees; however, intrafamilial heterogeneity may still exist. We previously described Shared Genomic Segment (SGS), a large pedigree mapping method that identifies subsets of cases within an extended pedigree that share segregating chromosomal regions. Here, we have expanded this strategy to incorporate evidence from multiple large, HRPs while maintaining statistical rigor. Briefly, using a dense, genome-wide map of single nucleotide polymorphisms (SNPs), excessively long runs of identical-by-state SNPs are identified. Chance occurrence is established empirically using gene-drop simulations. All non-trivial subsets of cases are considered within the pedigree. In the multi-pedigree SGS version, the best region/case-subset at each locus is optimized across pedigrees. Statistical significance for SGS segments at the same genomic location across pedigrees are combined using Fisher's method. Genomewide significance thresholds are determined through distribution-fitting and the Theory of Large Deviations based on the multi-pedigree optimized results. We applied our multi-pedigree strategy to 11 extended, multiple myeloma (MM) HRPs. One genomewide significant region was identified at 18q21.33. This region contains one gene, *CDH20*, that facilitates cell-cell adhesions to form cadherin junctions. Three genomewide suggestive regions were identified at 6q26, 18p11.22, and 7p14.3. Together these regions contain 10 genes: *PARK2*, *TWSG1*, *RALBP1*, *PPP4R1*, *RAB31*, *ADCYAP1R1*, *NEUROD6*, *CCDC129*, *PPP1R17*, and *PDE1C*. Our results provide statistically compelling risk loci for MM with evidence from multiple high-risk pedigrees. Further analyses will interrogate exome sequencing in these regions for functional variants in the SGS carriers and high-risk and sporadic cases. **Table. Significant and suggestive regions with evidence from 2 HRPs.**

Site	Region	Length (Mb)	p-value
18q21.33	58,208,260-59,059,262	0.9	6.1x10 ⁻⁹
6q26	162,474,893-162,676,564	0.2	1.3x10 ⁻⁸
18p11.22	9,285,824-9,751,587	0.5	3.3x10 ⁻⁸
7p14.3	31,091,235-31,806,797	0.7	5.4x10 ⁻⁸

777W

A Bayesian framework to detect differentially methylated loci in both mean and variability with next generation sequencing. *H. Xu¹, S. Li², D. Ryu³, V. George¹.* 1) Augusta University, Augusta, GA; 2) University of Texas Medical Branch, Galveston, TX; 3) North Illinois University, DeKalb, IL.

DNA methylation at CpG loci is the best known epigenetic process involved in many complex diseases including cancer. In recent years, next-generation sequencing (NGS) has been widely used to generate genome-wide DNA methylation data. Although substantial evidence indicates that difference in mean methylation proportion between normal and disease is meaningful, it has recently been proposed that it may be important to consider DNA methylation variability underlying common complex disease and cancer. We introduce a robust hierarchical Bayesian framework with a Latent Gaussian model which incorporates both mean and variance to detect differentially methylated loci for NGS data. To identify methylation loci which are associated with disease, we consider Bayesian statistical hypotheses testing for methylation mean and methylation variance using a twodimensional highest posterior density region. To improve computational efficiency, we use Integrated Nested Laplace Approximation (INLA), which combines Laplace approximations and numerical integration in a very efficient manner for deriving marginal posterior distributions. We performed simulations to compare our proposed method to other alternative methods. The simulation results illustrate that our proposed approach is more powerful in that it detects less false positives and it has true positive rate comparable to existing methods.

778T

Functional annotation enrichment based weights for multiple comparisons correction in genome-wide gene-environment interaction studies. *C.B. Haas^{1,2}, X. Wang^{1,2}, H. Chen³, P. Auer³, S. Lindstrom^{1,2}.* 1) Department of Epidemiology, University of Washington, Seattle, WA; 2) Fred Hutch Cancer Research Center, Seattle, WA; 3) Department of Biostatistics, School of Public Health, University of Wisconsin Milwaukee, Milwaukee, WI.

Bonferroni correction assumes equal prior probability of association for all variants tested in a genome-wide association study (GWAS), limiting power to detect true but weak associations. This is especially a concern for gene-environment (GxE) interaction studies, which require even larger sample sizes. Existing methods for multiple comparison adjustment often penalize each variant equally, ignoring any prior information about the biological importance of a specific variant. A weighted Bonferroni adjustment based on enrichment has been shown to increase power compared to standard Bonferroni correction in GWAS, but this has not been applied to genome-wide GxE interaction studies. We performed a case-only genome-wide GxE interaction analysis of age at menarche in 12,366 breast cancer cases from the UK Biobank. We leveraged results from a recent large-scale GWAS of breast cancer in the Breast Cancer Association Consortium (BCAC) to estimate functional enrichment of breast cancer associations across six functional categories: coding, DNaseI Hypersensitive Sites (DHS), intronic, promoter, UTR, and other. Enrichment estimates were calculated using a p-value thresholds of 10^{-8} . Annotation-specific enrichments were 2.58 (coding), 2.14 (UTR), 1.88 (DHS), 1.46 (intronic), 1.37 (promoter) and 0.87 (other). We then applied the annotation-specific enrichments to the results obtained from the GxE interaction scan. In total, we tested 610,758 variants for their interaction with age at menarche. In the UK Biobank results, 4.2% of SNPs were categorized as coding, 22.1% as DHS, 23.5% as intronic, 1.8% as promoters, 1.2% as UTR, and the remaining 47.1% as other. No SNP reached genome-wide significance at a threshold of 5×10^{-8} . We observed nine variants that had a p-value $< 10^{-5}$ of which one was a coding variant, two were located in DHS, three were located in introns and three were other. Adjustment for multiple comparisons using weighted p-values based on enrichment with a threshold of 10^{-8} promoted two additional variants, one intronic variant and one variant located in DHS. As next step, we will pursue replication of these variants in independent datasets. Our results shows the feasibility of incorporating functional annotations into GxE interaction studies to capture potentially interesting variants that do not reach conventional thresholds for genome-wide significance.

779F

Fine-mapping of loci in polyunsaturated fatty acid metabolism genes.

N.K. Khankari¹, J.N. Hellwege^{1,2}, H.J. Murff¹, M.J. Shrubsole¹, Q. Cai¹, J.C. Denny^{2,3}, T.L. Edwards^{1,2}, W. Zheng¹. 1) Division of Epidemiology, Vanderbilt University Medical Center, Nashville, TN; 2) Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN; 3) Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN.

Background: Metabolism of arachidonic acid results in production of the inflammatory eicosanoid, prostaglandin E2 (PGE2), which has been implicated in colorectal cancer. Levels of arachidonic acid, and other polyunsaturated fatty acids are influenced by the fatty acid desaturase (*FADS1* and *FADS2*; Chromosome 11) and elongase (*ELOVL2*; Chromosome 6) genes. In this analysis, fine-mapping of loci involved in polyunsaturated fatty acid metabolism was conducted in order to identify additional variants that may be associated with red blood cell polyunsaturated fatty acids levels. **Methods:** This study utilized resources from the Tennessee Colorectal Polyp Study (TCPS), a colonoscopy-based case-control study. A total of 1,784 white participants were available for this fine-mapping analysis. These participants had available genotype data for 23 variants in loci related to polyunsaturated fatty acid metabolism (i.e., *FADS1*, *FADS2* and *ELOVL2*), and had red blood cell polyunsaturated fatty acid levels measured. Variants were phased and imputed (via SHAPEIT/IMPUTE2) in these loci using the 1000 Genomes Phase 3 Reference Panel. A total of 716 variants were imputed and used for association analysis after filtering on imputation quality (quality score ≥ 0.3). Association tests were conducted using SNPTTEST to identify statistically significant associations with red blood cell polyunsaturated fatty acid levels measured as a percentage of total membrane-bound fatty acid content. **Results:** After Bonferroni correction for multiple comparisons, several variants in the *FADS1* and *FADS2* locus were associated with linoleic acid [$n = 115$, percent variation explained (%VE) = 3.18], arachidonic acid ($n = 53$, %VE = 0.68), α -linolenic acid ($n = 23$, %VE = 7.67), and eicosapentaenoic acid ($n = 5$, %VE = 21.6). No variants reached the Bonferroni threshold for statistical significance for the *ELOVL2* locus. The index variants associated with each of the traits were rs102275 for linoleic acid (T allele, $\beta = 0.61$, $p = 3.6 \times 10^{-11}$), rs174562 for arachidonic acid (A allele, $\beta = -0.18$, $p = 7.3 \times 10^{-7}$), rs187598088 for α -linolenic acid (C allele, $\beta = 10.6$, $p = 9.2 \times 10^{-79}$), and rs189781536 for eicosapentaenoic acid (G allele, $\beta = 27.7$, $p = 2.3 \times 10^{-224}$). **Conclusions:** This study suggests that several variants surrounding the *FADS1* and *FADS2* loci are likely involved in polyunsaturated fatty acid metabolism, in particular for linoleic, arachidonic, α -linolenic, and eicosapentaenoic acids.

780W

Characterization of candidate functional variants in CASP8 from a melanoma locus at chr2q33-q34. H. Kong, J. Choi, T. Zhang, M. Kovacs, A. Vu, M. Xu, K.M. Brown. Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institute of Health, Bethesda, MD.

Previous genome-wide association studies have identified a melanoma susceptibility locus at 2q33-34 near *CASP8*. This region appears to be a multi-cancer risk locus, as a similar association signal has been observed in this region for susceptibility to cutaneous basal cell carcinoma, breast cancer, esophageal squamous cell carcinoma, and chronic lymphocytic leukemia. To perform an eQTL analysis for this locus in a melanoma-relevant tissue, we collected primary cultures of human melanocytes from 106 newborn males, measured transcript levels via RNA sequencing, and assessed genotypes by direct genotyping and imputation. *CASP8* is a significant expression quantitative trait locus (eQTL) gene for this locus, with the risk-associated allele (rs7582362[A]) correlated with lower *CASP8* levels ($P = 4.43 \times 10^{-8}$) in melanocytes. No significant *cis*-eQTL of other genes with the melanoma lead SNP was observed. An association between this locus and *CASP8* is also found in 21 out of 44 the Genotype-Tissue Expression (GTEx) Project tissues, however the direction of effect appears to be opposite that observed in melanocytes for four GTEx tissues (19%). In order to identify potentially functional risk-associated sequence variants with allele-specific transcriptional activities, we performed a massively parallel reporter assay (MPRA) in UACC903 melanoma cells. As a part of a larger MPRA assay, we chose 21 high-LD variants ($r^2 > 0.4$ with the lead SNP, rs7582362) with promoter/enhancer-relevant ENCODE annotations in either primary melanocytes or melanoma cells. MPRA confirmed that two sequence variants had FDR-significant allele-specific transcriptional activity (rs1861270, FDR = 1.40×10^{-6} ; rs3769823, FDR = 3.73×10^{-4}), however rs1861270 failed to replicate in an independent assay, and the observed direction of effect for rs3769823 was opposite the direction of the melanocyte eQTL. We are presently evaluating rs3769821 (FDR = 0.07), which also showed allelic differences in activity as a potential causal SNP for melanoma. We have also performed splice QTL analysis of our melanocyte eQTL data using LeafCutter, and have found that the protective allele of a melanoma-associated SNP (rs10804111[T]; $r^2 = 0.65$ to rs7582362) is correlated with usage of an alternative splice donor for exon 8, resulting in a potentially non-functional *CASP8* transcript. We are presently testing associated sequence variants for an effect on gene splicing using a minigene assay.

781T

Pathway-based analysis of prostate cancer integrating eQTL data and GWAS summary statistics from the UK Biobank. *T.J. Meyers¹, S.R. Rashkin², J.S. Witte^{1,2}.* 1) Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA, USA; 2) Institute for Human Genetics, University of California, San Francisco, San Francisco, CA, USA.

Genome-wide association studies (GWAS) have reported approximately 170 risk variants for prostate cancer. However, the functional mechanisms relating these associations to disease development are not well described. Pathway-based analyses offer potential insight into these mechanisms and confer an advantage when associations of individual variants or genes are modest. Here, we related gene pathways to prostate cancer susceptibility by integrating GWAS summary statistics estimated from the UK Biobank, expression quantitative trait loci (eQTL), and known gene sets using the weighted adaptive sum of powered score pathway-based test ('aSPUPath2'). The UK Biobank study population includes 7,441 prostate cancer cases and 169,970 cancer-free controls, and summary statistics include typed SNPs and those imputed to the Haplotype Reference Consortium (HRC), UK10K, and 1000 Genomes panels. The aSPUPath2 program performs the weighted sum test of the gene variants on gene expression, then tests the association of each gene pathway with the trait by combining associations across the genes. Furthermore, the program is robust to pathways containing genes or SNPs with associations that may be weak and/or in different directions. We obtained eQTL weights from prostate tissue from the GTEx program website and gene sets from the Molecular Signatures Database version 6.1. Results from this analysis will be presented. By leveraging GWAS summary statistics and eQTL data, we can detect biological pathways implicated in prostate cancer development.

782F

Analysis of the *CDKN2A* gene in FAMMM syndrome families reveals early age of onset for additional syndromic cancers. *C.D. Middlebrooks¹, M. Stacey², Q. Li³, C. Snyder⁴, T. Shaw⁵, M. Rendell⁶, C. Ferguson², T. Townley⁴, P. Silberstein², M.J. Casey^{2,6}, S. Lanspa⁷.* 1) National Human Genome Research Institute, Baltimore, MD; 2) Hereditary Cancer Center, Creighton University, Omaha, NE; 3) The Rose Salter Medical Research Foundation, Newport Coast, CA; 4) Internal Medicine Department, Creighton University, Omaha, NE; 5) Department of Hematology/Oncology, Creighton University, Omaha, NE; 6) Department of Obstetrics and Gynecology, Creighton University, Omaha, NE; 7) Internal Medicine Department - Gastrointestinal Division, Creighton University, Omaha, NE.

Familial atypical multiple mole melanoma (FAMMM) syndrome is a familial cancer syndrome that results from mutations in several genes including the *CDKN2A* gene. The syndrome is currently comprised of dysplastic nevi as well as melanoma, breast, pancreatic and lung cancer which are referred to as "concordant" cancers. As families with known *CDKN2A* mutations have been studied longitudinally, clinicians observed an abundance of other cancers or "discordant cancers." However, it was unknown whether these cancers were related to the syndrome. We sought to determine whether these discordant cancers also occur at higher frequencies and at earlier age of onset in carriers than non-carriers. We studied 10 FAMMM syndrome families (N = 1085 individuals) in which a causal mutation in the *CDKN2A* gene was identified. We performed survival analysis as well as a mixed effects cox regression with age at follow-up or cancer event as our time variable and presence or absence of a concordant or discordant cancer as our censoring variable. The survival curves showed a significant age effect with carriers having a younger age at cancer onset for concordant (as expected) as well as discordant cancers than that of non-carriers. The cox regression models were also highly significant (P = 1.24E-27 and P = 5.00E-13 for the concordant and discordant cancers, respectively). As melanoma screening in these high risk families will identify non-melanoma skin cancers but not the other discordant cancers observed here, we sought to determine if there was evidence that new sites should also be commonly screened. Hence, we repeated the discordant cancer analysis with non-melanoma skin cancers removed to ensure that this group was not driving the effects seen. The result was still highly significant. These analyses support the hypothesis that carriers of mutations in *CDKN2A* in FAMMM syndrome have increased risk for early onset of several additional cancer types, suggesting that early screening for these cancers would be beneficial to carriers and clinical guidelines may need to be updated. Based on the data and results, we also recommend that genetic counseling with intensive cancer screening education should be offered to family members beyond the usual first-degree relatives.

783W

Integrative latent unknown cluster assignment using multi-omics data with phenotypic traits. C. Peng¹, J. Wang¹, I. Asante², S. Louie², G. Casey¹, D.C. Thomas¹, D.V. Conti¹. 1) Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA; 2) School of Pharmacy, University of Southern California, Los Angeles, CA.

Epidemiologic, clinical, and translational studies are increasingly generating multiplatform omics data, such as chemical profiles, metabolomes, transcriptomes, genomes and epigenomes. Methods that can integrate across multiple high-dimensional data types while accounting for differential patterns are critical for uncovering novel associations and underlying relevant subgroups. We propose an integrative model to estimate latent unknown clusters (LUCid) that aims to both distinguish unique genotypes and informative biomarkers/omic effects while jointly estimating subgroups relevant to the outcome of interest. Leveraging ideas from the theory of exposure measurement error, the conceptual framework for the integrated analysis views measured omic data as flawed measurements of currently unobserved, latent subgroups. In turn, these subgroups are both influenced by long-term factors (i.e. genotypes or previously encountered exposures) and are related to the outcome. To estimate the overall integrated model, we treat subtype estimation as a missing data problem, where we can use the Expectation-Maximization (E-M) algorithm for estimation with the E-step using the conditional probabilities, and M-step using the expected log-likelihood from the E-step maximized with respect to the parameters. We have implemented this model in an R package *LUCid* (<https://github.com/USCbiostats/LUCid>) capable of analyzing a large number of measured omic data and exposure or genotypes with either a continuous or dichotomous outcome. To accommodate high dimensional omics data, we use regularized regression techniques to obtain sparse estimates for the corresponding coefficients. Simulation results indicate that we can obtain consistent estimates reflective of the true simulated values and we can estimate subgroups and recapitulate subgroup-specific effects. We apply this approach to real data with measured folate biomarkers among colorectal cancer (CRC) cases and controls. We identify four latent subgroups with one noteworthy group having high CRC risks (OR=6.7) and characterized by high levels of pyridoxal and homocysteine (HCY) in the methionine cycle and low levels of 5-MTHF in the folate cycle. The proposed method offers a promising solution to identify factors and estimate molecular subtypes of disease for efficient targeted treatments and successful personalized medicine.

784T

BRA-STRAP: Personalized medicine to precision public health. T. Nguyen-Dumont^{1,2}, K. Tucker³, J. Kirk⁴, P. James⁵, A. Trainer⁶, I. Winship⁶, N. Pachter⁷, N. Poplawski⁸, S. Grist⁹, D.J. Park¹⁰, F. Hammett¹¹, M. Mahmood¹², H. Tsimiklis¹³, J. Steen¹⁴, D. Theys¹⁵, J. Burke^{1,2}, E. Thompson³, I. Campbell⁶, A. Willis³, J. Weitzel¹¹, F. Couch¹², J. Leary¹³, J.L. Hopper¹⁴, D.E. Goldgar¹⁵, A. Morrow¹³, M.C. Southey^{1,2}. 1) Precision Medicine, Monash University, Clayton, VIC, Australia; 2) Genetic Epidemiology Laboratory, The University of Melbourne, VIC, Australia; 3) Prince of Wales Hospital, Sydney, NSW, Australia; 4) Familial Cancer Service, Westmead Institute for Cancer Research, University of Sydney, NSW, Australia; 5) Peter MacCallum Cancer Centre, VIC, Australia; 6) Royal Melbourne Hospital, VIC, Australia; 7) King Edward Memorial Hospital, WA Australia; 8) Adult Genetics Unit, South Australian Clinical Genetics Service, SA, Australia; 9) SA Pathology, Flinders Medical Centre, SA, Australia; 10) Melbourne Bioinformatics, The University of Melbourne, VIC, Australia; 11) City of Hope, Duarte CA, USA; 12) Mayo Clinic Cancer Center, Rochester, MN, USA; 13) Westmead Institute for Medical Research, NSW, Australia; 14) Center for Epidemiology and Biostatistics, School of Population and Global Health, The University of Melbourne, VIC, Australia; 15) Huntsman Cancer Institute, University of Utah, Salt Lake City, UT, USA.

BRA-STRAP is performing targeted-sequencing of 24 genes commonly included on panel tests for breast cancer predisposition in 30,000 Australian women of all ages across the cancer risk spectrum, affected and unaffected with breast cancer, and their families. Contributing to *BRA-STRAP* are i) large research cohorts of aging populations and of women above population risk, and ii) the clinical genetics community and women found negative for *BRCA1* and *BRCA2* mutations following testing in an Australian Familial Cancer Centre (FCC) over the last 2 decades. *BRA-STRAP* is also engaged with other similarly designed studies set outside of Australia (e.g. BRIDGES and CARRIERS). Data on this scale represents the spectrum of genetic variation observed in these genes and exemplifies the opportunities and challenges for realizing precision medicine and precision public health for breast cancer. To date, *BRA-STRAP* has identified over 240 women carrying genetic variants in genes supported by Australian best practice guidelines: 133/2,744 (4.8%) in the Australian Breast Cancer Family Registry (ABCFR) and 106/4,911 (2.1%) in the participating FCCs. These genetic variants include pathogenic variants in *BRCA1* and *BRCA2* (as classified by *ENIGMA*), truncating variants in *PALB2* and *TP53*, and *ATM*:c.7271T>G (p.Val2424Gly). Women found to carry these variants have the opportunity to receive this information, free of charge, to inform their risk management, treatment options and enable predictive testing in family members. In the other genes, *BRA-STRAP* has so far identified 49/1,189 (4.1%) and 238/4,911 (4.8%) carriers of truncating variants in affected women participating in the ABCFR and women attending an FCC, respectively. A comprehensive analysis of the data generated to date is underway. A large number of genetic variants will need further study to determine their clinical utility. These include variants in genes that are not well characterized outside a specific cancer syndrome or for which there is currently insufficient evidence to be called *bona fide* breast cancer susceptibility genes, and variants of uncertain clinical significance in known breast cancer susceptibility genes. For instance, we have observed 85/7,655 (1.1%) carriers of rare (MAF <0.1%) missense variants in *PALB2* that are predicted to be deleterious by multiple prediction tools. More information is urgently required so these variants can be included in evidenced-base clinical management and genetic counselling.

785F

Fine-mapping of 150 breast cancer susceptibility regions to single-variant resolution and prediction of target genes. J. Beesley¹, L. Fachal², J. Allen², S. Karr², D. Barnes², H. Aschard³, V. Kristensen⁴, A. Lemaçon⁵, A. Antoniou², J. Simard⁶, P. Kraft⁶, D.F. Easton², G. Chenevix-Trench¹, A.M. Dunning². 1) QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia; 2) Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK; 3) Program in Genetic Epidemiology and Statistical Genetics, Harvard T.H.Chan School of Public Health, Boston, MA, USA; 4) Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital Radiumhospitalet, Oslo, Norway; 5) Genomics Center, Centre Hospitalier Universitaire de Québec Research Center, Laval University, Québec City, QC, Canada; 6) Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA.

Genome-wide Association Studies (GWAS) have identified breast cancer risk alleles at 150 genomic regions. Here, we use dense genotyping and imputation to fine-map all these regions to identify independent risk signals, reduce the number of candidate causal variants (CCVs), and predict target genes. Using stepwise multinomial logistic regression with genotype data for >90,000 cases and 95,000 controls from 147 studies contributing to the Breast Cancer Association, Discovery, Biology, and Risk of Inherited Variants in Breast Cancer, and Investigators of Modifiers of *BRCA1/2* Consortia, we identify 352 independent risk signals comprising 7,394 CCVs, defined as variants with marginal *P* values within two orders of magnitude of the most significant SNP. We find 28 signals with a single CCV, and 96 signals with less than ten. Sixty six signals are more strongly associated with risk of estrogen receptor positive (ER+) and 29 with ER- disease. Enrichment analysis reveals significant overlap of CCVs in cisomes of transcription factors (TFs) associated with estrogen signalling with far larger effect sizes observed for ER+ variants. Significant overlap of ER+ CCVs is further supported by localisation of variants to corresponding TF recognition motifs. We use Integrated expression Quantitative trait and In Silico prediction of GWAS Targets (INQUISIT) to predict target genes finding 1,355 unique targets for 93% signals. The 155 most robustly predicted targets are strongly enriched for known breast cancer driver genes ($P < 10^{-6}$) but also include many genes with no previously known role in breast cancer. Modelling the number of signals per gene by negative binomial versus Poisson regression suggests that a larger than expected number of genes are impacted by >1 independent signal at proximal genomic regions (likelihood ratio test $P = 10^{-3}$). Pathway analysis using high confidence target gene predictions reveals specific themes for ER+ candidate targets including Ephrin signalling, and FOXM1 signalling for ER- risk genes. These results provide logical prioritization of risk loci for functional follow-up to unequivocally identify target genes and elucidate the molecular and cellular mechanisms underlying the genetic basis of breast cancer. 1. Michailidou *et al.*, Association analysis identifies 65 new breast cancer risk loci. *Nature* 2017.

786W

Judging genetic loci by the company they keep: Comparing network-based methods for biomarker discovery in familial breast cancer. H. Climente-González^{1,2,3}, C. Lonjou^{3,1,2}, F. Lesueur^{3,1,2}, D. Stoppa-Lyonnet^{4,5}, N. Andrieu^{3,1,2}, C.-A. Azencott^{2,1,3}, GENESIS investigators. 1) Institut Curie, PSL Research University, F-75005 Paris, France; 2) MINES ParisTech, PSL Research University, CBIO-Centre for Computational Biology, F-75006 Paris, France; 3) INSERM, U900, F-75005 Paris, France; 4) Service de Génétique, Institut Curie, Paris, France; 5) Inserm, U830, Université Paris Descartes, Paris, France.

In this work we explore and evaluate the ability of network-based methods to find genetic markers for complex diseases. The three surveyed algorithms (PINBPA, dmGWAS and SConES) leverage on prior knowledge to overcome the lack of power of GWAS studies. We explore the GENESIS case-control dataset, composed of 1,282 French women affected by familial breast cancer, and 1,272 unaffected women from the general population. Genotyping data were obtained using the iCOGS SNP array (200,000 SNPs). Network methods found biomarkers that showed high, albeit non-significant, scores of association with the disease, while being connected in a network that represents the underlying biology. Most of them are in or close to genes well known to be related to breast cancer susceptibility (such as *FGFR2*, *TOX3*, *NEK10*). We compared our results to those reported for European populations by the Breast Cancer Association Consortium (BCAC), who have performed the most exhaustive meta-GWAS of breast cancer. Network methods retrieved 23 and 7% of BCAC significant SNPs and genes respectively with a hundredth of the sample size. On top of that, they recovered 211 SNPs and 12 genes candidates to be specific for the French population. While all three algorithms successfully leverage prior knowledge on our data set, they differ in their methodology, aim, and results. PINBPA and dmGWAS require combining SNP scores into gene scores. They output a list of ranked subnetworks of the provided gene-gene network, where strongly associated genes are interconnected through other genes, if not directly. By contrast, SConES returns a single list of SNPs and works directly on genotypes, using a SNP-based network. We explored three SNP-network architectures: genome structure, gene membership and a network derived from the protein-protein interaction network we used for PINBPA and dmGWAS. Two challenges all three tools have in common are the choice of their parameters, which can notably alter their output, and the visualization of the results. In summary, dmGWAS, PINBPA, and the three variants of SConES all give different importance to gene-gene interactions. We found that although the multiplicity of subnetworks from PINBPA and dmGWAS makes interpretation more challenging, it provides useful complementary information. We conclude that dmGWAS and PINBPA are interesting for generating hypotheses about the etiology of the disease, while SConES' strength is biomarker discovery.

787T

Phenotype-driven burden testing for identifying rare and common variants influencing the somatic mutation landscape in 2,642 cancer genomes. F.M. De La Vega^{1,2}, S. Nohzadeh Malakshah³, S. Flygare³, B. Moore³, E. Hernandez³, E. Kiruluta¹, C.D. Bustamante², M. Yandell³. 1) Fabric Genomics, Oakland, CA; 2) Stanford University School of Medicine, Stanford, CA; 3) University of Utah School of Medicine, Salt Lake City, UT.

Whole-genome and whole-exome sequencing (WGS/WES) are increasingly used to identify rare genetic variants involved in the etiology of complex disease and novel drug targets. Analysis of such datasets requires specialized statistical methods that need to deal with allelic and locus heterogeneity and the scale of the data. To address these needs, we have developed an integrated pipeline that includes an improved version of the VAAST algorithm, which implements a gene-level burden test statistic and provides empirically calculated *p*-values for every gene-hit. The pipeline starts from variant calls in VCF format, bundles clean-up and normalization steps, scores variants with a new algorithm (VVP), performs the VAAST burden test, and then optionally runs Phevor, a phenotype-driven method that leverages prior knowledge from the Human Phenotype and Gene Ontologies to increase power. To validate our pipeline, we spiked known disease variants from ClinVar into WES data from 2,000 controls from unrelated studies. Variants were spiked on a range of allele frequencies, carrier rates, inheritance models, and allelic heterogeneities. We also benchmarked our software against the commonly used SKAT-O method. Our methods consistently identified known disease-genes with genome-wide significance and delivered more power than SKAT-O, alongside fast compute times. We applied our pipeline to a dataset of 2,642 cancer patients across 39 cancer types, where their germline and cancer genomes were sequenced as part of the Pan-Cancer Analysis of Whole-Genomes ICGC project. Since this cohort does not include non-cancer control samples, for analysis we split samples with different mutational signatures, different germ layers, or used external WGS controls. Previous analysis uncovered germline variants in *APOBEC3B* associated with somatic mutational signatures derived from this enzyme activity; however, even when the analysis was limited to 298 candidate cancer genes, other analyses were underpowered. Our pipeline is applied to replicate the previous finding with *APOBEC3B*, and additional associations with *BRCA1/BRCA2* in ovarian and breast cancer, respectively, and *POLK* in lymphoma, without limiting to candidate genes. Our results show an increased power to identify genes associated with complex traits in the presence of allelic and locus heterogeneity with modest sample sizes, suggesting our pipeline could be valuable to analyze large prospective and case/control cohorts.

788F

Whole exome sequencing in lung cancer families identifies significantly linked loci on multiple chromosomes. A.M. Musolf¹, H. Sun², B.A. Moiz¹, C.W. Pikielny³, M. de Andrade⁴, D. Mandal⁵, C. Gaba⁶, P. Yang⁴, Y. Li³, M. You⁷, R.K. Wilson⁸, E.Y. Kupert⁹, M.W. Anderson⁷, A.G. Schwartz³, S.M. Pinney¹⁰, C.I. Amos¹¹, R. Govindan¹², J.E. Bailey-Wilson¹. 1) Computational and Statistical Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Baltimore, MD; 2) Laboratory of Medical Genetics, Harbin Medical University, Harbin, China; 3) Geisel School of Medicine, Dartmouth College, Lebanon, NH; 4) Mayo Clinic, Rochester, MN; 5) Department of Genetics, Louisiana State University Health Sciences Center, New Orleans, LA; 6) Department of Medicine, University of Toledo Dana Cancer Center, Toledo, OH; 7) Medical College of Wisconsin, Milwaukee, WI; 8) Institute for Genomic Medicine, Nationwide Children's Hospital, Columbus, OH; 9) Karmanos Cancer Institute, Wayne State University, Detroit, MI; 10) Department of Environmental Health, University of Cincinnati College of Medicine, Cincinnati, OH; 11) Baylor College of Medicine, Houston, TX; 12) Division of Oncology, Washington University School of Medicine, St. Louis, MO.

Lung cancer (LC) kills more people in the United States than any other cancer; it is responsible for 25% of all cancer deaths nationwide. LC risk is known to be increased by a variety of environmental factors, especially tobacco smoking, however the disease has a significant genetic component as well. We performed whole exome sequencing (WES) on 204 individuals from 25 extended families. These families have a strong history of LC and are highly aggregated for the lung cancer phenotype. Individuals from each family were chosen for sequencing to maximize information in a genetic linkage analysis. The WES was recalled with PICARD and GATK and standard quality controls on depth, genotype quality, missingness and Mendelian errors were performed. We also calculated identity-by-descent (IBD) values to confirm all familial relationships. After quality control, we were left with approximately 500,000 SNVs and indels per family for subsequent analysis. We performed two-point parametric linkage analysis on LC and the WES data including phenotype data for both sequenced and unsequenced family members (493 total) using an autosomal dominant model with disease allele frequency of 1%, 10% penetrance for carriers and a 1% phenocopy rate. Our analyses were affected only analyses, meaning individuals were that were affected were coded as cases, all others were coded as having unknown (or missing) phenotypes. We performed linkage analysis both on the individual markers (SNVs or indels) and on individual genes. The gene-based markers were constructed from haplotypes of the rare variants (MAF < 0.01) located in each gene. Two-point linkage was then performed on each gene-based marker. Initial analyses found multiple significant gene-based linkages. The highest HLOD scores centered on chromosome 1q. The significant HLOD scores clustered around known cancer genes, including the implicated genes *OBSCN* and *RYR2*. The signals also centered on five genes in the neuroblastoma breakpoint family (NBPF) genes, a cluster of recently duplicated genes that have been found to be associated with a variety of cancers, including lung cancer. We have an additional 200 individuals from additional extended families that are currently being recalled and will be ready for analysis shortly. These individuals will be analyzed as well to increase power on our current findings and possibly identify new linked regions.

789W

Multivariable Mendelian randomization study of adiposity measures, puberty onset, and prostate cancer risk. C. Zhang¹, L. Kachuri¹, S. Rashkin¹, T.J. Meyers¹, K.M. Walsh², J.S. Witte¹. 1) University of California, San Francisco, San Francisco, CA; 2) Duke University, Durham, NC.

Observational studies suggest links between adiposity measures and prostate cancer risk, including lower risk with elevated body mass index (BMI) and, paradoxically, higher risk with greater BMI-adjusted waist hip ratio (whradBMI). However, Mendelian randomization studies do not support the effect of genetically elevated BMI or whradBMI on prostate cancer. Furthermore, some of the genetic variants associated with the two adiposity measures are pleiotropic, i.e. associated with both BMI and whradBMI, which violates the assumptions of traditional univariable MR. A previously described protective effect of delayed puberty onset on prostate cancer risk has been attributed to reduced androgen levels, which has also been suggested as the underlying mechanism for the protective effect of obesity on prostate cancer risk. To explore these hypotheses, we performed a *multivariable* MR analysis of the independent causal effects of BMI, whradBMI, and puberty onset (based on the Tanner scale) using GWAS summary statistics for prostate cancer risk from the UK Biobank (cases=7,441), and previously published GWAS data on the three risk factors. Results from univariable MR show a weak protective effect of increased BMI on prostate cancer risk (odds ratio (OR)=0.87, 95% confidence interval (CI)=0.71-1.05), a weak protective effect of later puberty onset (OR=0.76, 95% CI=0.47-1.22), and a weak risk effect of higher whradBMI (OR=1.25, 95% CI=0.93-1.67). In a multivariable MR analysis that included all three factors, the inverse associations for BMI and puberty onset strengthened, reaching statistical significance: OR=0.77 (95% CI=0.63-0.95) and OR=0.77 (95% CI=0.60-0.99), respectively, while the multivariable MR estimate for whradBMI revealed a non-significant protective effect of greater BMI-adjusted waist hip ratio on prostate cancer risk (OR=0.84, 95% CI 0.65-1.09). These results suggest that BMI and puberty onset exert independent causal effects on prostate cancer but does not exclude the possibility that both operate via the effect of diminished androgen exposure during discrete phases of life. Multivariable MR analysis can be useful in evaluating the role of multiple exposures on disease risk when these exposures share overlapping genetic architecture.

790T

Re-interpreting PAM50: Intrinsic tumor dimensions, prognosis and response to paclitaxel. N.J. Camp^{1,2}, M.J. Madsen¹, J.H. Valera³, E. Carrasco³, R. Caballero³, M. Martin⁴, P.S. Bernard^{1,5}. 1) Huntsman Cancer Institute, University of Utah, Salt Lake City, USA; 2) Department of Internal Medicine, University of Utah, Salt Lake City, USA; 3) Spanish Breast Cancer Research Group, GEICAM, Madrid, Spain; 4) Department of Medical Oncology, Hospital General Universitario Gregorio Marañón, Madrid, Spain; 5) Department of Pathology, University of Utah, Salt Lake City, USA.

Background. We recently showed that PAM50 gene expression data can be represented by five quantitative, orthogonal, multi-gene breast tumor traits. These novel tumor 'dimensions' were superior to intrinsic subtypes for differentiating tumors from high-risk breast cancer pedigrees, indicating potential to represent underlying genetic susceptibilities and biological pathways. Here we explore the prognostic and predictive utility of these tumor dimensions in a sub-study of GEICAM/9906, a Phase III randomized prospective clinical trial of the addition of paclitaxel in breast cancer. **Methods.** Intrinsic tumor dimensions PC1 – PC5 were calculated using pre-defined coefficients. Univariable and multivariable Cox proportional hazards (PH) models, Kaplan-Meier curves and log-rank tests for disease-free survival (DFS) were used to identify associations between dimensions and prognosis or response to the addition of paclitaxel. **Results.** Dimensions PC1 and PC5 were associated with DFS (Cox PH $p=6.7 \times 10^{-7}$ and $p=0.036$), remaining significant after correction for standard clinical-pathologic prognostic characteristics. Both dimensions were selected in the optimal multivariable model, together with nodal status and tumor size (Cox PH $p=1.4 \times 10^{-12}$). Interactions with treatment were identified for PC3 and PC4. Response to addition of paclitaxel was restricted to tumors with low PC3 and PC4 (log-rank $p=0.0021$). Women with tumors high for PC3 or PC4 showed no survival advantage. **Conclusions.** Intrinsic dimensions illustrate clinical utility beyond standard clinical-pathologic characteristics for prognosis and predicting chemotherapy response in locally-advanced breast cancer patients. The consideration of expression data as quantitative tumor dimensions offers new potential to identify clinically-important patient subsets, increase power in clinical trials and improve precision medicine.

791F

Mosaic Y loss and incident cancer risk in the UK Biobank. *M.J. Machie-la¹, E. Lofffield¹, W. Zhou^{1,2}, M. Yeager^{1,2}, S.J. Chanock¹, N.D. Freedman¹.* 1) Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD; 2) Cancer Genomics Research Laboratory, Frederick National Laboratory for Cancer Research, Leidos Biomedical Research, Inc., Frederick, MD, USA.

Mosaic loss of the Y chromosome (mLOY) in peripheral leukocytes is a somatically occurring event in which a fraction of leukocytes lacks the entire Y chromosome. The frequency of mLOY is higher among cigarette smokers and increases with age, with mLOY commonly observed in men over age 50. mLOY may reflect poor genomic maintenance as well as clonal imbalances in normal immune function, making mLOY an attractive candidate marker for cancer risk. Some, but not all, previous studies have suggested that mLOY may be associated with higher risk of solid tumors. Therefore, we investigated the relationship between mLOY and incident cancer in a large sample of 207,603 cancer-free men from the UK Biobank. We identified mLOY by scanning for deviations in genotyping array log R intensity ratios (LRR) across the male-specific chromosome Y region, classifying mLOY into low (median LRR < -0.15, ~10%) and high (median LRR < -0.40, ~24%) proportions of cells affected. In total, 14,947 men developed an incident cancer during follow-up with prostate (N=4,345), colon (N=848) and lung (N=783) being the most common incident non-cutaneous solid tumors. Overall, we detected low proportions of cells with mLOY in 3,358 (1.6%) men and high proportions of mLOY in 524 (0.3%) men. Adjusting for age, comprehensive smoking history, ethnicity, body mass index, alcohol consumption and education, we found no association of mLOY with overall cancer incidence using either the low or high mLOY thresholds ($HR_{low}=1.05$, 95% $CI_{low}=0.96-1.16$, $P\text{-value}_{low}=0.31$; $HR_{high}=1.22$, 95% $CI_{high}=0.99-1.51$, $P\text{-value}_{high}=0.07$), but we did observe associations for two smoking-related cancers: bladder ($HR_{low}=1.67$, 95% $CI_{low}=1.07-2.59$, $P\text{-value}_{low}=0.02$) and lung ($HR_{high}=1.97$, 95% $CI_{high}=1.19-3.24$, $P\text{-value}_{high}=0.008$) cancer. Stronger associations were observed without adjustment for smoking, suggesting that smoking may be an important confounder. Alternatively, mLOY may mediate, at least to some extent, the association of cigarette smoking to these cancers. In summary, mLOY was not a strong risk factor for incident solid tumors in adult men in the UK Biobank. Modest associations were observed for bladder and lung cancer, although these findings may reflect residual confounding by smoking.

792W

Association testing in admixed populations using aggregated data. *E. Ziv¹, D. Hu¹, J. Weitzel², S. Neuhausen².* 1) Medicine, University of California San Francisco, San Francisco, CA; 2) Beckman Research Institute, City of Hope, Duarte, CA.

Modern sequencing studies often depend on the use of aggregate data for controls, such as ExAC and gnomAD. However, a significant limitation of this approach, particularly in admixed populations, is that genetic ancestry is not known for individual participants and therefore there is no method to adjust for ancestry differences between cases and controls. Here we develop a method to adjust for ancestry differences between cases and controls using aggregate data. The components required include (a) an estimate of the overall genetic ancestry in the aggregate dataset/s and (b) an estimate of the allele frequency in each ancestral population for each allele tested. Methods for the former have previously been developed and can be implemented as long as a fraction of the variants have also been genotyped in ancestral populations. For the latter, estimation of allele frequencies in the ancestral populations, we develop a method that uses several admixed populations to infer the ancestral allele frequencies. We show that as long that for any N admixed populations derived from M ancestral populations with different proportions, as long as $N \geq M$, we can infer ancestral allele frequencies for any variant typed in all N populations. This inferred allele frequencies can then be used as part of a likelihood ratio test conditional the overall ancestral proportions in cases and controls. We demonstrate the method avoids bias due to population stratification using simulations. We also utilize the approach to test for allele frequency differences between Latina breast cancer cases and Latino controls from ExAC and gnomAD.

793T

TWO-SIGMA-geneset: TWO-component SInGle cell Model-based Association method for Gene Set Testing. E. Van Buren, Y. Li, D. Wu. University of North Carolina at Chapel Hill, Chapel Hill, NC.

Gene set testing (GST) methods allow omnibus testing for a pre-specified set of genes, often by assigning an overall p -value to the set of genes. By considering such sets of genes, GST methods can improve statistical power and help to understand how genes in a common pathway jointly regulate biological processes. There are many GST methods for bulk RNA-seq data, however no such methods are specifically designed for use with single-cell RNA-seq data, which often exhibits excess zeros and overdispersed counts compared to traditional bulk RNA-seq data. Here, we propose TWO-SIGMA-geneset to conduct gene set testing using single-cell RNA-seq data. We utilize our previously developed TWO-SIGMA, a flexible TWO-component SInGle cell Model-based Association method. The first component of TWO-SIGMA models the “drop-out” probability with a mixed effects logistic regression, and the second component models the (conditional) mean read count with a log-linear negative binomial mixed-effects regression. TWO-SIGMA accommodates excess zeros and overdispersed counts, and can, using random effects, explicitly account for the possible dependency introduced by measuring multiple single cells from the same sample. The focus of TWO-SIGMA-geneset is on conducting competitive gene set testing, in which the genes in a given set are compared to the remaining collection of genes. Each gene in the set is assigned an individual statistic, such as a likelihood ratio statistic from the joint test of both components, produced from TWO-SIGMA. These gene-level statistics are then combined into one set-level statistic, with options including the average, median, or sum of the ranks. Finally, the set-level statistic is compared to the complement set of genes using, for example, the t -test, Wilcoxon rank-sum test, or a Monte Carlo p -value. Simulation studies show that type-I error is well controlled in a variety of representative scenarios. Power is improved over state-of-the-art methods, including CAMERA, for a variety of scenarios consistent with real single-cell RNA-seq data, including when varying: gene set sizes, the scales of the differential expression, the proportion of drop-out events, or the presence of individual-level random effects. Extensions to allow for inter-gene correlation are discussed. We demonstrate our method with an application to a real dataset consisting of 36,849 single cells from 10 mice and gene sets from the Molecular Signatures Database.

794F

Genome partitioning with region-based analysis adapted to linkage disequilibrium (LD) structure in lung cancer case-control association analysis. M. Brossard¹, S.A. Kim², D. Roshandel³, Y. MacMillan¹, A.D. Paterson^{3,4}, R. Hung^{1,4}, Y.J. Yoo², S.B. Bull^{1,4}. 1) Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, ON, Canada; 2) Seoul National University, Seoul, Korea; 3) Hospital for Sick Children Research Institute, Toronto, ON, Canada; 4) Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada.

To improve region discovery in genome-wide association studies of complex traits, we investigate a novel haplotype block detection method (BigLD, Kim 2017) to identify non-overlapping, quasi-independent LD block regions for subsequent region-based testing. When individual effects are too small to be detected by conventional single-SNP analysis, global region tests can improve power to identify genomic regions that harbor multiple SNPs with joint effect on disease risk (Yoo 2017). We investigated a region-based approach in chromosome 8 data from the Toronto site (1359 cases/949 controls) of the ILCCO large scale meta-analysis of lung cancer case-control studies (McKay, Nat Gen 2017), based on a reported *NRG1* susceptibility locus. We applied BigLD on 335K 1000Genomes (1KG)-imputed SNPs ($MAF \geq 0.05$) and assessed sensitivity to the BigLD input parameters and the source population. Because cases may have different LD structure than controls, we recommend that control samples be used for block construction; blocks identified in controls showed good agreement with recombination hotspots and with blocks identified in 1KG European samples. In controls, BigLD identified 5266 blocks with a variable number of SNPs per block (median:15, range: 2-1071); pairwise SNP correlation was high within- and low between-blocks (mean: 0.60 and 0.33 respectively). We found BigLD to be robust to clustering parameter values within the neighborhood of the default value. In each of the BigLD blocks, we compared three regression-based tests to assess global association with lung cancer risk: the generalized Wald test with df equal to the number of SNPs in region and two reduced df tests adapted to local correlation structure. The multiple linear combination test (MLC, Yoo 2017) constructs a constrained test statistic from regression coefficients with df equal to the number of correlated SNP clusters. The regression of SNP principal components (PC80, Gauderman 2007) first reduces the number of regression variables and then constructs a Wald test. MLC and PC80 reduce df of region-based tests (reduction of Wald-df by mean factor of 3 and 5 respectively (ranges 1-13, 1-65)) and provide similar test results (Pearson correlation of $-\log_{10}$ P-values of 0.84). Because LD varies across the genome, LD-defined region size is right-skewed and high-df global tests tend to become conservative as region sizes increase, suggesting the need for additional dimensional reduction methods in very large regions.

795W

iMEM: A novel data statistical method for integrating mutation, gene expression and methylation with an application to a paired oral cancer dataset. S. Das¹, P.P. Majumder², I. Mukhopadhyay¹. 1) Human Genetics Unit, Indian Statistical Institute, Kolkata, West Bengal, India, PIN 700108; 2) National Institute of Biomedical Genomics, P.O.: N.S.S., Kalyani, West Bengal, India, PIN 741251.

Alteration in regulatory activities of a gene may be attributed to various genetic and epigenetic factors, like, mutation in DNA, methylation at various CpG sites etc. Traditional methods of analysing omics data separately would incur loss of essential information that might be beneficial in differentiating between the genetic code of a tumor and a normal sample. Recently data integration through joint analysis of various omics data is focused to obtain robust inferences. But it is imperative to understand and include the available downstream information and use more practical assumptions in the model development, in order to capture the diversity in the genomic profiles of patients. Using the idea of analysis of covariance, we propose a general statistical method that identifies the regulators affecting the expression of a gene, through integration of three omics data viz. somatic mutation, gene expression, and methylation. In our model, we consider various downstream categories of mutation classified according to their functional severity that is induced by change in the amino acid and subsequent alteration in the protein. On the other hand, methylation values at adjacent CpG sites are strongly correlated and this correlation weakens as the methylation sites move apart. Our model summarises hundreds of methylation values in the vicinity of the gene incorporating the correlation structure among the methylation sites. Thus our method includes a biological insight from mutation and methylation to assess its effect on the gene expression based on tumor-normal paired sample. We also derive the asymptotic distribution of our test statistic to calculate p-values fast. We perform extensive simulations to analyse the performance of our method mainly in terms of size and power of the test. The type I error rate, calculated on the basis of the distribution of the test statistic, is controlled at 5% level of significance. Our proposed method is powerful and consistent as the power of the test increases with the increase in sample size. Our method is robust as it shows consistent results for different genetic models that we consider in our simulation study. Application to real dataset identifies some significant genes, of which a few are already reported. Based on the results obtained from simulated and real data on oral cancer, our method looks very promising in differentiating gene expression profiles of patients through integrated analysis of three omics data.

796T

Polygenic prediction of breast cancer: Comparison of genetic predictors and implications for screening. K. Läll^{1,4}, M. Lepamets¹, M. Palover¹, T. Esko², A. Metspalu¹, N. Tõnisson¹, P. Padrik^{2,3}, R. Mägi¹, K. Fischer¹. 1) Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu, Tartumaa, Estonia; 2) Institute of Clinical Medicine, University of Tartu, Estonia; 3) Cancer Center, Tartu University Hospital, Tartu, Estonia; 4) Institute of Mathematics and Statistics, University of Tartu, Estonia.

As the heritability of breast cancer (BC) is known to be around 20-30%, but only small fraction of BC cases has rare inherited component identified, the ability to predict BC polygenic forms may improve current screening and risk assessment strategies. We compare four genetic risk scores (GRS): two (GRS₇₀, GRS₇₅) have been recommended by previous publications and include less than 100 single-nucleotide polymorphisms (SNPs) and two newly developed GRSs (GRS_{ONCO} and GRS_{UK}) based on available summary statistics of two large-scale GWAS analyses. Two cohorts - Estonian biobank with 32557 women and a subset of 46,984 women from the UK Biobank cohort- were used to study associations between GRSs and BC. We show that combining GRS_{ONCO} and GRS₇₅ into one metaGRS produces the score with strongest association with prevalent disease in both cohorts. Women in the top 5% metaGRS category have 4.2 times higher hazard (95% CI 2.8-6.2, $p = 8.1 \times 10^{-13}$) for incident breast cancer than women in the lowest 50% of metaGRS and almost 3 times higher hazard (HR=2.73, 95% CI 1.92-3.9) of developing BC compared to the rest of the population in the Estonian Biobank. While the overall BC incidence in the Estonian population reaches 5% by the age of 70, women in the highest 5% of the metaGRS reach the same risk level almost 20 years earlier. We also show that GRSs that are based on summary statistics of different GWASs are only moderately correlated with each other and discuss the possible implications of that to personalized genetic feedback.

797F

Quantifying uncertainty in causal biological networks. *E.A. Martin, A.Q. Fu.* Bioinformatics and Computational Biology, University of Idaho, Moscow, ID.

Introduction: Many methods have been developed to infer a causal network under the principle of Mendelian randomization (PMR), by jointly analyzing genotype and gene expression data. In particular, MRPC, our recently developed algorithm, is one such method and outperforms similar methods in inference accuracy. Due to noise in the data, however, not all edges in the inferred network have the same level of confidence, and none of the existing methods quantifies the uncertainty in the inferred edges. **Methods:** With the causal network inferred by MRPC, we develop a genetic algorithm (GA) to estimate for each inferred edge the probability of the (directed or undirected) edge being correctly identified. The fitness is the log likelihood of the graph given the orientation of the edges and the data. Our algorithm starts by generating a population of individuals, whose 'DNA' are the inferred edges, with their directions randomly assigned. In each generation of the GA a winner and loser are selected from two subsets of the population. The winner is the individual who has the highest fitness among the first subset of individuals. The loser is the individual who has the lowest fitness among the second subset of individuals. The fitness is the log likelihood of the graph given the orientation of the edges and the data. After the winner and loser are selected the DNA of the winner is copied, mutated (reversing edge direction), and then replaces the DNA of the loser. The GA is run until most of the individuals in the population have a similar likelihood. We next count the number of times each direction occurs in the population and divide it by the population size. This gives a probability of an inferred edge being correct. **Conclusions:** We have simulated data under multiple graphs with different numbers of nodes and edges. In the initial results the probabilities of the edges are what we would expect even accounting for Markov equivalent graphs, which have different topologies but the same likelihood. With these promising results we will continue to expand the algorithm to include more data types (eg genetic variants, different distribution models, etc.) and test it on larger graphs.

798W

Using Bayes model averaging for admixture mapping. *L.C. Moss, X. Sheng, C.A. Haiman, D.V. Conti, African Ancestry Prostate Cancer Consortium and the Ellipse Game-On Consortium.* Preventive Medicine, University of Southern California, Los Angeles, CA.

Admixture mapping is typically performed in recently admixed populations to detect genetic risk loci for diseases that have differential risk by ancestry. Admixture mapping relies on the deviation of local ancestry, alleles inherited from one ancestral population at a particular locus, from global ancestry, the proportion of the entire genome inherited from one ancestral population. The approaches most commonly used are case-only or case-control models which respectively compare ancestry deviations in cases only or test the effect of the deviation between cases and controls on the disease outcome. Although the case-only approach has potentially more power than the case-control approach, the case-only approach is prone to spurious associations when deviations of ancestry are not solely due to noise within controls. In this study, we use Bayes model averaging from estimates obtained from case-only and case-control models to yield a novel statistical test for admixture mapping. The novel approach offers more power than a case-control method while remaining robust in scenarios where the case-only method is most susceptible to false positives. We use simulations to demonstrate that our approach detects admixture signals with increased power and robustness over case-control and case-only methods and illustrate the approach in the African Ancestry Prostate Cancer (AAPC) and the Ellipse Game-On Consortia.

799T

Classification of variants without prior penetrance estimates and the effect of ascertainment bias on classification. *J. Ranola, G. Tsai, B. Shirts.* University of Washington Laboratory Medicine, Seattle, WA.

Classification of variants of unknown significance is an increasingly important aspect of personalized medicine. Current methods require a penetrance estimate for a gene prior to classification of a new variant in that gene. When available, these penetrance estimates are often based on highly deleterious variants binned together under the assumption that they all have the same pathogenic effect. New variants are then classified according to whether they more closely resemble this set of pathogenic variants or benign ones. The assumption that all pathogenic variants in a gene have the same effect is problematic when classifying new variants. Requiring the penetrance of a similar variant in order to determine the pathogenicity of a new variant is paradoxical. The problem of penetrance assumptions can be exacerbated by the fact that initial penetrance estimates are often from selected families leading to biased over estimates of penetrance. These assumptions put limits on the type of variants that can be correctly classified. Variants with low to moderate risk could never be correctly classified regardless of the amount of data. Likewise, protective variants would be misclassified. As a solution to this problem, we present a new variant assessment method for hereditary cancer risk genes that uses family data and SEER cancer incidence data, which is readily available. The method, based on the Poisson binomial model, compares incident cancers in a family with expected incident cancers while utilizing information from both genotyped and ungenotyped relatives. The model may facilitate information sharing and ongoing improvement of estimates as identifiable pedigree information no longer needs to be shared to combine variant-specific penetrance estimates. Our method, like previous methods, takes into account the ascertainment bias inherent in the proband and can also account for known bias in ascertainment of other relatives. This method can also be used to detect bias in a control dataset when making initial risk estimates. We apply this method to classify hypothetical variants in simulated families across a range of risk while varying the percent of individuals genotyped. We also evaluate actual families tested for hereditary cancer University of Washington database and show the magnitude of clinical bias in ascertaining cancer families.

800F

Rare variant tests for association in affected sib pairs with application to breast cancer. *R.G. Romanescu¹, G. Gos¹, I.L. Andrulis^{1,2}, S.B. Bull^{1,3}.* 1) Lunenfeld-Tanenbaum Research Institute, Toronto ON, Canada; 2) Department of Molecular Genetics, University of Toronto, Toronto ON, Canada; 3) Dalla Lana School of Public Health, University of Toronto, Toronto ON, Canada.

Recent progress in sequencing technologies has made it possible to investigate the role of rare variants (RVs) in disease etiology. Compared to case-control designs, tests based on affected sib pairs (ASPs) can be more powerful, as RVs tend to be enriched in families. With the aim of discovering novel breast cancer susceptibility genes, we conducted a whole exome sequencing (WES) study of sisters with early onset breast cancer (proband age-at-onset <45 years). The sisters, ascertained from high-risk families in the Ontario Familial Breast Cancer Registry, screened negative for known susceptibility genes (including *BRCA1/2* and *CHEK2*1100delC* variants), thereby increasing the chances of finding rare familial mutations. Candidate genes identified in the ASP study are deemed eligible for potential validation in large case-control consortia. Our objective is to develop tests of association for RVs at the single region, as well as the pathway level, a feature which is expected to improve power and biological interpretability. The key idea is that RVs will be found preferentially on haplotypes shared identical by descent (IBD) versus haplotypes not shared IBD. We construct two test statistics which measure the departure of the joint distribution of RV allele count for the pair and their IBD state (0-2) under ascertainment compared to expectation under simple Mendelian transmission. To mimic the motivating study design, we evaluate our methods on simulated sister pair data under early age-at-onset ascertainment. We assume a genetic model that reflects a scientific hypothesis of locus heterogeneity, i.e., that different families tend to harbor different susceptibility variants. We demonstrate good type I error properties for the tests under a null scenario in which the ASPs do not share causal RVs captured in the test statistic, because they are sporadic cases or because they share a causal RV not included in the test. Under the alternative hypothesis that 15% of RVs captured in the test are causal, we show that the proposed tests are as, or more powerful than existing methods for testing association in ASPs. Finally we apply our tests to sets of variants classified as rare in our sister pairs data, and compare the results to those of a conventional filtering approach we applied in these data that prioritized variants based on mutation type, gene function, and pathway membership.

801W

Y-chromosome lineages and its contribution to prostate cancer in Mestizo population from Mexico City. *E. Álvarez-Topete¹, R. Gómez¹, L. Torres-Sánchez², S.J. Trujillo-Cáceres².* 1) Departamento de Toxicología, Cinvestav, Mexico City, Mexico; 2) Instituto Nacional de Salud Pública, Centro de Investigación en Salud Poblacional, Cuernavaca, Morelos, Mexico.

Prostate cancer (PC) is the second most common malignant neoplasm in males worldwide and the fifth leading cause of death for cancer in men. Risk factors such as age, diet, smoking, and the genetic background are keystones in this malignant hyperplasia. Y-chromosome genes play a meaningful role in male-specific organs such as prostate glands and have been related to incidence and PC progression. Certain paternal lineages have also been related to PC development and prognosis in different populations. Nonetheless, the Y-chromosome haplogroups have been poorly studied in this context in the Mexican population, which presents a complex genetic architecture related to the historical migrations of the last 500 years. The tremendous diversity within the Mexican paternal lineages provides a unique opportunity to investigate the contribution of Y-chromosome haplogroups and the PC in a multi-ethnic population. Therefore, the non-recombining region of Y-chromosome (NRY) of 533 unrelated mestizo men (150 PC incident cases and 383 population controls) was genetically characterised with seventeen-Y-STRs through capillary electrophoresis. Each individual signed an informed consent validated by the Ethics Committee of National Public Health Institute. The Native American lineage (Q) was the most prominent (~ 42%), followed by R1b (~ 30%) and E1b1b (~ 8%) haplogroups. Of note, the E1b1b carriers exhibited a marginal association of twice risk to PC (OR = 1.80; CI95% 0.965 - 3.998, $p = 0.06$) than the rest of the haplogroups. Stratifying the carriers of this lineage by age-onset and PC aggressiveness degree, a significant association was found (OR = 2.30; CI95% 1.06 - 4.97, $p = 0.03$) for late-onset PC (≥ 60 years old) and Gleason ≤ 6 at PC diagnosis (OR= 4.96; CI95% 1.6 - 15.5, $p = 0.006$). Our results suggest a possible contribution of E1b1b with PC, reinforcing previous findings in Ashkenazi Jewish ancestry, and African ancestry where this haplogroup is highly prominent. Our preliminaries results suggest a potential role of Y-chromosome in late-onset PC with a slowly growing grade in the Mexican population.

802T

A survey of germline mutations with epithelial ovarian cancer in Japanese patients. *A. Abe¹, I. Imoto², M. Nishimura¹, M. Irahara¹.* 1) Tokushima university, Tokushima, Japan; 2) Aichi Cancer Center Hospital, Aichi, Japan.

Objectives: In Japan, the number of BRCA testing is increasing annually. But, till recently, genetic counselling for epithelial ovarian cancer (EOC) has not been generalized according to hereditary breast and ovarian cancer (HBOC). We investigated the awareness survey and the rate of gBRCA1/2 mutation with EOC, regardless of their family histories. **Methods:** 127 patients with EOC who had treated at our hospital were enrolled for 2years. After genetic counselling, Three patients refused result disclosure of gBRCA1/2 testing because of no relatives or depression. For 108 patients, analyses for gBRCA1/2 test were performed. DNA was extracted from whole blood, and all coding exons and their flanking intron regions of 11 genes included BRCA1 and BRCA2 were sequenced with next-generation sequencing in our laboratory. **Results:** 99 patients (80%) wished for gBRCA1/2 result disclosure, and other 21 patients did not wish to know that result because of worrying. Seventeen of the 108 patients (15.7%); 6 patients in gBRCA1 mutation, 10 patients in gBRCA2 mutation, and one patient in gRAD51C mutation. 10 of 16 with gBRCA1/2 mutation were high grade serous carcinoma, two; adenocarcinoma, and 2; others. **Conclusion:** Almost patients with EOC wanted to know genetic risk after genetic counselling. The rate of gBRCA1/2 mutation in tested Japanese EOC patients with an inherited risk, regardless of a family history, was 15.7%. This finding indicates that the cooperation of genetic counselling with other department would be important for patients with ovarian cancers.

803F

Quantifying the varying selective processes of pre-tumor and tumor evolution from high-coverage sequence data. A. Ang Houle^{1,2}, K. Skead¹, D. Soave¹, P. Mehanna^{1,3}, M. Agbessi^{1,3}, V. Bruat^{1,3}, J. Bartlett⁴, P. Boutros^{1,5}, S. Wright⁶, L. Stein^{1,2}, P. Awadalla^{1,2}. 1) Computational Biology, Ontario Institute for Cancer Research, Toronto, Ontario, Canada; 2) Molecular Genetics, University of Toronto, Toronto, Ontario, Canada; 3) Montreal Heart Institute, Montreal, Quebec, Canada; 4) Diagnostic Development, Ontario Institute for Cancer Research, Toronto, Ontario, Canada; 5) Medical Biophysics, University of Toronto, Toronto, Ontario, Canada; 6) Ecology and Evolutionary Biology, University of Toronto, Toronto, Ontario, Canada.

Cancer progression is an evolutionary process: somatic mutations can confer selective advantages to certain cells, eventually causing them to abnormally proliferate. Important aspects to consider when studying the evolutionary processes of cancer include the accurate characterization of somatic mutations. While positively selected alleles may lead to selective sweeps within the cancer population, and reduce the diversity of somatic variation, damaging mutations may occur and be swept along during the process. The signature of the selective sweeps may hinder the detection of negative selection within the cancer cell population. Our ability to capture these evolutionary signatures is dependent on sample size, which we infer from the depth of sequence coverage in the cases of tumor evolution within a single patient. Average sequencing depths are often insufficient to detect mutations found at low variant allele frequency within a population of cancer cells, and sequencing at higher depths of coverage allows for a more complete picture of the full evolutionary spectrum. Here, we aim to accurately identify evolutionary processes in 1735 cancer and pre-cancer samples sequenced with a minimal coverage of 100x from 5 cancer types. Using the read depth of somatic mutations to quantify the allelic frequency of a mutation within each population of cancer cells, we employ an Approximate Bayesian Computation framework to quantify the proportions of beneficial and deleterious mutations, and the timing of driver and passenger events throughout the life of a cancer. Owing to the absence of programmed recombination allowing for the exchange of alleles, non-recombining evolutionary models may be applied to cancer cells, and the richness of high coverage cancer data provides a unique opportunity to accurately quantify evolutionary processes in non-recombining systems. Using these same models, we can capture the impact of ploidy changes on evolutionary inferences. Interestingly, by applying these models, and stratifying the mutational spectrum within patient samples, we capture non-neutral evolutionary processes previously undetected in tumors. We show varying rates of damaging as well as beneficial mutations accumulating across patients within many tumor types. Our work demonstrates the heterogeneity of evolutionary processes within a cancer type, and further outlines the possibility of dubious inferences of cancer evolutionary processes caused by low sequencing depth.

804W

Genetic testing for the Tier 1 genomic conditions in a population-level cohort. C.L. Neben, A. Zimmer, W. Stedden, J. Van den Akker, A.Y. Zhou. Color Genomics, Burlingame, CA.

An estimated 2 million people in the United States are at increased risk for the Tier 1 genomic conditions hereditary breast and ovarian cancer (HBOC), Lynch syndrome (LS), and familial hypercholesterolemia (FH). Recently, there has been momentum in the field towards offering genetic testing for these conditions at the population-level because such screening leads to early detection and intervention. Color Genomics has been partnering with organizations to offer genetic testing for these conditions as an employer-sponsored, voluntary, confidential employee health benefit. Here, we report the frequency and spectrum of pathogenic variants in HBOC (*BRCA1*, *BRCA2*), LS (*MLH1*, *MSH2*, *MSH6*, *PMS2*, *EPCAM*), and FH (*APOB*, *LDLR*, *PCSK9*) in employees who represent an average-risk population-level cohort. This cohort was almost equal parts men (5735, 47.5%) and women (6351, 52.5%), the majority of which were under age 40 (6854, 56.7%). Nearly half of individuals were non-Caucasian (5566, 46.1%) with Asians accounting for 25.1% (3028). Only 0.7% (90/12,073) of individuals reported a personal history of breast, ovarian, colorectal, or endometrial cancer, and 21.3% (342/1605) reported having high cholesterol. These demographics are in stark contrast to typically studied multi-gene panel cohorts of women, over age 50, with high clinical risk. In this cohort, we identified 122 pathogenic variants in HBOC or LS genes in 12,073 individuals for a pathogenic frequency of 1.01% and 7 pathogenic variants in FH genes in 1605 individuals for a pathogenic frequency of 0.44%. Thus, we estimate an overall pathogenic frequency for the Tier 1 genomic conditions of 1.45%. This frequency is double the estimated prevalence from previous studies, but closely matches that from the recent Geisinger MyCode cohort (1:78, 1.28%). In our cohort, of those with a pathogenic variant in HBOC or LS genes, only 2.5% (3) of individuals had a personal history of cancer, which is likely attributed to a younger age. Furthermore, only 50.0% (61) of individuals would have met NCCN criteria for genetic testing, which is consistent with recently published findings in population-level testing. Taken together, these data suggest that twice as many people are at increased risk for HBOC, LS, and FH and that carriers may be missed by current guidelines. Identifying these individuals by population screening at a young age, such as through an optional employee benefit, can improve health outcomes and prevent disease.

805T

Quantifying the evolutionary parameters of age-related clonal hematopoiesis (ARCH). K. Skead^{1,2}, A. Ang Houle^{1,2}, D. Soave¹, S. Abelson³, M. Agbessi⁴, P. Mehanna¹, M. Minden^{3,4}, L. Shlush^{3,5}, J.E. Dick^{2,3}, S. Wright⁶, P. Awadalla^{1,2}. 1) Department of Computational Biology, Ontario Institute for Cancer Research, Toronto, Ontario, Canada; 2) Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada; 3) Princess Margaret Cancer Centre, Toronto, Ontario, Canada; 4) Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada; 5) Weizmann Institute of Science, Rehovot, Israel; 6) Department of Ecology & Evolutionary Biology, University of Toronto, Toronto, Ontario, Canada.

Cancers share common traits that are thought to accumulate progressively as a function of somatic evolution. We have a limited understanding of how these mutations accumulate and the evolutionary forces which catalyze the transition of a cell from a healthy to a cancerous state. Multiple studies reveal that hematopoietic stem cells (HSCs) accumulate somatic mutations as a function of regular aging in a process referred to as Age-Related Clonal Hematopoiesis (ARCH). Some of these acquired mutations are believed to confer a slight fitness advantage causing particular HSCs to undergo clonal expansions and contribute disproportionately more to the mature blood cell pool. However, ARCH has been found to associate with blood-based cancers and cardiometabolic conditions and, accordingly, as a pre-disease state, affords us the unique opportunity to capture and characterize the evolutionary forces contributing towards disease development. Here, we develop computational models to characterize tumour evolution by repurposing classic population genetic tools, and explore their utility in quantifying the evolutionary parameters associated with ARCH development. To test our models, we use longitudinal deep-coverage sequencing data and paired clinical outcomes from 100 pre-AML samples and 400 age-matched controls. We employ a comparative Bayesian approach to simulate a quantitative framework to ascertain which models and selection parameters shape the intra-individual accumulation of somatic mutations. Our simulation parameters are modelled on observed hematopoietic stem cell niche characteristics and we vary probabilities of mutations being beneficial, rates of selection and the distribution from which they are drawn. We are able to identify instances in both pre-AML and control samples where evolutionary parameters: 1) depart from neutrality, 2) are characteristic of deleterious selection, and 3) calibrate the rate and timing of driver mutations and their association with cancer outcomes. These models will be used to inform analytical approaches in deep coverage sequencing, and provide a richer description of temporal evolutionary processes so as to predict which clinical outcome a sample is destined towards. Our work establishes a paradigm for determining the mutational and deterministic forces shaping tumour evolution and stands to transform diagnostic pipelines through facilitating the incorporation of evolutionary analyses into standard clinical practice.

806F

Intratumor heterogeneity of non-small-cell lung cancer with EGFR mutations. X. Zhao¹, S. Qiao¹, X. Zhou¹, Y. Zhang², L. Zhang², K. Wu¹. 1) BGI-Shenzhen, Shenzhen, China; 2) Sun Yat-sen University Cancer Center, Guangzhou, China.

The mutation frequency of epidermal growth factor receptor (EGFR) is approximately 40% among Asian patients with NSCLC. Guidelines for the diagnosis and treatment of primary lung cancer in China recommended epidermal growth factor receptor-tyrosine kinase inhibitors (EGFR-TKIs) as a first-line therapy for advanced NSCLC patients carrying EGFR mutation including exon 19 deletion (19Del) and L858R mutation on exon 21 (L858R). Although survival periods in clinical practice of EGFR-TKIs treatment have been getting longer, the patients with different subtype of EGFR mutations exhibited diverse drug response to EGFR-TKIs. The reasons why are still unclear. In this study, we performed whole genome and panel sequencing of 239 spatially distinct regions from 42 lung adenocarcinoma of Chinese patients, of which 14 patients carried EGFR 19del mutation (EGFR^{19del}), 15 patients carried L858R mutation (EGFR^{L858R}) and 13 patients did not carry any of these two mutation types (Control). We found the intra-tumor heterogeneity (ITH) levels were significantly higher in EGFR-mutant patients compared with control ($P < 0.05$). Average fraction of non-ubiquitous mutation was 93.63%, 93.87% and 71.34% in EGFR^{19del}, EGFR^{L858R} and Control, respectively. Moreover, EGFR^{19del} patients showed slightly higher ITH levels than EGFR^{L858R} which non-silent mutations were considered as a unique parameter. Remarkably, two distinct evolutionary patterns were observed in the EGFR mutant subtypes. Cell locomotion and biological adhesion promoted oncogenesis and progression in EGFR^{19del} and the mutant of CTNNB1, COL2A1 and SMAD4 were the later driver events. Correspondingly, the evolutionary model of EGFR^{L858R} showed a diversity driver events in our cohort. These preliminary results suggested that intratumor genetic diversity provided a substrate for tumor adaptation and evolution in the subtype of EGFR mutation. It could provide genetic basis for diagnosis and prognosis in EGFR-TKIs therapeutic response and to guide stratification of lung cancer treatment and clinical research in future.

807W

ScaR - A tool for sensitive detection of known fusion transcripts: establishing prevalence of fusions in testicular germ cell tumors. A.M. Hoff, S. Zhao, J.M. Strømme, R.I. Skotheim. Institute for Cancer Research, Oslo University Hospital, Oslo, Norway.

Bioinformatics tools for detection of fusion transcripts from RNA-sequencing data are in general developed for identification of novel fusions, often from data with a high level of background noise. Therefore, these demand a high number of supporting reads and strict filters are applied to ensure high specificity and low false discovery rate. However, as our knowledge of bona-fide fusion genes become more saturated and these are also implicated as biomarkers in cancer, there is a more pressing need to pinpoint the presence and prevalence of these fusion transcripts with high sensitivity. Here, we report the development of *ScaR*, using a scaffold realignment approach for sensitive fusion detection in RNA-seq data. We show that *ScaR* detects a set of 50 synthetic fusion transcripts from simulated data at a higher sensitivity compared to the established fusion finders deFuse and FusionCatcher. We further apply *ScaR* to explore on previously established fusion genes in testicular germ cell tumours (TGCTs) in larger cohorts. From RNA-sequencing data of 150 TGCTs, available from The Cancer Genome Atlas, *ScaR* detects the fusion transcripts *RCC1-ABHD12B* and *CLEC6A-CLEC4D* in 9% and 28% of the samples, respectively. In comparison, none out of 198 normal testis tissue samples were positive for the fusion transcripts, using RNA-sequencing data from the GTEx consortium. Thus, we demonstrate high prevalence of the *RCC1-ABHD12B* and *CLEC6A-CLEC4D* fusion transcripts in clinical TGCT samples, and that they are highly cancer specific. Further, within TGCTs, we find that *RCC1-ABHD12B* and *CLEC6A-CLEC4D* are predominantly expressed in the seminoma and embryonal carcinoma histological subtypes. In conclusion, we have developed an approach for sensitive detection of fusion transcripts from RNA-seq data, *ScaR*, which will be useful for establishing the frequency of known fusion transcripts in larger data sets and detecting clinically relevant fusion transcripts with high sensitivity.

808T

Isoform expression signatures associated with epigenetic interactions between miRNA and DNA methylation in bladder cancer. M. Shivakumar¹, S. Han², Y. Lee^{2,4}, D. Kim^{1,3}. 1) Biomedical and Translational Informatics Institute, Geisinger Health System, Danville, PA; 2) Department of Biomedical Informatics, University of Utah, University of Utah School of Medicine, Salt Lake City, UT; 3) The Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, PA; 4) Huntsman Cancer Institute, University of Utah, Salt Lake City, UT.

Various epigenetic factors are responsible for the non-genetic regulation on gene expression. The epigenetically dysregulated oncogenes or tumor suppressors by miRNA and/or DNA methylation are often observed in cancer cells. Each of these epigenetic regulators has been studied well in cancer progressions; however, their mutual regulatory relationship in cancer still remains unclear. In this study, we propose an integrative framework to systematically investigate epigenetic interactions between miRNA and methylation at the alternatively spliced mRNA level in bladder cancer, which is a way of a higher resolution for interpreting miRNA's regulation at the mRNA level. DNA methylation, miRNA, and RNA-seq datasets for bladder cancer were retrieved from TCGA and analyzed using a rMAT tool to identify alternatively spliced events. Then, a likelihood-ratio test was used to determine if the interaction terms between miRNA and methylation for the evaluated models were statistically significant (Bonferroni corrected $P < 0.05$). Cox regression adjusting for age and gender was performed on statistically significant pairs. We identified 125 significant isoform signatures (103 genes) associated with epigenetic interactions as well as overall survival. As intragenic DNA methylation may have a different molecular function by their specific location in gene, we classified the interaction pairs according to the regions of DNA methylation (i.e., promoter, exon, intron, and 3' UTR). We observed that the methylation status is more dominant regulators than miRNAs in all interaction pairs. Additionally, we found that there were more genes in the intron region showing significant associations with overall survival. Our study provides a higher resolution of molecular insight into the crosstalk between two epigenetic factors, DNA methylation and miRNA. Furthermore, our work will be an addition to our knowledge on how intragenic methylation may play with miRNA for the mRNA regulation.

809F

Isoform expression analysis reveals distinct subset of cancer: Role of alternative splicing patterns of *LARS* in colon adenocarcinoma. H. Song^{1,2}, N.H. Kwon³, H.S. Yun⁴, S.S. Yoon^{1,5}, S. Kim³, Y. Koh^{2,5}. 1) Cancer Research Institute, Seoul National University, Seoul, South Korea; 2) Center for Medical Innovation, Seoul National University Hospital, Seoul, South Korea; 3) Medical Bioconvergence Research Center, Seoul National University, Seoul, South Korea; 4) Department of Bioinformatics, Seoul National University Hospital, Seoul, South Korea; 5) Department of Internal Medicine, Seoul National University Hospital, Seoul, South Korea.

Aminoacyl tRNA synthetases are essential enzymes for attaching appropriate amino acids to its tRNAs. Leucyl-tRNA synthetase (*LARS*) functions as an intracellular leucine sensor; when it senses sufficient amounts of leucine, it activates mammalian target of rapamycin (mTOR) pathway that regulates protein translation, cell size, and autophagy. We searched for alternative splicing in *LARS* gene in colon cancer samples. We aligned 46 whole transcriptome sequencing (WTS) data of TCGA colon adenocarcinoma (COAD) using STAR algorithm and quantified expression using RSEM algorithm. We used isoform-level expression of 16 isoforms followed by StringTie and Ballgown to re-assemble, merge, and quantify exon-level expression to explore unique expression pattern of *LARS*. We first performed unbiased clustering using 16 *LARS* isoform-level expression values and then performed another clustering based on the differentially expressed genes (DEG) from whole human gene set. CMScaller was used to annotate the samples in terms of the consensus molecular subtypes (CMSs) of colon cancer. We first classified 46 COAD samples into 4 clusters (A, B, C, and D) by *LARS* isoform expression, and identified that 5 samples in cluster A define a separate group. In order to verify consistency of the classification, we applied this cluster group annotation to the second clustering (DEG). As a result, we have found that in cluster A, six samples are clustered adjacent to each other in terms of distance between each sample. Interestingly, we noticed that these samples belong to either CMS 2 or CMS 4 which has high somatic copy number alterations. In addition, we found that these samples have mutations of *ATRX* (essential in normal development) and *STAG2* (a tumor suppressor) in common. Survival analysis revealed a tendency for short survival in *cluster A* compared to the other clusters. In conclusion, we demonstrated that the clustering by *LARS* isoform expression distinguishes certain biological similarities and distinct subsets in colon adenocarcinoma.

810W

Analysis of gene fusions and simple nucleotide variants using molecular bar coded chemistry. J. Balan, A. Nair, J. Voss, E.B. Fritcher, C. Zysk, T. Kollmeyer, N. Mattson, R. Gonzalez, E. Kischell, C. Ross, A. Raghunathan, C. Ida, B. Kipp, C. Giannini, R. Jenkins, E. Klee. Mayo Clinic, Rochester, MN.

Cancer is a result of tumor accretion resulting from genomic mutations, genomic rearrangements and epigenetic alterations. Detecting these alterations in the driver oncogenes is pivotal for clinical therapy. The developments in the Next Generation Sequencing (NGS) analysis technologies provide excellent means to study such alterations. Most sequencing techniques designed to study the driver oncogenes involve Polymerase Chain Reaction (PCR)-based sequencing methods that makes it challenging to identify true events over PCR bias artifacts. However, adding molecular bar codes (MBCs) to each DNA/RNA molecule permits traceability of PCR artifacts/duplicates. We have developed a neuro-oncology assay that utilizes the QIAseq (Qiagen Inc., Germantown, MD) targeted chemistry and Illumina sequencing. In our neuro-oncology assay, we target 150 genes for the DNA panel and 80 genes in the RNA panel to detect simple nucleotide variants (single nucleotide polymorphisms and small indels less than 50 base pairs in length) and gene fusions, respectively. We have developed UMIDNA and UMIFUSION pipelines to detect these simple nucleotide variants and fusions. UMIFUSION (a Cromwell/WDL based pipeline) uses an approach that involves denovo assembly of the targeted reads using CAP3, mapping the assembled contigs to human genome reference using BLAT, a second pass alignment using BWA-mem, false positive filtering module, annotation module and calculation of unique molecules that support the fusions. The UMIDNA pipeline involves trimming the reads to remove non-reference bases, alignment and variant calling using CLC and a post processing step that annotates variants with unique molecules that support the variant. Using these pipelines, we have analyzed 52 DNA samples and 61 RNA samples in neuro-oncology assay verification studies. For DNA samples, we achieved >98% concordance to an alternative method (Ion Torrent with QIAseq chemistry, other NGS Illumina chemistries, and Sanger) for detecting simple nucleotide reportable variants. For RNA data samples, we achieved 100% concordance to alternative confirmatory methods (RT-PCR/Sanger, other NGS Illumina chemistries and FISH) in detecting gene fusions. Leveraging the chemistry with the MBCs and pipelines that handle the MBCs, we have established high sensitivity of the DNA and RNA assays. Importantly, the simple variants and fusion data from DNA and RNA of a patient helps us advance the clinical diagnosis of cancer.

811T

Deep learning approach to automate somatic variant refinement. *E.K.*

Barnell¹, B.J Ainscough^{1,2}, K.M. Campbell¹, T.E. Rohan³, R. Govindan^{1,2}, M. Griffith^{1,2,4,5}, E.R. Mardis^{6,7}, S.J. Swamidass⁸, O.L. Griffith^{1,2,4,5}. 1) McDonnell Genome Institute, Washington University School of Medicine, St. Louis, MO, USA; 2) Siteman Cancer Center, Washington University School of Medicine, St. Louis, MO, USA; 3) Department of Epidemiology and Population Health, Albert Einstein College of Medicine, New York, NY, USA; 4) Division of Oncology, Department of Medicine, Washington University School of Medicine, St. Louis, MO, USA; 5) Department of Genetics, Washington University School of Medicine, St. Louis, MO, USA; 6) Institute for Genomic Medicine, The Research Institute at Nationwide Children's Hospital, Columbus, USA; 7) Department of Pediatrics, The Ohio State University College of Medicine, Columbus, USA; 8) Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, Missouri, USA.

Introduction: Cancer genomic analysis requires accurate identification of somatic variants in sequencing data. Manual review to refine somatic variant calls is required as a final step after automated processing. However, manual variant refinement is time-consuming, costly, poorly standardized, and non-reproducible. Here we describe a systematized and standardized method for somatic variant refinement using a machine learning approach. **Methods:** A deep learning model was developed using 41,000 variants from nine tumor subtypes including both hematologic tumors and solid tumors. Data was derived from multiple different sequencing platforms (whole-exome-, whole-genome-, and capture-sequencing) and variants were called by a variety of automated variant callers (VarScan, Strelka, etc.). Model performance was analyzed by using internal cross-validation, comparing model predictions to orthogonal sequencing data, and comparing model predictions to manual review calls on independent testing sets. **Results:** Internal cross-validation showed a receiver operating characteristic (ROC) area under the curve (AUC) of 0.96. When the model was employed on a set of 192,241 putative somatic variants derived from a single AML case with orthogonal sequencing data, the model attained a ROC AUC of 0.96. When the model was employed on a cohort of 17,356 variants derived from 11 lung carcinomas, 14 follicular lymphomas, 19 oral squamous cell carcinomas, and 1 breast cancer cell line, the model recapitulated manual review labels with a slightly reduced accuracy (ROC AUC range = 0.72-0.92). Reduction in accuracy was attributed to batch effects associated with the new data features including: different manual reviewers, new somatic variant callers, unseen tumor subtypes, and updated sequencing platforms. When retraining a new model that incorporated ~250 data points from the independent testing sets, model accuracy was recovered (ROC AUC >0.9). **Conclusions:** The deep learning model for somatic variant refinement (DeepSVR) provides an unbiased method for identifying true somatic variants to avoid observed high inter-reviewer variability. It also reduces the existing annotation bottleneck hindering widespread adoption of genomic analysis for every tumor workflow. The described model and methods for customizing a new model have been made publicly available at <https://github.com/griffithlab/DeepSVR>.

812F

Bioinformatics and DNA-extraction strategies to reliably detect genetic variants from FFPE breast tissue samples. *A.V. Bhagwate, S. Winham, S. McDonough, M. Stallings-Mann, J.M. Cunningham, E.P. Heinzen, J. Davila, R.A. Vierkant, D.C. Radisky, A.C. Degnim, C. Wang.* Mayo Clinic, Rochester, MN.

Archived FFPE samples are valuable clinical resources to examine tissue features (including protein expression and genetic changes) associated with clinical outcomes. However, the DNA quality and quantity from FFPE samples are sub-optimal and degraded, and resulting NGS-based genetics variant detections are prone to false positives. With the advancement of new NGS protocols for FFPE DNA, an overall evaluation of wet-lab and bioinformatics approaches are needed to optimize variant detection from FFPE samples. Focusing on breast tissue samples as pilot study, we designed within-subject replicates consisting of DNA derived from four paired FFPE and fresh frozen samples to highlight FFPE-specific artifacts. We tested two FFPE DNA extraction methods to determine the impact of wet-lab procedures on variant calling: Qiagen QIAamp DNA Mini Kit ("QA"), and Qiagen GeneRead DNA FFPE kit ("QGR") containing a repair step, resulting in a total of four triplicates (N=12 samples). The fresh frozen samples were prepared using Qiagen Gen-ta Puregene Kit ("QGP"). Additionally, we used one negative control sample (NA12891) and one positive control sample (Horizon Discovery) with known mutations. All DNA sample libraries were prepared for NGS according to QIAseq Human Breast Cancer Targeted DNA Panel protocol as per vendor's instructions, and sequenced on the HiSeq 4000, 150bp per lane, 12 samples multiplexed per lane. The variant calling was performed using Qiagen Gene Globe Data Portal, following which variant concordance comparisons were carried out within replicates. In this study, we found more variants on average called with FFPE samples (QGR=1119; QA=9592 variants), compared to their paired fresh-frozen tissue samples (QGP=273 variants), even after applying molecular barcoding error-correction and default bioinformatics filtering recommended by the vendor. We also found that an optimized FFPE-DNA extraction approach (QGR) leads to much fewer false variants. Approximately 92 % of the uniquely called FFPE variants were of low allelic frequency range (<5%), and collectively shared a "C>T|G>A" mutational signature known to be representative of FFPE artifacts. Based on control samples and replicates, we derived effective filtering strategy with associated empirical false-discovery estimates. Through this study, we demonstrated the feasibility of calling genetic variants from FFPE samples, and recommend a FFPE-fresh frozen paired design for controlling false variants.

813W

Integration of biological information into the CIViC database cancer variant representation. A. Danos, K. Krysiak, A. Wagner, S. Kiwala, J. McMichael, A. Coffman, E. Barnell, C. Ramirez, M. Griffith, O. Griffith. McDonnell Genome Institute, Washington University, Saint Louis, MO.

With the growing availability of clinical cancer genome sequencing coupled with growing knowledge on the vast array of cancer variants, there is great need for data resources focusing on the subset of clinically important cancer variants. The Clinical Interpretations of Variants in Cancer (CIViC, www.civicdb.org) database is designed to meet this need. As construction of large datasets requires massive amounts of curation effort, some databases maintain this curated knowledge behind a paywall, presumably using fees for this curation. In contrast, CIViC's approach is to crowdsource and expertly moderate curation effort, which allows it to be a completely free, open-access public resource, able to quickly respond to new directions in the field as the active user base grows. To date CIViC has built a community of over 100 contributors and over 16,000 clinical and research users worldwide. The fundamental unit of knowledge in CIViC is the evidence item (EID). The EID associates a clinically relevant statement to a specific variant in a specific cancer type using evidence derived from peer-reviewed publications and linked to a PubMed ID. EIDs are used to build summary clinical assertion statements for a variant in a specific cancer context, to which an AMP Tier and Level is associated. Here we introduce a new EID type called Biological Evidence. The CIViC EID currently consists of Predisposing, Predictive, Diagnostic or Prognostic evidence, and variants that are known to play an important role in cancer, but do not have directly actionable clinical information associated to them, are not admissible in CIViC. Upon user request for CIViC to be able to handle this type of evidence, and internal desire to expand the data model, Biological EIDs will be added to CIViC. This new evidence type will allow CIViC to gather information on important cancer driver mutations, and furthermore admit important new functionality enabling classification of AMP Tier III (VUS) and AMP Tier IV (known benign) variants. Additionally, Biological Evidence allows CIViC EIDs to incorporate PM1, PM4 and PM5 codes from the ACMG guidelines for assessment of variant pathogenicity at the assertion level. Current debates are underway regarding methods for assessment of somatic cancer variant oncogenicity, and integration of biological evidence will allow CIViC's evolving conception of the cancer variant to incorporate and help establish new standards for variant representation as they emerge.

814T

GCC-Calc, a method and tool to identify, compare, normalize and interpret genomic copy number variation in longitudinal bone metastases and xenografts. T. Gaasterland^{1,2,3}, M.T. Muldong^{4,5}, A.A. Kulidjian⁶, C.J. Kane⁴, C.A.M. Jamieson^{4,5}. 1) Institute for Genomic Medicine, University of California, San Diego, La Jolla, CA; 2) Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA; 3) Bioinformatics & Systems Biology Program, University of California, San Diego, La Jolla, CA; 4) Department of Urology, University of California, San Diego, La Jolla, CA; 5) Moores Cancer Center, University of California, San Diego, La Jolla, CA; 6) Department of Orthopaedic Surgery, University of California, San Diego, La Jolla CA.

Rationale: Prostate cancer (PrCa) metastasizes to bone in 50-90% of patients with advanced disease yet little is known about genome-wide alterations in the bone metastases. To identify changes that may lead to progressive therapy resistance, we measured genome variation and gene expression in a longitudinal series of surgical PrCa bone metastasis samples and xenografts derived from one patient. To interpret the data, we developed methods to compare copy number variants across samples and measurement platforms, determine gene structures impacted over time, and assess expression changes. **Data:** Surgical samples from one patient's serial metastasis events during 3 years were saved as flash-frozen tissue and dissociated cells. Xenograft models were generated from the cells and subjected to the patient's treatments. Blood provided a baseline genome. Healthy bone marrow provided baseline gene expression. Data from each sample included copy number variant (CNV) regions measured with the Oncoscan CNV platform, whole exome sequencing (WES), and gene expression measured with the GeneChip RNA platform. Exome data were analyzed to compute CNV regions using CNVkit and single nucleotide polymorphism (SNP) sites with samtools and GATK. **Approach:** We developed a method to generate a merged CNV profile for a given sample. It compares CNV regions measured with different platforms in the same sample, determines concordances and differences, and generates a consistent CNV profile. The method then compares CNV profiles across serial samples, determines emerging variants, and associates new variants with genes. Gene expression data is then queried to assess expression patterns for genes affected by CNVs. The methods address the following data challenges: normalization of CNV gain scores across samples, resolving lower-confidence CNVs in patient samples with differing proportions of normal and diseased cells, and distinguishing true CNV gains and losses from mis-measurement. A tool suite, GCC-Calc (Gene Copy Comparison Calculator), accepts tables of CNV regions for serial samples and reports affected genes and expression changes. **Results:** With GCC-Calc, merged, serial CNV profiles revealed a therapy resistant sub-clone present at low levels in the early patient bone metastasis. GCC-Calc provides a foundation to profile genomic diversity in recurrent bone metastatic cancers. It enables comparison of CNV platforms and resolves differences in CNV profiles obtained for one sample.

815F

A sensitive and robust pipeline for analyzing 5hmC patterns from low coverage oxidative bisulfite sequencing. *R.L. Goldfeder, A.C. Lau, C. Wei.* The Jackson Laboratory for Genomic Medicine, Farmington, CT.

Over 600,000 Americans die each year of cancer. Cancer is so deadly because it often cannot be identified until symptoms are very advanced, when the cancer has likely metastasized and treatment is less successful. Liquid biopsy is a new, non-invasive technique that enables early cancer detection, thereby facilitating earlier life-saving treatment. This technique involves simply drawing blood, sequencing the cell-free DNA (cfDNA) in the blood, and analyzing the DNA for biomarkers that indicate the presence of cancer. Until recently, cfDNA analysis has focused on sequence mutations or patterns of 5-methylcytosine (5mC). Now, researchers are beginning to investigate 5-hydroxymethylcytosine (5hmC), a tissue-specific epigenetic mark that has been linked to cancer. This mark can be assayed from bisulfite and oxidative bisulfite sequencing. Our group and others aim to study 5hmC patterns from cfDNA and solid tumor biopsy; our goal is to identify patterns in cfDNA that recapitulate the patterns seen in tumor biopsy so that 5hmC can be utilized as a biomarker in liquid biopsies for early cancer detection. cfDNA is fragmented, in low abundance, and is a representation of the tumor-derived tissue as well as other tissues in the body, creating several challenges with sequencing and analysis. Currently, the available tools for detecting 5hmC rely on high coverage datasets or are inappropriate for meeting the challenges that arise with cfDNA. For example, we see short sequence reads with paired ends that overlap and high duplication rates that ultimately lead to low coverage and low quality 5hmC calls. Additionally the mapping rate is low due to the nature of bisulfite sequencing. To address these challenges, we present Convinced, a sensitive and robust pipeline for detecting 5hmC patterns from oxidative bisulfite sequencing. We evaluate Convinced using publically available data and benchmark against a commonly used pipeline for analyzing 5hmC, BsExpress. We find that Convinced provides a higher mapping rate and more accurate 5hmC calls. Our pipeline improves 5hmC detection from bisulfite and oxidative bisulfite sequencing libraries. This will ultimately enable researchers and physicians to use 5hmC patterns from liquid biopsies to aid in the non-invasive early detection of cancer.

816W

Development and analytical validation of highly customizable next-generation targeted sequencing technology for the detection of gene fusions at a very low frequency. *R.K. Gottimukkala, N. Li, Y. Fu, A.G. Hatch, C. Van Loy, C. Bai, A. Erwing, D.J. Mazur, F.M. Hyland.* Thermo Fisher Scientific, South San Francisco, CA.

Gene fusions caused by chromosomal rearrangements play an important role in oncogenesis, the progression of cancer and the selection of guided targeted therapies. Next-generation sequencing enables sensitive, specific and precise detection of particular fusion isoforms for defined gene pairs. AmpliSeq™ HD technology is new targeted sequencing chemistry that allows highly sensitive variant detection down to $\leq 0.1\%$ mutant allele fractions in both blood and FFPE samples. We developed cloud-based software to support the design of a custom AmpliSeq™ HD gene fusion panel, comprising 1 to 523 fusion isoform assays and any gene expression assays for normalization. These include high quality curated and annotated fusion isoforms involving ALK, RET, ROS1, NTRK1 and other genes. We designed AmpliSeq™ HD primer pairs for each of these fusions using advanced assay design algorithms. We developed cloud-based analysis software to analyze the BAM file resulting from amplification and sequencing of custom AmpliSeq™ HD fusion panels on an Ion GeneStudio™ S5 sequencer. The sequencing reads generated using this chemistry contain unique molecular tags that allow identification of reads which originated from the same molecule. We developed an algorithm that models errors accumulated during amplification and sequencing, and accurately reconstructs the sequence of original molecules based on these reads. We developed a Fusions calling method that uses the reconstructed sequences to enable accurate detection of gene fusions. To validate the AmpliSeq™ HD chemistry and the analysis workflow for the detection of gene fusions, we created three custom panels: Panel 1 with 95 fusion isoform assays and 4 gene expression assays, panel 2 with 40 fusion isoform assays and 19 gene expression assays and panel 3 with 17 fusion isoform assays and 19 gene expression assays. We sequenced the Total RNA from Tri-Fusion control sample diluted to $\leq 1\%$ with the background of wild-type RNA, using all the three panels and detected all the expected fusions; EML4-ALK, SLC34A2-ROS1, and CCDC6-RET with 100% specificity. We sequenced the Seracare FFPE samples using panel 2 and detected all the 8 expected fusions with 100% sensitivity and specificity. We also sequenced the negative FFPE samples using panel 2 and as expected, detected no fusions. This technology enables researchers to create large multiplex custom FFPE and cfDNA panels to detect gene fusions at an ultra-low frequency with high accuracy.

817T

Combining single-cell RNA and DNA sequencing analysis to measure genomic instability at individual clone resolution in gastric cancer. S.M. Grimes¹, N. Andor², B.T. Lau¹, H.P. Ji^{1,2}. 1) Stanford Genome Technology Center, Stanford University, Palo Alto, CA; 2) Division of Oncology, Department of Medicine, Stanford University School of Medicine, Stanford, CA.

Genomic instability is a hallmark of cancer and gives rise to the diversification which allows tumors to evolve into multiple genetically distinct subclones. Traditional bulk sequencing analysis averages out the signals from the individual clones, and therefore is often unable to distinguish the unique markers which characterize smaller clones. These unidentified clones may be resistant to prescribed treatment, leading to subclonal expansion and tumor growth. To help decipher intratumor heterogeneity we present a method to measure genomic instability at the individual clone resolution, using a combination of single cell DNA and RNA sequencing analysis. Copy number variation (CNV) and clustering of cells in G0G1 cell cycle phase with similar CNV profiles, is performed using both techniques. For scDNA the cell cycle was inferred by analysis of cell ploidy and breakpoints; CNV calls were made by standard open-source software and clustering was based on CNV segmentation. For scRNA the cell cycle was determined by the activity of pathways related to the different phases of the cell cycle, copy number was inferred by gene expression across the transcriptome and clustering was based on consolidated gene expression across large (>10Mb) segments of the genome. This methodology was validated on several gastric cancer cell lines, and resulted in good correlation between number and size of clones identified independently by scDNA and scRNA on the same samples. This method was then used to analyze a primary gastric tumor sample, with scDNA enabling a more granular identification of clones and phylogeny, and scRNA categorizing clones in more detail via cell surface markers and pathway analysis. The ability to interrogate both the genome and transcriptome of single cells and identify clonal characteristics provides a very powerful analysis tool with regard to tumor evolution and cell fitness, enabling potential identification of new therapeutic targets.

818F

Exome-wide analysis of bi-allelic alterations identifies a Lynch phenotype in the Cancer Genome Atlas. O. Harismendy¹, A.R. Buckley², T. Ideker¹, H. Carter¹, N.J. Schork^{1,2,3}. 1) UC San Diego, San Diego, CA; 2) J. Craig Venter Institute, La Jolla, CA; 3) The Translational Genomics Research Institute, Phoenix, AZ.

Cancer susceptibility genes generally require the somatic alteration of the remaining allele to drive oncogenesis and, in some cases, tumor mutational profiles. Whether combined germline and somatic bi-allelic alterations are universally required for germline variation to influence tumor mutational profile is unclear. Here we performed an exome-wide analysis of the frequency and functional effect of bi-allelic alterations in The Cancer Genome Atlas (TCGA). We integrated germline variant, somatic mutation, somatic methylation, and somatic copy number loss data from 7,790 individuals from TCGA to identify germline and somatic bi-allelic alterations in all coding genes. We used linear models to test for association between mono- and bi-allelic alterations and somatic microsatellite instability (MSI) and somatic mutational signatures. We discovered significant enrichment of bi-allelic alterations in mismatch repair (MMR) genes, and identified 6 bi-allelic carriers with elevated microsatellite instability (MSI), consistent with Lynch syndrome. In contrast, we find little evidence of an effect of mono-allelic germline variation on MSI. Using MSI burden and bi-allelic alteration status, we reclassify two variants of unknown significance in *MSH6* as potentially pathogenic for Lynch syndrome. Extending our analysis of MSI to a set of 127 DNA damage repair (DDR) genes, we identified a novel association between methylation of *SHPRH* and MSI burden. We find that bi-allelic alterations are infrequent in TCGA, but most frequently occur in *BRCA1/2* and MMR genes. Our results support the idea that bi-allelic alteration is required for germline variation to influence tumor mutational profile. Overall, we demonstrate that integrating germline, somatic, and epigenetic alterations provides new understanding of somatic mutational profiles.

819W

Identification of critical cis-regulators of differential gene expression between EGFR mutant and wild-type lung adenocarcinomas. H. Ho¹, S. Yuan¹, G. Chang^{2,3,4}, S. Yu⁵, K. Li^{1,6}. 1) Institute of Statistical Sciences, Academia Sinica, Taipei, Taiwan; 2) Division of Chest Medicine, Department of Internal Medicine, Taichung Veterans General Hospital, Taichung, Taiwan; 3) Faculty of Medicine, School of Medicine, National Yang-Ming University, Taipei, Taiwan; 4) Institute of Biomedical Sciences, National Chung Hsing University, Taichung, Taiwan; 5) Department of Clinical Laboratory Sciences and Medical Biotechnology, College of Medicine, National Taiwan University, Taipei, Taiwan; 6) Department of Statistics, University of California, Los Angeles, USA.

Introduction Lung cancer is the most common cause of cancer deaths worldwide with lung adenocarcinomas being its predominant subtype. EGFR-activating mutation is the most common driver mutation for lung adenocarcinomas in East Asia and also has a high carrier frequency in the West. Several studies have established that EGFR-TKIs are in general more effective for EGFR mutant (MT) patients than wild-type (WT) patients. The cause of the response heterogeneity at the molecular level remains unclear. We hypothesized that a better understanding of the heterogeneity of the gene regulation between EGFR MT/WT tumors may help understand the response heterogeneity. In cancer genome, gene expression is regulated by factors at multiple molecular levels, including point mutation, indels and copy number alteration (CNA) at the DNA level; transcription factor and miRNA at the RNA level; and methylation at the epigenetic level (hereinafter referred to as “regulators”). Integrating multi-level molecular profiling is urgently needed for studying gene regulation. **Methods** Multi-level molecular profiling, including MASS spectrometry for EGFR mutation, array CGH for CNA, gene expression array and real-time PCR for miRNA expression was performed on a cohort of 177 lung adenocarcinomas from National Taiwan University Hospital and Taichung Veterans General Hospital. We proposed a modified liquid association score to quantify the strength of differential regulations (DR) between EGFR MT/WT lung adenocarcinomas. We derived a novel statistical framework combining differential expression (DE) analysis and DR analysis to form an enrichment test for identifying the critical regulators of the DE genes. **Results** Through the DE analysis, we identified 507 DE genes between EGFR MT/WT patients. By utilizing our approach, we found that four DE genes were differentially regulated by cis-CNAs and 109 DE genes were differentially regulated by seven DE miRNAs. In total, 22% of the DE genes were regulated to at least one critical cis-regulators. Among them, five critical cis-regulators were associated with patient's overall survival. **Conclusions** Integrative analysis of multi-level molecular profiling shed light on the critical regulatory disparity between EGFR MT/WT lung adenocarcinomas associated with patient's outcome. Our finding provides a better understanding of the heterogeneity of gene regulation between EGFR MT/WT patients and may help improve patient management in lung adenocarcinoma.

820T

Deconvoluting pathway dynamics using big data. S.Chloe. Jang, B. Han. Department of Medicine, Seoul National University College of Medicine, Seoul, South Korea.

Complex diseases such as cancer are usually the result of a combination of various biological pathways consisting of sets of genes. It is potentially invaluable to identify causal pathways of diseases, as they give understanding of pathogenesis of diseases. For this reason, pathway analysis has become a popular method for revealing new insight into the underlying biology of differentially expressed genes, as it identifies functional associations among genes. The current pathway datasets include the snapshot of each pathway that consists of the genes in the pathways (nodes) and the relation between them (edges). The limitation of these datasets is that they give only the static snapshot, and not the dynamics under different conditions. For example, a subset of edges may be “open” under specific conditions and “closed” under different conditions. The existing datasets lack this information because the edges in the pathway were confirmed by many different experiments that differ by time, cell type, and condition. This study proposes a new approach that decomposes the biological pathways into subgroups that dynamically act together by using big expression data. Specifically, we assess how pathway relations (edges) are associated and correlated using the datasets from Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) data sets. Colorectal Cancer(CRC) is one of the most frequent cancers and accounts as the second leading cause of death from cancers in the world. We investigated CRC expression data from 24 gene expression datasets derived from GEO and TCGA. We parsed all the human pathway data in major public repository, Kyoto Encyclopedia of Genes and Genomes (KEGG), and filtered out the cancer-related pathways. We then assessed the correlation between genes denoting edges in the pathways. We confirmed that the correlations of the edges were higher than the null subsets (random 1000 gene sets) as expected. Then we analyzed the dynamics of the edges in each pathway by calculating associations between the correlation coefficients of edges and applying clustering algorithms. We found that many pathways show clustering that may give clue how edges act together under different conditions. We expect that our findings may give exploration of the effective subnetworks of pathways, which may reduce the complexity in pathway mapping analysis and increase power to detect sub-pathways that contain genetic signals associated with complex diseases.

821F

PathoMAN: An automated germline cancer variant classification tool for pathogenicity and clinical actionability. V. Joseph¹, V. Ravichandran¹, Z. Shameer¹, Y. Kemel¹, M. Walsh¹, K. Cadoo¹, S. Lipkin², D. Mandelker¹, L. Zhang¹, Z. Stadler¹, M. Robson¹, K. Offit¹. 1) Medicine, MSKCC, New York, NY., USA; 2) Weill Cornell Medical College, New York, N.Y., 10065, USA.

Clinical Cancer Gene testing is one of the cornerstones of cancer prevention. Cancer Genetics testing has transformed from single gene test to multiplexed gene panel testing that extend beyond traditionally tested cancer types. The trend towards complex gene panels to exomes is unmistakable. While, data generation is rapid, the data interpretation has lagged behind. Even identifying known gene mutations is usually tenuous for many end-users. There is demand for a germline genomics platform that can perform high throughput variant interpretation and gene discovery. Similarly, there is no other algorithm that can perform variant classification in a uniform, rapid, unbiased fashion for key cancer genes. PathoMAN, a variant classification system, is perhaps the only system that brings rich germline cancer genetics domain expertise, intuitive ease of use and clear interpretation in the same application. PathoMAN delivers variant classification and interpretation with high concordance and low discordance to those curated by domain experts. It provides these classifications in six categories: Pathogenic (P), Likely-Pathogenic (LP), Benign (B), Likely-Benign (LB), Risk Factor (RF) and Variant of Uncertain Significance (VUS). In addition, it also tags select variants as a founder mutation in certain population such as the Ashkenazi Jews. We compared expert manually curated variant data from studies on (i) prostate cancer (ii) breast cancer and (iii) ClinVar to assess performance. PathoMAN achieves high concordance (83.1% pathogenic, 75.5% benign) and negligible discordance (0.04% pathogenic, 0.9% benign) when contrasted against expert curation. Some loss of resolution (8.6% pathogenic, 23.64% benign) and gain of resolution (6.6% pathogenic, 1.6% benign) was also observed. We propose new literature mining features and functional evidence for mutations from saturation mutagenesis studies that is used in the variant classification. Discovery mode enables users to perform gene burden test against ExAC population. PathoMAN will reduce the manual workload of domain level experts. It provides a substantial advance in rapid classification of genetic variants by generating robust models using a knowledge-base of diverse genetic data. <https://pathoman.mskcc.org>.

822W

Improving variant detection with unique molecular index technology: The case of biomedical genomics workbench. F. Lescai¹, B.J. Vilhjalms¹, A.M. Hein¹, D. Kamaev¹, M.M. Johansen¹, L. Husted¹, R. Samara², T. Rask¹, R.G. Madsen¹, Z. Etches¹, V. Gregersen¹, S. Tian², E. Lader², L. Schausen¹, M. Simonsen¹. 1) QIAGEN, Aarhus, Denmark; 2) QIAGEN, Frederick, MD, United States.

In numerous applications, and particularly in detecting somatic mutations at low allele fractions, the capability to distinguish between sequencing or amplification errors and real variants is crucial. Unique Molecular Indices (UMIs), also known as molecular barcodes, address these challenges and allow an unbiased estimation of allele-fractions, by providing a one-to-one representation of the originally captured sample DNA molecules. However, bioinformatics functionality capable of exploiting the added value of molecular barcodes is needed. To provide a sample-to-insight solution, QIAGEN combines a powerful chemistry based on UMIs and a UMI-aware bioinformatics workflow in the Biomedical Genomics Workbench. We assess the impact of UMI information on the sensitivity and precision of variant detection at several coverage levels. We prepared artificial mixtures of Genome-in-a-Bottle and Horizon reference DNA standards to evaluate performances in validated high confidence regions, and in known difficult-to-detect-variants regions. Our analysis quantifies the effect of integrating UMI data into the variant calling pipeline: at 1000X UMI-coverage, the sensitivity in detecting variants with expected allele fraction >5% is 100% when using UMIs, while is 93.3% if UMI information is not used; when detecting variants with expected allele fraction >1%, the sensitivity in UMI-aware workflow is 90.9% and 83.3% without UMI. At 2000X, the precision with UMI variant calling is 99.45% and only 81.55% if UMIs are not used, highlighting their value when analysing high-coverage datasets, crucial in applications like liquid biopsy. Our analysis further investigates this effect in relationship to coverage, and provides evidence of the benefits of combining this new barcoding technology with specifically designed bioinformatics solutions, such as the one offered in QIAGEN Biomedical Genomics Workbench. Finally, we provide suggestions on appropriate sequencing coverage levels in order to reach optimal calling accuracy, and we compare the performance of the Biomedical Genomics Workbench's resequencing workflow with other state-of-the-art pipelines.

823T

Ensemble machine learning for identifying the important biomarkers in proteogenomics cancer study. Y. Liang¹, A. Gharipour², A. Kelemen³, A.L. Kelemen⁴. 1) Family and Community Health, University of Maryland, Baltimore, Baltimore, MD; 2) School of Information and Communication Technology, Griffith University, Gold Coast campus, Queensland 4222, Australia; 3) Department of Organizational Systems and Adult Health, University of Maryland, Baltimore; 655 W. Lombard Street, Baltimore, MD 21201; 4) Department of Computer Science, University of Maryland, College Park, MD.

Identification of important biomarkers such as proteins is essential for medical diagnosis and prognosis. Accurate biomarker or disease signature selection is critical for drug targets in common diseases such as cancer. Standard statistical models with single feature selection involve the multi-testing burden of low power if the limited samples available. Furthermore, high correlations among the proteins with high redundancy lead to unstable selections and may cause reproducibility issues based on p values. The ensemble feature selection approach incorporates stability consideration into the algorithm design stage and has the advantage to alleviate and compensate such drawbacks. It aims at identifying a stable set of disease signatures or biomarkers that could improve the prediction performance of classification models, and thereby also simplify their interpretability. Mass spectrometry based proteomic technologies have enabled protein global expression profiling in order to examine the linkages between mRNA, proteins, and disease status (i.e. ovarian cancer). In this work, we developed and conducted two stage Ensemble Machine Learning approaches for both identifying biomarkers and predicting binary outcomes (putative homologous recombination deficiency, HRD positive or negative) and multiple mRNA classes (differentiated, proliferative, immunoreactive, mesenchymal, and unknown) using ovarian cancer proteogenomics data. In stage one, a homogeneous ensemble biomarker selection with data perturbation based on random forests (RF) approach is developed and implemented to identify important biomarkers. In stage two, we utilized the selection result of stage 1, and compared various ML methods (gaussprLinear, gaussprRadial, LogitBoost, MLP, MLPML, parRF, RF, SVM, treebag, WSRF) for predicting binary HRD, and multiple mRNA classes. Results show that our ensemble approach identifies the consistent and reproducible set of important markers out of 9606 proteins linked to the binary HRD status, and with good prediction accuracy, i.e. Treebag provided 83% prediction accuracy (85% sensitivity and 81% specificity); while for mRNA multi-classes, RF provided 81% prediction accuracy. One drawback of the proposed ensemble approach is that it is computationally expensive (i.e. in the ensemble of RF with homogeneous feature selection algorithm using 10000 trees), therefore requiring constant tuning of the model parameters for an improved performance.

824F

Amplicon Reconstructor: Optical mapping aids the reconstruction of complex, large cancer genome rearrangements. J. Luebeck¹, V. Deshpande², C. Coruh³, D. Pai⁴, S. Wu⁵, C. Zhang¹, K. Turner⁵, J. Law², P. Mischel⁵, V. Bafna². 1) Bioinformatics & Systems Biology Graduate Program, UC San Diego, La Jolla, CA; 2) Department of Computer Science, UC San Diego, La Jolla, CA; 3) Salk Institute for Biological Studies, La Jolla, CA; 4) BioNano Genomics, San Diego, CA; 5) Ludwig Institute for Cancer Research, UC San Diego, La Jolla, CA.

Focal amplifications (< 5 Mb) are a hallmark of the cancer genome, and differ from large scale chromosomal aneuploidy or simple segmental tandem duplication. Focal amplifications have been shown to reside not only on the linear chromosomes as homogeneously staining regions (HSR), but also as circular extrachromosomal DNA (ecDNA). ecDNA amplicons show high copy count variation, ranging from 10 to 100+ (Turner, Nature 2017). Extrachromosomal structures tend to be highly rearranged and may include segments from multiple chromosomes. Recent estimates suggest ecDNA may be found in ~40% of cancers (Turner). Previously, we developed a tool, Amplicon Architect (AA) that combines coverage and breakpoint information from short reads to output a breakpoint graph containing possible amplicon structures. Due to the complexity of the breakpoint graph, it is often impossible to reconstruct the complete architecture using short reads alone. Optical mapping provides long restriction maps (map length ~200kb) which can be assembled into ultra-long contigs (assembly N50 ~2.5Mb). While useful for ordering large segments of DNA, optical maps struggle to map smaller fragments individually. We developed a method, Amplicon Reconstructor (AR) that uses BioNano optical mapping to scaffold segments in AA-reconstructed breakpoint graphs from NGS data, disambiguating amplicon architecture. AR uses multiple algorithmic strategies (contig recruitment, graph matching and dynamic programming) to obtain candidate reconstructions. AR also maps possible integration sites of ecDNA into chromosomal DNA, and calls structural heterogeneity of amplicons. On five samples with cytogenetically validated ecDNA or HSR amplicons, AR revealed the structure of ecDNA in 3 glioblastoma samples, and identified sites where ecDNA had reintegrated back into chromosomal DNA. In one sample, AR reported structurally heterogeneous ecDNA reconstructions with differing status of EGFRV8 mutation. In a hematopoietic cell line, AR produced the first complete reconstruction of a Philadelphia chromosome (BCR-ABL gene fusion), involving segments from three chromosomes. Additionally, it enabled the first scaffolding of a breakage fusion bridge (BFB), in a lung cancer cell line, by identifying assembled optical map contigs spanning multiple junctions found in the 18.3 Mb BFB structure. Thus we demonstrated a combination of optical mapping and NGS in concert with AR produces unambiguous reconstruction of complex structures.

825W

Classification of prostate cancer patients as indolent or aggressive using machine learning approach. T. Mamidi, C. Hicks, J. Wu, E. Nicklow. Department of Genetics, Louisiana State University Health Sciences Center, School of Medicine, 533 Bolivar Street, New Orleans, LA 70112.

Background: Prostate cancer (PCa) is the second most common cancer in men in the US. Many PCas are indolent and do not result in cancer mortality, even without treatment. However, a significant proportion of patients with PCa have aggressive tumors that progress rapidly to metastatic disease and are often lethal. Currently, treatment decisions for PCa patients are guided by various stratification algorithms. Among these parameters, the most potent predictor of PCa mortality is the Gleason score (ranges from 6-10). Although current risk stratification tools are moderately effective, limitations remain in their ability to distinguish truly indolent from aggressive and potentially lethal disease. Here we propose the use of machine learning (ML) for classification of PC patients as having either indolent or aggressive tumors using transcriptome data. We hypothesize that genomic alterations could lead to measurable changes distinguishing indolent from aggressive tumors. **Method:** We used gene expression data and clinical information on 546 samples (494 tumors and 52 normal) from The Cancer Genome Atlas (TCGA). Tumor samples were distributed according to Gleason score (indolent 45, Gleason score = 6; intermediate grade 245, Gleason score 7 and aggressive 204, Gleason score 8-10). Because intermediate grade has more variable clinical course, we classified it as either indolent (3+4) or aggressive (4+3) based on current treatment protocol. We performed supervised analysis to discover genes distinguishing the two patient groups. We selected a small subset of significantly differentially expressed genes for use as classifiers. We used three ML algorithms including Bayes classifier, Logistic model tree (LMT) and Hoeffding tree classifiers and evaluated their performance and classification accuracy.

Results: Preliminary results have shown that using current classification protocol of 3+4 for indolent or 4+3 for aggressive, for patients with intermediate Gleason grade 7, patients were misclassified by ~50% by all the three classifiers. We plan to develop new classification algorithms that integrate transcriptome data with mutation information for accurate classification of patients as having indolent and aggressive tumors using ML. **Conclusion:** These results show that current classification protocol could misclassify PC patients and that ML algorithms could be used to accurately classify PCa patients to indolent or aggressive using transcriptome data.

826T

Analysis of targeted sequencing data of hereditary lung cancer families identifies germline copy number variations (CNVs) in multiple genes. D. Mandal¹, D. Tran¹, K.W. Termine¹, T. Mamidi¹, A. Musolf², M. Baddoo³, M. de Andrade⁴, C. Gaba⁵, P. Yang⁶, M. You⁷, M.W. Anderson⁸, A.G. Schwartz⁹, S.M. Pinney⁹, C.I. Amos⁹, J.E. Bailey-Wilson². 1) Genetics, LSU Health Sciences Center, New Orleans, LA; 2) National Human Genome Research Institute, National Institutes of Health, Baltimore, MD; 3) Tulane Cancer Center, New Orleans, LA; 4) Mayo Clinic, Rochester, MN; 5) University of Toledo Dana Cancer Center, Toledo, OH; 6) Medical College of Wisconsin, Milwaukee, WI; 7) University of Cincinnati College of Medicine, Cincinnati, OH; 8) Karmanos Cancer Institute, Wayne State University, Detroit, MI; 9) Baylor College of Medicine, Houston, TX.

About 10% of lung cancer (LC) cases (22,000 cases per year) in the U.S. have at least one first-degree relative affected with LC, and 25% of cases have at least one first- or second-degree affected relative. Previous multipoint linkage analyses of hereditary lung cancer (HLC) families (≥3 LC/family) collected by the Genetic Epidemiology of Lung Cancer Consortium detected significant linkage to 6q25. To aid in further investigation, targeted sequencing of 6q23-6q27 was obtained and we analyzed these data to identify germline copy number variations (CNVs). CNVs have been reported on lung cancer cell lines and somatic samples; however, no germline specific CNVs have been reported in HLC families. The objective of the present study is to detect germline CNVs from the targeted sequencing data in HLC families that might be responsible for increased susceptibility to LC in these high-risk families. The nine most strongly linked HLC families (≥4 LC cases/family) on whom we have targeted sequencing data included 25 affected and 55 informative unaffected family members. Phenotype, genotype, and smoking data are available on all family members. Two CNV-specific algorithms as incorporated in CANOES (<http://www.columbia.edu/~ys2411/canoes/>) and XHMM (<https://atgu.mgh.harvard.edu/xhmm/>) were used in order to successfully call and identify germline CNVs. To confirm, CNVs were then visualized independently using Integrative Genomics Viewer (IGV). Using both CANOES and XHMM we have identified CNVs in more than 30 genes that are deleted or duplicated in multiple families. Of those, CNVs in four cancer related genes (RGS17, SASH1, ARID1B, and PARK2) are present in more than three families and are detected by both algorithms. These genes were identified previously as candidates using either linkage analysis of the HLC families or by targeted sequencing analysis of 6q23-27. CNVs in three other cancer related genes, PACRG, RPS6KA2, and ESR1 were identified by both algorithms in five or more families. Detection of these genes by both CNV calling algorithms in multiple HLC families warrants further investigation in additional HLC families. In addition to identifying CNVs, transmission of germline CNVs in these families in multiple generations, including both affected and informative unaffected family members, will be useful for biomarker identification for prevention and early detection of LC in high-risk individuals.

827F

A high throughput software pipeline for NGS-based association studies: From assembly to candidate variants and their effect on 3D protein structures. K. Maxfield, S. Baldwin, D. Nash, M. Larson, M. Keyser, S. Darnell, T. Durfee. DNASTAR, Inc., Madison, WI.

Advances in DNA sequencing have made identifying genetic variation from any individual routine and allowed variation across populations or cohorts to be analyzed for candidate mutations causing diseases or traits of interest. However, the computational complexity of the analysis together with the enormous amount of data pose daunting challenges in terms of bioinformatics, data management and computing resources that are beyond all but the most well equipped laboratories. We have developed a seamlessly integrated software pipeline to address this problem. NGS sequencing reads from each sample are aligned to a reference genome using SeqMan NGen, a fast non-memory bound assembler for data sets of any size. A Bayesian modeled probabilistic variant caller analyzes the gapped alignments in-stream to produce high accuracy single nucleotide variant (SNV) and small indel calls for each sample. Assemblies can be done on standard desktop computers or on the DNASTAR Cloud which allows the samples to be processed in parallel. Variant profiles from each sample are automatically combined into a single project file for analysis in ArrayStar where various searching and statistical methods are available for identifying candidate genes and/or variants of interest. ArrayStar also allows gene and variant level annotations to be added aiding in the identification and prioritization of candidates. Further, for genes with known 3D protein structures, the effect of candidate missense mutations can automatically be predicted through integration our molecular structure visualization and analysis module, Protean 3D. As a demonstration of the pipeline, we will present results from the reanalysis of 96 targeted resequencing samples from a Chinese cohort with lung squamous cell carcinomas (LSCC, Li et al., Sci Rep. 2015. 5:14237). Results will be compared to the published observations highlighting new candidate genes/variants that may have been initially missed. We will also show examples of how predicted protein structure changes caused by missense mutations can aid in prioritizing variants of interest.

828W

Expanding GEMINI to annotate and prioritize subclonal mutations in heterogeneous tumors. T. Nicholas^{1,2}, B. Pedersen^{1,2}, Y. Qiao^{1,2}, X. Huang^{1,2}, G. Marth^{1,2}, A. Quinlan^{1,2}. 1) Human Genetics Department, University of Utah, Salt Lake City, UT; 2) USTAR Center for Genetic Discovery, University of Utah, Salt Lake City, UT.

DNA sequencing has unveiled extensive tumor heterogeneity in several different cancer types, with many exhibiting substantial clonal substructures. Considerable genetic variation exists within individual tumors and between metastatic sites from shared phylogenies, highlighting ongoing tumor growth and evolution. These differences can be further magnified over time by selective pressures introduced by different treatments. Properly identifying and tracing this variation throughout the expansion and progression of a tumor represents a significant challenge. Furthermore, being able to prioritize and focus on mutations most likely to contribute to tumor evolution or that could serve as potential therapeutic targets represents an ongoing problem, but is crucial towards determining possible clinical care. GEMINI is an effective tool for exploring genetic variation by enabling the querying of multiple genome annotations into a single database. Here we describe our progress towards an adaptation to the GEMINI framework that is more ideally suited for the complex patterns of genetic variation inherent to the study of changes in tumor heterogeneity across multiple biopsies over the course of treatment. This is accomplished by enhancing GEMINI variant annotation to include tumor clonal specifications and allowing for filtering methods that reflect specific tumor properties. Additionally, by incorporating existing tools and databases that facilitate the interpretation of cancer mutations (e.g., CIViC, DGIdb, CRAVAT) into the GEMINI framework, the identification of mutations that may be driving clonal evolution is simplified. The advances to the GEMINI framework that we are developing will better enable cancer scientists and research oncologists to discern genetic details, at a clonal level, of how a tumor evolves over time, and better interpret these mutations in the context of potential patient care.

829T

Identification of disruptive germline and somatic variants in intronic splicing sequences and their confirmation by RNA expression in 2,489 cancer patients.

A. Polley¹, J. Little¹, S.C. Benz¹, S. Rabizadeh², J.Z. Sanborn¹. 1) NantOmics, LLC., Santa Cruz, CA; 2) NantOmics, LLC., Culver City, CA.

Pre-mRNA splicing is a complex process involving conserved sequence elements within the intron. The 5' and 3' splice sites are critical for proper splicing and their consensus sequences are highly conserved. Variants at these sites can result in disruptive events such as exon loss and intron retention, and are consequently implicated in many human cancers and diseases. Intronic sequences upstream of the 3' acceptor splice site, including the polypyrimidine tract (PPT) and branch point (BP) region, also play a critical role in the splicing reaction. However, the consequences of variants within these regions are less predictable due to lower sequence conservation in the human genome. We have developed a pipeline to study the impact of splice site mutations in both germline and somatic DNA, detect and investigate intronic mutations adjacent to the 3' splice site, and confirm their downstream effect on RNA splicing and expression. We ran 2,489 cancer patients' germline and somatic DNA and tumor RNA sequencing data through the pipeline against 328 cancer genes, including a subset of 22 cancer predisposition genes. Germline splice site variants were identified in cancer genes in almost all patients, and somatic variants in over 15% of patients. Additionally, 15% of patients had germline splice site mutations in the subset of cancer predisposition genes. Low RNA exon coverage was detected for splice site mutations in 5 patients, with a confirmed exon 16 deletion in the *ERBB2* gene of a patient with breast cancer. We identified the cause of this event to be a 3' splice site mutation (c.1854-2A>G). To estimate which intronic sequences could function as the BP and PPT, we created and searched for characteristic sequence identifiers in the reference genome and used this data to rank the strength of potential BPs in the set of cancer genes. We identified over 25,000 potential BPs, which we observed in 90% of genes in our set. Around 15% of the predicted BPs were categorized as being high-confidence. We then investigated variation in the high-confidence sequences and their impact on splicing. Our analysis demonstrates that the prediction of intronic splicing elements is an important supplement to a full splice site mutation analysis. Additionally, we highlight the benefit of using RNA exon coverage data to confirm abnormal splicing, which is an important driver of oncogenesis and plays an under-appreciated role in human disease.

830F

Beyond BRCA: Discovery of novel drivers of homologous recombination deficiencies.

N. Sahni¹, D.J. Mcgrail², Y. Li³, B. Feng³, L. Hu², G.B. Mills², S.Y. Lin², S. Yi⁴. 1) MD Anderson cancer center, Houston, TX. Baylor College of Medicine, Houston; 2) MD Anderson cancer center, Houston, TX; 3) TESARO Inc, Waltham, MA; 4) Dell Medical School, The University of Texas at Austin TX.

Introduction: In the two decades since the discovery of *BRCA1* and *BRCA2* mutations as risk factors for breast cancer, our knowledge of their function in maintaining genomic integrity through homologous recombination (HR) DNA repair and other pathways has greatly advanced, ultimately culminating in development of PARP inhibitors that can specifically target patients who harbour these mutations. However, BRCA mutations have only shown relevance in a subset of cancers and the discovery of additional drivers of HR deficiencies across cancers remains minimal. **Methods:** We integrated large-scale omics data from The Cancer Genome Atlas across 18 cancer types that display genomic evidence of HR defects to elucidate novel genomic drivers. Using network analysis, we predicted novel drivers of HR defects and experimentally validated these predictions across multiple experimental systems including HR reporter assays and Rad51 foci formation. **Results:** In addition to prediction of classical HR genes such as *BRCA1*, *BRCA2*, and *PALB2*, our analysis indicates that over half of HR deficiencies originate outside of canonical DNA damage response genes. Notably, the novel identified drivers were particularly enriched for genes involved in RNA splicing and processing. The predicted candidates were significantly enriched for cancer risk loci, over half of which were related to RNA processing. Experimental techniques validated over 90% of our predictions in a panel of 50 genes tested by siRNA as well as 31 additional engineered mutations identified from TCGA patients. **Discussion:** Taken together, our integrated computational analysis identified and validated over 50 drivers of HR deficiencies across numerous lineages. These novel drivers may serve as important biomarkers to help understand what patients may benefit from targeted therapies such as PARP inhibition. Moreover, this work may enable better understanding of cancer etiology, and advance screening approaches aimed at identifying individuals at risk for developing malignancies.

831W

An integrated systems analysis of miRNA target network associated with natural killer cell activation. G. Thillaiampalam¹, K. Bednarska¹, Q. Cui¹, F. Vari^{1,3}, M. Gandhi^{1,2}, A. Cristino¹. 1) University of Queensland Diamantina Institute, Translational Research Institute, University of Queensland, Brisbane, QLD, Australia; 2) Princess Alexandra Hospital, Brisbane, QLD, Australia; 3) QIMR Berghofer Medical Research Institute.

Natural killer (NK)-cells are immune effector cells that have the ability to kill cancer cells without prior sensitization. A number of cell surface receptors and transcriptional factors are known to control NK-cell activation. However little is known about the regulatory molecules involved in NK-cell activation. microRNAs (miRNAs) are regulatory RNAs (19 – 24 nt) critical for the control of gene expression via direct binding to 3'UTR of messenger RNAs (mRNAs) to suppress the protein synthesis. In this study we used the human NK-cell lines as a model system to identify the regulatory miRNAs by comparing the transcriptome of resting NK cells and two activation systems: direct cytotoxicity and antibody-dependent cell-mediated cytotoxicity. We identified genes involved in ER stress (XBP1), cell motility (XIRP1) and checkpoint receptors (PD1) up-regulated during activation. Our transcriptomic analysis revealed 83% of mRNA transcripts were up regulated while 81% of miRNAs were down regulated during both activation conditions which resulted in a putative miRNA - mRNA network. miR-34a-5p has been identified as a regulatory hub binding to 3'UTR of genes including XBP1, XIRP1 and PD1. We validated the direct binding of miR34a-5p to the predicted target sites in these genes by a dual-luciferase reporter system and the *in vivo* regulation over PD1 protein was validated by flow cytometry of NK cells transfected with mimics miR-34a-5p. Our findings provide detailed characterization of a miRNA regulatory network underpinning NK-cell activation against cancer cells and could open new avenues for miRNA-based therapeutics to enhance NK cell anti-tumoural function.

832T

Coordinating variant interpretation knowledgebases improves clinical interpretation of genomic variants in cancers. A. Wagner¹, B. Walsh², G. Mayfield², D. Tamborero³, D. Sonkin⁴, K. Krysiak¹, J. Deu Pons⁵, R. Duren⁶, J. Gao⁶, J. McMurtry⁶, S. Patterson⁷, C. Del Vicchio Fitz⁸, S. Li⁹, D. Ritter⁹, R. Dienstmann¹⁰, D. Chakravarty⁶, J. Beckmann¹¹, M. Baudis¹², M. Haendel¹, M. Lawler¹³, S. Mockus⁷, O. Elemento¹⁴, N. Schultz⁶, N. Lopez-Bigas³, J. Goecks², A. Margolin², M. Griffith¹, O. Griffith¹. 1) Washington University School of Medicine, St. Louis, MO; 2) Oregon Health and Science University, Portland, OR; 3) Institute for Research in Biomedicine, Barcelona, Spain; 4) National Cancer Institute, Rockville, MD; 5) Molecular Match, Houston, TX; 6) Memorial Sloan Kettering Cancer Center, New York, NY; 7) The Jackson Laboratory for Genomic Medicine, Farmington, CT; 8) Dana-Farber Cancer Institute, Boston, MA; 9) Baylor College of Medicine, Houston, TX; 10) Vall d'Hebron Institute of Oncology, Barcelona, Spain; 11) University of Lausanne, Lausanne, Switzerland; 12) University of Zurich, Zurich, Switzerland; 13) Queen's University Belfast, Belfast, Northern Ireland; 14) Cornell University, Ithaca, NY.

The promise of precision medicine in which a cancer patient's treatment is tailored to their specific cancer variants requires concise, standardized, and searchable clinical interpretations: biomarker-disease associations that provide therapeutic or prognostic value. While many institutions curate interpretations for clinically relevant cancer variants, these efforts are siloed and non-interoperable. We formed the Variant Interpretation for Cancer Consortium (VICC, cancervariants.org) as a Driver Project of the Global Alliance for Genomics and Health (GA4GH, ga4gh.org), and assessed six established resources of cancer variant interpretations (CIViC, OncoKB, JaxCKB, MMatch, PMKB, CGI). We compared the coverage of the biomedical literature between them, finding that 87% of the 6034 publications from the six resources were unique across the collective. This illustrates both the value gained from aggregating the knowledge of these resources and the enormity of the biomedical literature describing cancer variants. We integrated guidelines and ontologies describing genes, variants, diseases, drugs, and evidence to harmonize clinical interpretations of cancer variants. From this, the VICC collectively established a set of minimum elements required for describing a cancer variant interpretation, and used them to harvest and harmonize knowledge from the six founding resources. Our results demonstrate that the knowledge from the constituent resources of the VICC are highly heterogeneous across every element defining these interpretations (genes/variants, drugs, diseases, evidence). To address this, we have developed a harmonized database and a prototype web interface for exploring and querying this breadth of knowledge. We evaluated 110,320 somatic mutations of 17,005 patient donors in the AACR Project GENIE cohort against our VICC aggregate interpretation database. When stratified by evidence level (as described by the joint consensus recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists), the interpretations within the VICC aggregate knowledgebase provided 2-4x breadth of coverage across each level of evidence when compared to the median coverage from the constituent resources. Finally, we demonstrated how the clinically actionable interpretation of a patient mutation may change dramatically when considering the knowledge from the VICC resources in aggregate instead of individually.

833F

Machine learning-based genome-wide interrogation of copy number aberrations in circulating tumor DNA for non-invasive detection of early-stage hepatocellular carcinoma. Q. Wei¹, Z. Bian², Z. Ye³, Y. Wang², C. Wang³, C. Yin², K. Tao², H. Yang³, B. Li¹, J. Xing². 1) Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN; 2) State Key Laboratory of Cancer Biology and Experimental Teaching Center of Basic Medicine, Fourth Military Medical University Xi'an, China; 3) Department of Medical Oncology, Thomas Jefferson University, Philadelphia, PA.

Hepatocellular carcinoma (HCC) is a leading cause of cancer death worldwide. For HCC patients, detecting them in early-stage is a much higher prognosis. Recently the discovery of ctDNA (circulating tumor DNAs) in cancer patients makes it possible to develop blood-based non-invasive approaches of cancer detection. In this study, we carried out low-depth (5X) whole-genome sequencing (WGS) to analyze somatic copy number aberration (SCNA) profiles in 392 plasma samples of hepatitis B virus (HBV)-related HCC and cancer-free HBV patients, using three different cohorts, and developed a machine learning-based early-detection model to detect the HCC patients in early-stage. In a discovery cohort of 209 samples, leveraging the model we achieved an overall area under curve (AUC) of 0.894 (0.842 for stage I and 0.934 for stage II-IV). We then assessed the performance of our model in two independent validation cohorts with 78 and 105 samples respectively, consisting of mostly stage I patients (75%). Our model exhibited a robust predictive performance, with AUC of 0.898 and 0.788 in validation cohort 1 and 2, respectively. Meanwhile to increase the detection accuracy for early-stage HCC, we incorporated prior knowledge derived from external genomics data into the framework and improved the model performance in validation cohort 2. Overall, we developed an SCNA-based, machine learning-driven model in the non-invasive detection of early-stage HCC and demonstrated its performance through strict independent validations.

834W

Breast cancer type classification using machine learning approach. J. Wu¹, T. Mamidi¹, L. Zhang², E. Nicklow¹, H. Hicks¹. 1) Department of Genetics, Louisiana State University Health Sciences Center, New Orleans, LA; 2) Louisiana Tumor Registry, Louisiana State University Health Sciences Center, New Orleans, LA.

Background: Breast cancer (BC) is the most diagnosed cancer and the second leading cause of cancer-related death among women in the US. Majority of BC are ER positive (ERPBC) and respond to endocrine or targeted therapies. However, a significant proportion of BC are triple negative breast cancer (TNBC) the most aggressive form of BC with poor survival rates and poor clinical outcomes. Although current classification based on the three receptor markers (lack of estrogen receptor, progesterone receptor and lack of HER-2 amplification) is relatively effective in identifying patients with TNBC, significant challenges remain. BC is inherently a heterogeneous disease and accurate stratification of BC patients based on the molecular type remains a challenge for clinicians and oncologists. Accurate classification of breast cancer types could guide therapeutic decision making. Advances in high-throughput sequencing of BC genomes has led to an expansion in molecular classification of types of BC. Here we propose the use of machine learning (ML) for classification of BC patients as either having TNBC or ERPBC. We hypothesize that genomic alterations in individual patients with either TNBC or ERPBC could lead to lead to measurable changes distinguishing the two patient groups, and that this could be accomplished using ML. Methods: Our proposed classification method uses information from gene expression and clinical data we downloaded from The Cancer Genome Atlas (TCGA) consisting of 110 TNBCs, 922 ERPBCs and 53 control samples. We performed whole genome analysis comparing gene expression levels between the two patient groups (TNBC vs ERPBC) to discover significantly differentially expressed genes. From the significantly differentially expressed genes, we selected a small subset of genes for use as classifiers. We used two classification approaches, k-nearest neighbors (kNNs) and decision-trees (DTs) and evaluated their performance and classification accuracy. Results: The two methods accurately classified the patients as either TNBC or ERPBC, suggesting that ML could be used to accurately classify BC into the two types of BC. The overall classification accuracy was higher than 95% for both ML methods. Conclusions: The result demonstrate that ML could be used to accurately classify BC patients as either TNBC or ERPBC type using molecular markers from transcription profiling. Future work will involve mutation-based classification of BC into TNBC or ERPBC using ML. .

835T

Comprehensive analysis of genetic variations in patients with acute lymphoblastic leukemia. W. Yun¹, B. Kim², S.T. Lee², J.R. Choi². 1) Brain Korea 21 PLUS Project for Medical Science, Yonsei University; 2) Department of Laboratory Medicine, Yonsei University College of Medicine, Seoul, Korea.

Background Researches on genetic variations in patients with B-cell and T-cell acute lymphoblastic leukemia (ALL) are ongoing, but a comprehensive analysis of genetic variations including chromosomal and genic copy number variations is rarely reported. **Methods** We developed and applied next-generation sequencing (NGS)-based targeted gene panel and bioinformatics pipeline to profile the genetic changes in patients with acute lymphoblastic leukemia. We analyzed 185 genes in 87 patients with B-cell ALL (n=72) and T-cell ALL (n=15). Thirty five adults and 53 pediatric patients were included. **Results** Among 72 patients with B-ALL, 33 has chromosomal abnormality such as hyperdiploidy. A total of 197 genetic variants were detected and 149 of these were deletion or duplication of genes (75.6%). *CDKN2A/B* and *IKZF1* genes were most frequently deleted. Among single nucleotide variants and small indels, those of *TP53*, *RAS* and *PTPN11* were most frequently detected. Among 15 patients with T-cell ALL, variations in *PTEN*, *PHF6* genes were frequently observed along with *CDKN2A* deletion. Deletions of genes indicating poor prognosis, such as *IKZF1* and *EBF1* deletion was more frequent in adult patients. **Conclusion** Using NGS-based gene panel which can assess copy number variations as well as sequence variations, genetic variations of lymphoblastic leukemia could be better understood. Further study would be needed to connect the genetic variations to the clinical outcomes.

836F

Reconstruct pseudogene regulatory networks using natural language processing for novel drug target prediction. Y. Zhang¹, S. Tan^{1,2}, X.-M. Mo¹. 1) Department of Biomedical Informatics, College of Medicine, The Ohio State University, Columbus, OH; 2) School of Information Engineering, Hubei University of Economics, Wuhan, Hubei, China.

Background: Pseudogenes are non-coding DNA segments that are homologous to the coding genes. Although they have been thought of as "genomic fossils", recent studies have shown that a considerable number of pseudogenes are transcribed and some can even regulate gene expression. Some of the pseudogenes can regulate the parental genes' expression by competitively binding to microRNAs. Some others can generate competitive endogenous RNAs (ceRNAs) that can bind to the 3' UTR of the parental gene thus regulate gene expression. One of the regulatory pseudogenes includes PTENP1. The selective loss of PTENP1 can result in aberrant activation of PI3K/AKT signaling pathway and is associated with tumorigenesis in human. All these show that pseudogenes make a new class of modulators of gene expression. However, they are still understudied and there is a lack of functional annotations of pseudogenes. **Results and Conclusion:** We extract information about regulatory pseudogenes by analyzing literature using natural language processing (NLP) techniques. Based on this information, we reconstructed the regulatory networks involving pseudogenes and target genes with disease and tissue-specific annotation. We performed pathway analysis on the networks, based on which we were able to annotate the regulatory pseudogenes with enriched functional impact. We also overlaid drug-target network onto this regulatory network, and predict the potential treatment mechanisms that the pseudogenes are involved in. This not only filled the gap for functional annotation of pseudogenes but also can be used as a ground truth, upon which we can develop computational methods for pseudogene function prediction and cross-validation.

837W

Integrating genetic information with machine learning to predict which prostate cancer cases should not be immediately treated. *T.B. Cavazos¹, N.C. Emami¹, J.E. Cowan², M.R. Cooperberg^{2,3}, J.M. Chan^{2,3}, P.R. Carroll², J.S. Witte^{2,3}.* 1) Program in Biological and Medical Informatics, University of California, San Francisco, San Francisco, CA; 2) Department of Urology, University of California, San Francisco, San Francisco, CA; 3) Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA.

Advances in screening and treatment have greatly improved five-year survival for prostate cancer. Approximately 90% of men diagnosed with prostate cancer are characterized as having low-risk disease. Although important strides have been made towards accurate patient stratification, less has been done to help guide treatment decisions for this low risk group. While some patients may experience upgrading or upstaging (UGUS) of their tumor and benefit from more invasive treatments such as radical prostatectomy, others may have slower growing tumors that are unlikely to result in any increased risk of morbidity or mortality due to this disease. In this case, active surveillance may be the best treatment option as surgery or radiotherapy have a number of potential side effects. To better stratify and guide the medical decisions of low-risk prostate cancer patients, we created a random forest (RF) classifier to predict the likelihood of future UGUS of prostate cancer at diagnosis and prioritize germline variants associated with this phenotype. We analyzed genetic and clinical data from a cohort of 864 low-risk prostate cancer patients whom underwent radical prostatectomy 6 months after initial diagnosis. At radical prostatectomy, 448 patients did not experience a change in tumor stage or grade. The patients were split into a training (n = 690) and validation (n = 174) cohort. With clinical factors alone, including Gleason score and prostate specific antigen levels (PSA), UGUS was predicted with 70% accuracy using a generalized linear model. To identify prognostic signals independent of clinical factors, approximately 700,000 single nucleotide polymorphisms (SNPs) were genotyped in these subjects. Due to the large number of features and small sample size of this study, RF is an appropriate method for analysis given its ensemble learning approach that helps prevent model overfitting. By integrating genetic information and using a RF classifier, we will show the potential increase in prediction accuracy over the clinical model alone.

838T

Clinically actionable fusion genes and their detection through whole transcriptome sequencing (RNASeq™) from FFPE solid tumor biopsies. *H.K. Chandok, K. Kelly, D. Bergeron, G. Ananda, H.V. Reddi.* The Jackson Laboratory for Genomic Medicine, Farmington, CT., 06032.

Introduction: Fusion genes have played a major role in tumor initiation and progression. A number of Next Generation Sequencing (NGS) technologies have enabled detection of fusions from both DNA and RNA. Transcriptome analysis allows improved detection of expression changes, somatic aberrations and gene fusions. Here we describe the development of an in-house RNAseq-based approach for detection of a larger number of gene fusions in solid tumors as an upgrade to our current 53 gene assay. **Methods:** 13 FFPE tumor samples with known fusions were selected, which included both pure cell lines and synthetic controls. Libraries were prepared from purified RNA using the KAPA RNA HyperPrep Kit with RiboErase (HMR) and were subsequently sequenced using the Illumina NextSeq. Sequencing data was processed through in-house bioinformatics pipeline including tools such as SOAPfuse, FusionCatcher, and EricScript for the detection of fusion transcripts. Samples were subjected to clinical quality control (QC) monitoring through the various sequencing and analysis steps. The fusion transcript detection algorithms were evaluated for their precision, recall trade-off, and computational cost. **Results:** Evaluation of QC metrics revealed that all 13 samples prepped using the developed protocol passed clinical quality cut-offs for library preparation, sequencing, and bioinformatics analysis. The samples had ~53-70M reads of which 39-55M reads were of high quality. FusionCatcher outperformed the others by providing the highest sensitivity of 87% compared to 40% and 47% by SOAPfuse and EricScript respectively. FusionCatcher was able to successfully identify 21 of the 23 known clinically actionable fusions in the samples tested. Additionally, the tool called 125-320 unknown fusions per sample. We imposed a series of filters to eliminate false positives and non-cancerous fusions such as gene fusions present in population databases, predicted effect in coding sequence regions, reads supporting spanning and junction regions etc., and the unknown fusions which passed out filters are currently under validation. **Conclusion:** Analyzing high quality data produced by RNAseq libraries through our robust pipeline allows for the detection of clinically relevant gene fusions as good targets for therapeutic applications and personalized medicine. With gene fusions implicated in a variety of cancer types, offering high quality, low cost molecular testing services will serve to best benefit the patient.

839F

Predicting melanoma survival with deep neural networks. *L. Cordeiro, O. Lupi.* UNIRIO, Rio de Janeiro, Brazil.

The worldwide incidence of cutaneous melanoma has been increasing at a rate higher than the rate of any other tumor over the last several decades. While novel systemic therapies have revolutionized melanoma treatment in the past few years and five-year survival rates have been improving, survival can be lower than 10% in advanced cases. Since the early prognostic models of Clark and Breslow, several attempts to develop accurate survival models have met with limited success. Recent models incorporating different omics datasets and employing standard machine learning algorithms have achieved mixed performance. Deep learning is a novel class of machine learning algorithms that found many applications in medicine recently. The objective of this study is to experiment with this technique and develop a predictive survival model for melanoma using public gene expression data. We developed a semi-supervised learning model, i.e., one that learns representative features from unlabeled data – without survival information – and uses this representation to train a supervised model to try and predict clinical outcomes of interest. We constructed the unlabeled dataset by systematically mining public gene expression data and building a unique dataset of over 650 samples of normal skin and primary and metastatic melanoma. The unlabeled dataset was used to train a deep stacked autoencoder which fed a three-layer classifier. We used TCGA data as the labeled dataset in the supervised learning task. Based on available survival data, we categorized each sample as “deceased” or “alive” in five discrete time horizons – from one to five years – to try and predict whether a patient would be alive at a particular time horizon, given the global expression profile of her tumor at the moment of diagnosis. To assess model performance, we performed 1000 iterations of a repeat cross-validation procedure. All computations were done in MATLAB. Prediction performance was poor in shorter horizons, most likely due to severe class unbalance, though in longer time horizons it reached accuracy levels currently seen in the literature. Model accuracy hovered around 65% in $T = 4$ and $T = 5$ years, with sensitivity and specificity in the 63%–65% range. Further modeling will integrate deep autoencoders with more sophisticated techniques such as regularized Cox regressions and random survival forests, allowing us to identify features that significantly impact survival.

840W

Immune cell tumor infiltration in TCGA: A genome-wide association study. *E.A. Daly, A. Mitchell, J. Hutz.* Human Biology and Data Science, Eisai Pharmaceuticals, Andover, MA.

Immunotherapy is a promising new treatment for many types of cancer that have not responded well to chemotherapy or traditional drugs. However, patient response to immunotherapy is still difficult to predict, and understanding underlying differences among these cancers is key to improving our treatment of them. In recent literature, infiltration of different immune cell types has been associated with mutation load and survival in multiple cancers. We sought to explore using GWAS whether there is a heritable contribution to an individual's immune response to cancer. We have explored through quantitative infiltration data generated by TIMER <<https://cistrome.shinyapps.io/timer/>> and germline genetic data from TCGA <<https://cancergenome.nih.gov/>> genetic variations in cancer patients leading to unusually high or low infiltration in six immune cell types: macrophages, neutrophils, dendritic cells, CD4 T-cells, CD8 T-cells, and B-cells. We utilized a standard pipeline including filtering for rare alleles (minor allele frequency > 0.02), missing genotyping rate < 0.02 , missing rate for loci < 0.05 , and Hardy-Weinberg equilibrium p -value $> 1e-6$, greatly reducing the noise in the results. As a positive control, we expected to see variants in the MHC controlling some variation in immune cell presence in tumors. Results for B-cells showed this most prominently with three variants in the MHC region with a p -value $< 1e-5$, but scattered signal for the MHC was present in all immune results, providing a level of confidence in the findings. Dendritic cell infiltration, which has been associated with mutation load in breast cancer, had a strong association with two SNPs very close in the same region of chr15 (99827439 and 99827447). Additionally, dendritic cell findings were very closely replicated in CD8 T-cell findings. This analysis is exploratory, but has promise in gaining a better understanding of the role of immune cells in cancer and the genetic profile of different tumors. We seek to replicate these findings using other methods of calculating tumor infiltration (outside of TIMER). With a better picture of genetic effects on immune cell infiltration in various cancer types, we may be able to better predict how patients will respond to immunotherapy.

841T

Multiscale mapping of the functional architecture of cancer systems. *M. Flagg^{1,3}, F. Zheng^{1,3}, M. Yu^{1,3}, K. Ono^{1,3}, M. Kim^{2,3}, D. Swanney^{2,3}, B. Demchak^{1,3}, J. Kreisberg^{1,3}, N. Krogan^{2,3}, T. Ideker^{1,3}.* 1) Department of Medicine, University of California San Diego, La Jolla, CA; 2) Department of Cellular and Molecular Pharmacology, University of California, San Francisco, CA; 3) The Cancer Cell Map Initiative (CCMI), La Jolla and San Francisco, CA.

The modular, interdependent multi-gene systems which govern cancer remain poorly understood. It is well-established that any given cancer gene operates in multiple distinct cellular contexts. However, it remains unclear how to represent these diverse roles in a single comprehensive model with predictive capabilities and robust multi-omic data integration. To that end, we integrate many physical and functional networks from a wide range of molecular studies, assembling a comprehensive multiscale map of human cancer's functional systems. This map organizes protein complexes, signaling pathways, and inter-pathway crosstalk in a single hierarchy of cellular subsystems. The model's many uncharacterized modules offer intriguing hypotheses with robust data support; in examining these modules, we predict novel functional roles for hallmark cancer genes, established biomarkers of unclear function, and wholly uncharacterized genes. Applied to somatic mutation profiles in The Cancer Genome Atlas (TCGA) patients, this map reveals patterns of subsystem alteration, identifying systems targeted by mutations at multiple scales above the individual gene. This map facilitates interpretation of omics data in terms of affected subsystems; we integrate new protein-protein interactions identified with AP-MS to elucidate the selective rewiring of these modules in multiple breast cancer cell lines. This map also enables integration of mutation and clinical data in TCGA patients across many cancer types, bridging genotype and phenotype to connect patient survival, treatment resistance, and tumor histology to subsystem alteration. By linking clinical course to underlying biology through the lens of this model, we identify novel cancer subtypes, offering a framework for personalized medicine approaches. To facilitate model interpretation, we created HiView, a web-based hierarchy browsing tool, which allows users to visualize and navigate hierarchies, contextualizing each module by displaying the underlying data and relevant entries in gold-standard databases.

842F

Tumor gene expression profiles segregate across primary sites. *F.G. Frost¹, D. Condon¹, P. Cherkuri¹, L. Charmichael¹, M. Dogramaci², S. Milanovich¹, C. Boerkoel¹, UCSC Treehouse Childhood Cancer Initiative.* 1) Sanford Imagenetics, Sioux Falls, SD; 2) Sanford Children's Hospital, Sioux Falls, SD.

Background: Cancer classification according to tumor primary site and histopathology have helped define tumor treatment and prognosis for over 200 years. With increasing knowledge and improving technology, many tumors are now subdivided according to histopathology as well as by protein and/or nucleic acid biomarkers. The latter has shown that tumors of apparently disparate histopathology can have similar gene expression signatures, and that tumors with similar histopathology can have distinct gene expression signatures. Based on either the Cancer Stem Cell (CSC) or the Dedifferentiation Theories of tumorigenesis, one might postulate that there are a finite number of ways for cancer to develop and that these mechanisms act across multiple primary tissue types. **Remaining Question:** This raises the question of whether such conserved cancer promoting pathways are identifiable and whether they do indeed transcend primary tissue type. **Hypothesis:** Transcriptomic phenotypes from diverse tumors segregate across tumor primary sites. **Methods:** Transcriptomes derived from both normal and tumor tissues were obtained from The University of California, Santa Cruz (UCSC) Treehouse Project's publicly available dataset. Principle component analysis and hierarchical clustering was performed with and without subgrouping using metadata. **Results:** Analysis of histopathologically-grouped tumors showed that no two groups were significantly similar. Subsequent analyses with agnostic clustering-based groups show incomplete independence of primary site and transcriptomic phenotype. The transcriptomic phenotypes of the clusters included classical cancer pathways and/or metabolomic signatures. These signatures were generally obscured in the absence of correction for primary tissue gene expression through tumor/normal ratios. **Conclusions:** Although there are exceptions, cancer-specific gene expression profiles frequently segregate across primary sites. Subtraction of remnant normal tissue expression appears to greatly improve both functional annotation of each cluster and separation of outgroups.

843W

Integrated genomic, epigenetic, and expression analyses of ovarian cancer cell lines. D. Hallberg¹, E. Papp¹, G.E. Konecny², D.C. Bruhm¹, V. Adleff¹, M. Noë¹, I. Kagiampakis¹, D. Palsgrove¹, D. Conklin², Y. Kinose³, J.R. White¹, M.F. Press⁴, R. Drapkin³, H. Easwaran¹, S. Baylin¹, D. Slamon², V.E. Velculescu¹, R.B. Scharpf¹. 1) Johns Hopkins School of Medicine, Baltimore, MD; 2) Jonsson Comprehensive Cancer Center, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA; 3) Penn Ovarian Cancer Research Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; 4) Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA, USA.

Ovarian cancer is the most common cause of death among gynecological cancers. Despite significant advances in therapies for other solid tumor malignancies, the overall survival of patients with late-stage ovarian cancer has remained dismal with few new options for treatment. The standard therapy involves debulking surgery followed by chemotherapy. Part of the reason for the lack of novel therapies for ovarian cancer has been an inadequate understanding of the underlying molecular characteristics of this disease, especially in the context of cancer cell models than can facilitate the development of various cancer treatments. To improve our understanding of ovarian cancer, we performed genome-wide sequencing analyses of 45 ovarian cancer cell lines of varying subtypes. Given the challenges of genomic analyses of tumors without matched normal samples, we developed a novel approach for detection of somatic sequence and structural changes called Trellis, and integrated these with epigenetic and expression alterations. Alterations not previously implicated in ovarian cancer that are biologically and clinically relevant included amplification or overexpression of *ASXL1* and *H3F3B*, deletion or underexpression of *CDC73* and TGF beta receptor pathway members, and rearrangements of *YAP1-MAML2* and *IKZF2-ERBB4*. Dose-response analyses to targeted therapies revealed novel molecular dependencies, including increased sensitivity of tumors with *PIK3CA* and *PPP2R1A* alterations to PI3K inhibitor GNE-493, *MYC* amplifications to PARP inhibitor BMN673, and *SMAD3/4* alterations to MEK inhibitor MEK162. Furthermore, we found that genome-wide rearrangements provided an improved measure of sensitivity to PARP inhibition rather than the currently used HRD score. This study provides a comprehensive and broadly accessible resource of molecular information for development of new therapeutic avenues in ovarian cancer.

844T

The genomic landscape and pharmacogenomic interactions of circadian genes in cancer chronotherapy. L. Han¹, Y. Xiang¹, J.S. Takahashi², G.B. Mills², S.H. Yoo¹, Y. Ye¹. 1) The University of Texas Health Science Center at Houston, Houston, TX; 2) The University of Texas MD Anderson Cancer Center, Houston, TX; 3) University of Texas Southwestern Medical Center, Dallas, TX.

Cancer chronotherapy, treatment at specific times during circadian rhythms, endeavors to optimize anti-tumor effects and lower toxicity. However, comprehensive characterization of clock genes and their clinical relevance in cancer is lacking. We systematically characterized the alterations of clock genes across 32 cancer types by analyzing data from TCGA, CTRP, and GDSC databases. Expression alterations of clock genes are associated with key oncogenic pathways, patient survival, tumor stage and subtype in multiple cancer types. Correlations between expression of clock genes and of other genes in the genome were altered in cancerous versus normal tissues. We identified interactions between clock genes and clinically actionable genes by analyzing co-expression, protein-protein interaction and ChIP-seq data, and also found that clock gene expression is correlated to anti-cancer drug sensitivity in cancer cell lines. Our study provides a comprehensive analysis of the circadian clock across different cancer types and highlights potential clinical utility of cancer chronotherapy.

845F

HLI Cancer Search: A scalable, cloud-based search engine for cancer genome analysis. A. Harley, E. Shinbrot, C. Lau, A. Gylfe, W. Delport, R. Trivedi, T. Lacey. Human Longevity Inc., San Diego, CA.

BACKGROUND: In the era of precision medicine with thousands of cancer genomes sequenced worldwide, information required to perform comprehensive genome analysis is scattered through publications, professional guidelines, and web-based sources. There is a growing need for informatics tools to efficiently centralize and present genomic and associated phenotypic data in a user-friendly fashion. Here, we described the development of HLI Cancer Search -- a multi-omics search engine that integrates various sources of genomic data and ranks cancer alterations based on a set of criteria. This tool allows the user to create their own inquiries with dynamically created results, in contrast similar tools only allow the user to query static data sets for specific questions. **METHODS:** Cancer Search consists of a proxy, query user interface (UI) and the index. The data indexed can come from either HLI somatic sequencing pipeline or from an external partner in the form of a VCF or other tabular formats. Cancer Search utilizes responsive web-based interface with two factor authentication to ensure data protection. **RESULTS AND CONCLUSIONS:** Cancer Search provides millisecond access to data for both individuals and cohorts of samples. Currently we have indexed 400 million distinct variants (out of which 361,000 have annotations related to clinical significant genes and pathways). Patient-level analyses include: 1) comprehensive assessment of single nucleotide variants, copy number alterations, RNA sequencing data, structural rearrangements in both tumor and normal genomes; 2) integrated annotations of 17 data sources including drug labels and guidelines; and 3) advanced genomic analytics and machine learning model predictions. Cohort-level analytics include 1) stratification based on integrative analysis of genomic and phenotypic data; 2) data quality assessment; and 3) advanced genomic analytics including cancer driver analysis, and comparison to other large cancer cohorts. HLI Cancer Search allows analysis to extend beyond both static clinical reports and pre-computed cancer portal analyses to offer versatile and flexible analyses. The system facilitates in depth cancer genome analysis, as well as provides a framework for clinical biomarker discovery by enabling interactive, integrated, large-scale analysis of cancer genomic data at the point-of-care.

846W

Inferring the evolutionary order of somatic mutations with RankNet. W. Hsieh, W. Lai. National Tsing Hua University, Hsinchu City, Taiwan.

Tumors are well-known for its massive accumulation of mutations along its development. Computational biologists are especially interested in reconstructing the evolutionary history of the somatic mutations in tumors. The issue differs from the phylogenetic/coalescence tree construction of human population because the data is a mixture of cells and there is no clear trace of the coexistence of variants in the same cells. We propose a procedure to decompose a tumor into well-separated subclones according to the allele frequencies of different sets of mutations. The mutations representing the emergence of the same subclone are arranged in an evolutionary tree with the neighbor-joining method according to the subclone proportions in each tumor. The root of the tree is selected by the orders derived from the RankNet across multiple samples. The strategy helps to identify the recurrent mutations across multiple tumors and to trace the driver mutations that potentially propagate the downstream variations.

847T

Biological and functional impacts of tumor suppressor gene inactivation events: A pan-cancer study. P. Jia¹, Z. Zhao^{1,2,3}. 1) Center for Precision Health, School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA; 2) Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA; 3) Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN 37203, USA.

In cancer, activation of oncogenes and inactivation of tumor suppressor genes (TSGs) are two major driving forces. While oncogenes have been intensively studied and served as promising candidate drug targets, a systematic investigation of the molecular profiles and functional impacts of TSGs, however, is lacking from large cancer studies such as The Cancer Genome Atlas (TCGA). TSGs are difficult to study because their impacts can be easily confused with or overwhelmed by oncogenes. So far, we have collected more than 1000 human genes that were previously reported as TSGs or playing tumor suppressor roles through our Tumor Suppressor Gene database (<https://bioinfo.uth.edu/TSGene/>). In this study, we proposed a novel mathematical model to detect TSG inactivation events based on the widely recognized Two-Hit model. Our method integrates somatic mutation and copy number variation status, accounts for both per-sample mutation load and per-gene mutation rate, and removes sensitive factors that often lead to TSG inactivation but without major driving impacts (e.g., neighbor gene co-deletion by large CNV segments and genes close to telomere or centromere). In total, we identified 276 high confident TSG×cancer events in 28 cancer types using ~8000 TCGA samples. The majority of these events showed significant downstream impact in the form of reduced expression of the TSG itself (*cis*-effect), disturbance on the transcriptome (*trans*-effect), or combinatory effects. Surprisingly, we found many TSGs were associated with large numbers of differentially expressed genes, implying a global disturbance to the transcriptome. Importantly, we found that the expression patterns associated with TSG inactivation events exceeded the strong restriction from cancer lineage. For example, inactivation of the same TSG in different cancer types often formed their own clusters rather than different TSGs in the same cancer types, indicating that the same TSG drove similar biological processes in different cancers. In addition to the well-known TSGs, such as *TP53*, *RB1*, *PTEN*, and *CDKN2A*, we also found a group of chromatin modeling genes that were frequently and recurrently inactivated in pan-cancer, such as *ARID1A*, *ARID2*, and *KDM6C*, although their functional impact is yet to be revealed. In summary, our work systematically quantified the landscape of TSG inactivation events and revealed their transcriptional and signaling footprints; the findings complemented with the oncogenesis in cancer.

848F

Molecular drivers of localized prostate cancer. K. Kasaian¹, T. Yamaguchi¹, H. Zhu¹, J. Livingstone¹, C. Jung², P. Georgeson², J. Penington², A. Papenfuss², N. Corcoran², J. Reimand¹, C. Hovens², P. Boutros¹, Canadian Prostate Cancer Network, Australian Prostate Cancer Network. 1) Ontario Institute for Cancer Research, Toronto, Canada; 2) Peter MacCallum Cancer Centre, Victorian Comprehensive Cancer Centre, Melbourne, Australia.

A comprehensive whole genome sequence analysis of several hundred prostate cancer tumors and their matched normal tissues were performed, the cohort was representative of all stages and grades of localized disease. We systematically identified all possible alterations of the nuclear genome, including single nucleotide variants, small insertions and deletions, large structural variants and regions of copy number alterations. These were supplemented by an analysis of the mitochondrial genome for recurrent somatic mutations. Small driver mutations were elucidated using a novel algorithm that corrects for the local background mutation rate and considering known functional elements of the genome. Previously unknown driver mutations in regulatory regions of known cancer genes were found. Similarly, recurrent regions of copy number alteration and loss of heterozygosity were examined for their potential role in tumorigenesis. Regions of hypermutation, defined as statistical over-representation of small mutations and double-stranded breaks, were flagged as putative drivers. Hypermutated regions demonstrating a skew in composition of base changes towards cytosine to thymine alterations in the context of TpC dinucleotides were indicative of kataegis; strong association of such regions with the presence of double-stranded breaks were observed suggesting that APOBEC-mediated repair of the DNA breaks can lead to an elevated number of mutations with defined signatures. Copy number alterations, loss of heterozygosity and structural variant calls were superimposed to look for regions of oscillating copy number and possibly zygosity states which also correlate with the presence of structural break points, such regions across chromosome arms or whole chromosomes are evidence of chromothripsis. Recurrent events were found in about 1% of localized prostate tumours. Association of mutational features and putative driver events with clinical data such as age at diagnosis, T category, PSA level, Gleason grades, biochemical recurrence and cribriform or IDC pathology were performed and somatic events indicative of more aggressive forms of disease were identified.

849W

Implementing the latest clinical variant guidelines through evolution of a cancer variant interpretation interface. K. Krysiak, AM. Danos, AH. Wagner, JF. McMichael, S. Kiwala, AC. Coffman, NC. Spies, L. Kujan, EK. Barnell, KA. Clark, YY. Feng, BJ. Ainscough, ZL. Skidmore, CA. Ramirez, JA. Neidich, M. Griffith, OL. Griffith. Washington University School of Medicine, Saint Louis, MO.

CIViC (Clinical Interpretations of Variants in Cancer) is a crowd-sourced, open-access knowledgebase for curating the diagnostic, prognostic, predisposing, and therapeutic utility of cancer variants in a structured, computationally-accessible manner. Evaluation of clinical utility requires staying abreast of the standards, regulations, and resources available for the use of genomic data in clinical applications as well as the expanding medical literature. While the literature alone has created an interpretation bottleneck, the complexity and pace of changes to clinical guidelines and standards have increased disparities between guidelines and variant interpretation resources. CIViC tackles this bottleneck through a mixed gatekeeper model, allowing significant crowd-sourced contributions from a wide range of trainees, researchers, and clinicians—all moderated by field experts. This model enables knowledge curation at a pace beyond what a single organization could achieve while reducing time and curation burden on domain experts. CIViC has contributions from >120 users around the world on 346 genes, 1852 variants, and 413 drugs across 226 diseases. These structured data provide coordinate representations for variants described in the literature and seamlessly describe both germline and somatic alterations with precise ontological annotations. Newly implemented *Assertions* are summarized interpretations that incorporate new standards (ACMG and AMP variant interpretations, FDA approvals, and NCCN guidelines), maintain evidence source provenance, enable real-time reassessment, and promote submission to ClinVar. Created with the same principles as other CIViC entities, Assertions are structured to be both human-readable and programmatically accessible, in contrast to many important constituent resources (e.g., NCCN guidelines). An Assertion can be created for highly specific variants that can be submitted to ClinVar (e.g., EGFR T790M) or more generic variants unsupported by many resources (ALK fusions), maximizing flexibility and guideline incorporation. In collaboration with fellows undergoing molecular genomics training at Washington University, we are improving our data model to support complex variant combinations (genotypes) and developing this system for better clinical workflow integration. The integration of clinical guidelines into Assertions as well as a revised data model, user interface design, and workflow supporting these updates will be presented.

850T

Tumor-only microsatellite instability classification. C.K. Lau, A. Harley, E. Shinbrot, A. Gylfe, R. Trivedi, T. Lacey. Human Longevity Inc., San Diego, CA., USA.

With recent approvals of immunotherapy drugs for the treatment of mismatch repair deficient colorectal cancer, it has become of outmost clinical importance to determine microsatellite instability (MSI) in tumors. Traditional methods to determine MSI status is by mismatch repair protein immunohistochemistry, as well as PCR and fragment analysis at the standard five NCI-recommended loci. Recently, several computational tools for determining MSI utilizing next-generating sequencing data have been introduced, among them are mSINGS (Salipante, Clin Can Res 2014), MSIsensor (Niu, Bioinformatics 2014), MOSAIC (Hause, Nature Medicine 2016), and MANTIS (Kautto, Oncotarget 2017). However, one limitation that these methods shared is the requirement of paired normal data. There are a variety of common scenarios in which matched normal tissue is not available for sequencing. **METHODS:** Here, we present a new MSI status detection method using tumor-only sequences. Our method was trained and tested using data from public resource -- the Cancer Genome Atlas (TCGA). First, we examined the distributions of microsatellites, then we built a cross-validated classifier for MSI detection. **RESULTS AND CONCLUSIONS:** Our tumor-only classifier achieves precision of 0.96, recall of 0.84, and the performance was robust with loci panels of varying size. When trained on tumor and normal data, the precision of our new method is 0.99, and recall of 0.98. Future analysis includes testing the classifier on MSI positive and negative patient tumors sequenced at HLI. The tumor only MSI classifier can be used to generate clinically meaningful results by identifying patients that might benefit from immunotherapy treatment.

851F

Comprehensive analysis of genomic alterations for lymph node metastasis in gastric adenocarcinoma revealed the critical role of PI3K/AKT/mTOR pathway. S. Lee¹, Y. Song². 1) Department of Hospital Pathology, Seoul St Mary's Hospital, The Catholic University of Korea, Seoul, Korea; 2) Department of Pathology, Hanyang University College of Medicine, Seoul, Korea.

BACKGROUND: Gastric cancer is the second cause of cancer mortality worldwide and one of the most frequent cancers in Far East Asian and Korean populations. Early diagnosis is crucial to the prognosis of stomach adenocarcinoma (STAD) patients, which constitute > 90% of gastric cancer. However, a large proportion of STAD patients are diagnosed at advanced stage, resulting in a poor survival rate. In STAD, > 50% of patients have lymph node metastases (LNM) at diagnosis. The presence or absence of LNM is one of the most important prognostic factors following curative resection. This study aimed to assess the differences of the gene expression, DNA copy number and methylation profiling between presence and absence of LNM in STAD samples. **METHODS:** The clinical data, gene expression, CNV and DNA methylation data of STAD were retrieved from the Genomic Data Commons Data Portal. A total of 443 STAD samples were enrolled. **RESULTS:** A total of 10153 differentially expressed genes were retrieved from 371 STAD cases, including 110 LNM(-) and 261 LNM(+) samples. Gene Ontology enrichment analysis revealed several significantly over-represented terms, including cancer related biological process such as PTEN signaling, PI3K/AKT signaling and G-protein coupled receptor protein signaling pathway. Kyoto Encyclopedia of Genes and Genomes pathway enrichment analysis showed that the gene sets in PIK3-AKT, RAS, MAPK signaling pathway were over-represented in LNM (+) samples. Through the analysis of mutation annotated files, we highlighted 118 highly recurrent mutated genes which are associated with PTEN signaling. *PIK3CA*, *PREX2*, *KRAS* and *MAGI2* were commonly mutated genes related to PTEN signaling. From somatic copy number profiles, the deleted genes were associated with pathways related to several infectious diseases, complement system including C1 complex and TGF-beta signaling cascades. The pathways enriched with amplified gene sets also showed association with pathways related to infectious diseases and immune system. Genome-wide view from the methylation analysis revealed no difference between the two groups. **CONCLUSION:** In summary, the PI3K/AKT/mTOR pathway is activated in STAD, which is consistent with previous reports. Our results showed that they might be also critical for LNM in STAD, which suggests a direction for further studies.

852W

Gene regulation and cell-cell communication underlying intra-tumoral heterogeneity in HCC liver cancer. B. Losic^{1,2,3,4}, A. Craig^{2,5}, J. Villacortes⁷, J. van Felden⁶, A. Villanueva^{2,3,5}. 1) Icahn Institute for Genomics and Multiscale Biology; 2) Department of Genetics and Genomics, Icahn School of Medicine at Mount Sinai; 3) Graduate School of Biomedical Sciences, Icahn School of Medicine at Mount Sinai; 4) Tisch Cancer Institute, Immunology; 5) Tisch Cancer Institute, Liver cancer; 6) Division of Liver Diseases, Department of Medicine, Liver Cancer Program, Mount Sinai School of Medicine; 7) Boston University.

Intra-tumoral heterogeneity (ITH) in HCC liver cancer is strongly implicated in its treatment resistance and relatively dismal prognosis. We recently initiated an ITH pilot study which multi-regionally sampled 14 treatment-naïve patients and harvested 79 tumor biopsies, along with matched adjacent normal samples. Twelve patients were analyzed using bulk omics to obtain readouts of HLA allele specific expression, neo-antigen burden from somatic mutations, and TCR profiling using VDJ sequencing. Immune editing was observed in the neo-antigen landscape with subclonal mutations recruiting the dominant TIL response, which suggestively correlated with clinically observed clonal de-differentiation. To further investigate these clonal ecosystem signals and deconvolve impurity and cell-type effect confounders, we examined 38,564 cells from the remaining two patients (7 biopsies) with 10X Chromium scRNA-seq. As expected the clonal ecosystem consisted of a complex balance of cancer associated fibroblasts, myeloid derived, endothelial, B and NKT cells, with quantifiable differences between the two patients. Notably, HCC cells clustered in multiple well-known HCC molecular classes in one patient, but only in one class in the other patient. To estimate the gene regulation ultimately responsible for this variability, we computed celltype specific co-expression networks enriched in the binding motif of key transcription factors, and binarized the activity of these networks in each cell. We found that HCC cells had remarkably regionally specific network activity and transcription factor activation, with pluripotency signaling orchestrated by the Yamanaka factor OCT4 dominating in a poorly differentiated region of the variable patient. Surprisingly, we also uncovered strong evidence that cancer-associated fibroblasts had regionally specific transcription factor activation patterns as well, suggesting highly dynamical and intertwined co-evolution between HCC clones and their microenvironment. We extensively confirm this via immunofluorescence, leading to improved, clinically important estimates for the sampling bias of a single biopsy in HCC. .

853T

HPV-related human cancers exhibit human-viral hybrid extrachromosomal circular DNA-like structures. *N. Nguyen¹, J. Pang¹, G. Xu¹, V. Deshpande², V. Bafna¹, J. Califano¹.* 1) University of California San Diego, La Jolla, CA; 2) Illumina, La Jolla, CA.

The integration of viral sequences into the host genome is a key driver of tumorigenesis in Human Papillomavirus (HPV) mediated cancers, such as head and neck squamous cell carcinoma (HNSC) and cervical cancer. Viral integration is reported to cause genomic amplification and structural variations near the integration site on the human chromosome, though mechanisms for these changes are unclear. A recent study found extrachromosomal circular DNA (ecDNA) in 40% of human cancer cells, but not normal cells, and that key driver oncogenes were typically amplified on ecDNA, providing a new model for oncogenesis (Turner et al., 2017). We propose that HPV-mediated cancers can contain hybrid human-viral ecDNA (HV-ecDNA) that not only allow for amplification of the human and viral oncogenes, but can also result in indiscriminate and unregulated transcription in proximal human genomic regions downstream of the viral regulatory elements. Previous work using The Cancer Genome Atlas cervical cancer (TCGA-CESC) dataset revealed that regions proximal to HPV integration had a 4-fold higher mRNA expression compared to the same regions without proximal integration (Nguyen et al., 2018). We applied computational methods developed in our lab on 68 cervical cancer samples from the TCGA-CESC dataset and 47 patient-derived HNSC samples to examine the prevalence of HV-ecDNA in cervical and HNSC samples, and whether HV-ecDNA had higher mRNA expression than expected. Our work reveals that 29% of TCGA-CESC samples with HPV integration displayed cyclic hybrid human-viral genomic structures, reminiscent of ecDNA. Human regions on these hybrid structures had an average 5-fold genomic amplification, but an average 18-fold higher mRNA expression compared to the same regions not on hybrid structures. The results on the patient-derived HNSC samples showed strong concordance with the TCGA-CESC results. Of the 23 HNSC samples with HPV integration, 26% displayed HV-ecDNA, with the structures having an average 7-fold genomic amplification, and average 15-fold higher mRNA expression relative to the same regions not on the structure, and a 7-fold higher mRNA expression than regions with proximal integrations but not on any hybrid structures. These structures can display complex human-viral chimeric transcripts with novel splicing patterns, high expression of ncRNA and intergenic regions, and may provide new insights in HPV-mediated tumor progression.

854F

Identification of profiles of mutation and CNA predictive of survival in patients with lung adenocarcinoma. *K.Y. Oróstica¹, A. Olivera-Nappa¹, G. Navarro², J.A. Asenjo¹, R. Armisen³.* 1) Department of Chemical Engineering and Biotechnology, University of Chile; 2) Department of Sciences of Computer Science, University of Chile; 3) Research Center of Excellence for Precision Medicine Pfizer.

Lung adenocarcinoma (LUAD) is one of the most frequent subtypes of Non-Small-Cell Lung Carcinoma (NSCLC), reaching up to 40% of all NSCLC. In addition, LUAD is one of the most aggressive subtypes, where the five-year survival is the lowest compared to other cancers. With the arrival of Next Generation Sequencing (NGS) new therapeutic strategies have been developed, based on the study of the genomic alterations present in tumour cells, which could have an important predictive value of the clinical response. Thanks to this approach, targeted therapeutic strategies have been implemented, based on specific driver mutations present in oncogenes, with lower levels of toxicity and more effective than traditional chemotherapy. However, cancer continues to progress due to the resistance acquired by the targeted treatment. Although the molecular classification of LUAD is a great advance for the development of new therapies, more systematic studies are needed for determining how genomic and clinical characteristics are related to the clinical response of patients with cancer. In this work, we propose to identify specific profiles of co-occurrence and mutual exclusion that alter the clinical response to treatment based on Mutational information and Copy Number Alteration (CNA) in patients with LUAD.

855W

Estimation of Microsatellite Instability (MSI) using Ion Torrent Sequencing technology. A. Pankov¹, A. Zheng¹, S. El-Difrawy¹, J. Au-Young¹, A. Kraltcheva², A. Kamat¹. 1) Thermo Fisher Scientific, South San Francisco, CA; 2) Thermo Fisher Scientific, Carlsbad, CA.

Introduction Immunotherapies to treat malignancies have shown great promise, however, only a fraction of cancer patients respond to immunotherapy, making it imperative to identify biomarkers that predict response. The only two FDA- approved companion diagnostics for immunotherapy treatment require assessment of PD-L1 expression by immuno-histochemistry and MSI status by CE or NGS. Microsatellites are tandem repeats of short DNA sequences (1-6 bp), abundant throughout the human genome. The MSI phenotype in tumors is observed as modifications of the length of repeats induced by DNA polymerase slippage and the inability to correct this slippage due to disruptions in the cancer's mismatch repair machinery. In particular, mononucleotide repeats of greater than 12 nucleotides in length and dinucleotide repeats greater than 10 repeats in length are shown to be most-highly associated with MSI phenotypes across various solid tumors types. Hence these long repeats are considered as candidate markers to detect MSI-H phenotype. **Methods** In this study, we identified 80 (71 dinucleotide and 9 mononucleotide) candidate MSI markers from the literature and selected additional 152 repeats from Exome and Tumor Mutation Load research panels. Concurrently, we developed a method to correct the signal errors in flow space and not in base space to allow for accurate length estimation of some dinucleotide repeats. Using a set of 40 CRC MSI-H research samples, we leveraged information theory to identify the recurrently aberrant repeat loci, and create a predictive algorithm. **Results** To show that our selected repeats may be informative of MSI status, we used statistical model to predict a set of useful sites and further demonstrated the informativeness of the markers using capillary electrophoresis on a separate data set. With a set of well-performing markers, we corrected the sequencing signal and developed a method of identifying aberrant mononucleotide markers against a control sample and an information theoretic approach for a tumor-only classifier using the dinucleotide markers. **Conclusions** In summary, we are developing a novel approach to identifying MSI loci and classifying tumor research samples based on the MSI status. This allows us to use Ion Torrent sequencing technology to identify changes in complex repeat regions of the genome.

856T

Linkage mapping and structural variant analyses using whole genome sequencing (WGS) identify a novel germline deletion in *ETV6* as the cause of acute lymphoblastic leukemia in an Australian pedigree. E.R. Rampersaud¹, D.Z. Ziegler², I.I. Iacobucci³, D.T. Payne-Turner⁴, M.C. Churchman³, I.M. Moore⁵, K.A.S. Schrader⁶, K.T. Tucker⁶, R.S. Sutton⁶, M.W. Warby², V.Q. Qu³, V.L. Levantaki³, H.B. Berns⁶, S.P. Pelletier⁷, G.W. Wu¹, C.M. Mullighan³. 1) Department of Computational Biology, St. Jude Children's Research Hospital; 2) Kids Cancer Centre, Sydney Children's Hospital; 3) Department of Pathology, St. Jude Children's Research Hospital; 4) Department of Medical Genetics, University of British Columbia; 5) Children's Cancer Institute Australia; 6) Transgenic/ Gene Knockout Shared Resource, St. Jude Children's Research Hospital; 7) Department of Immunology, St. Jude Children's Research Hospital.

Purpose: Recent studies have highlighted the importance of inherited predisposition to acute lymphoblastic leukemia (ALL), with identification of loss-of-function germline variants in *TP53*, *PAX5*, *IKZF1*, and *ETV6* in familial and sporadic ALL. We combined genome-wide linkage and analysis of germline SNV/indels and structural variants (SVs) of whole genome sequencing data to identify the genetic basis of familial pre-B ALL and lymphoma in a family with 6 affected individuals. **Methods:** Multipoint parametric linkage analysis was performed using Merlin, assuming an autosomal dominant low-penetrance model. SV detection was performed with CREST, based on de novo assembly of soft-clipped reads. WGS and RNA-sequencing were performed on tumor of one patient with ALL and one with lymphoma. The computationally discovered germline deletion was validated by RT-PCR of the patients' RNA. **Results:** No putative causal nucleotide or indel variant was identified. Linkage analysis identified a single 19cM peak on chromosome 12 with LOD score of 1.8. SV detection identified a 75 bp deletion at the intron 6-exon 7 junction of *ETV6*. This was supported by soft-clipped sequence reads and was experimentally confirmed to cause skipping of exon 7 of *ETV6* and protein truncation. We screened 4500 ALL patients for similar deletions, but no additional cases were found suggesting the deletion is rare. The ALL tumor exhibited hyperdiploidy, P2RY8-CRLF2 fusion and *CDH6*, *KRAS*, and *BTBD1* mutations, consistent with prior studies finding enrichment of germline *ETV6* mutations in hyperdiploid ALL. **Conclusions:** We identified a germline deletion which causes exon 7 skipping and protein truncation of *ETV6*. This is the first report of a potentially pathogenic germline SV in *ETV6*, and extends the known importance of *ETV6* germline variants in predisposing to ALL. A mouse model has been created and functional analyses are ongoing. Deletions of this size are usually missed by other coverage depth or discordant read-based SV detection algorithms, and retaining soft-clipped reads in the mapping is critical for this type of discovery. These data highlight the importance of WGS for identification of germline variants predisposing to ALL, particularly in kindred lacking putative causal variants from exome sequencing.

857F

Personalized drug combination prediction in triple-negative breast cancer using single-cell sequencing. Z. Safikhani^{1,2}, B. Haibe-Kains^{1,2}, M. Park^{2,4}, H. Kuasne^{3,4}, C. Martinez Ramirez^{3,4}. 1) Princess Margaret Cancer Research Centre, Toronto, Ontario, Canada; 2) University of Toronto, Department of Medical Biophysics; 3) Rosalind and Morris Goodman Cancer Research Centre; 4) Departments of Biochemistry, Medicine and Oncology McGill University.

Problem: Triple-negative breast cancer (TNBC), representing 20% of breast cancer, is an aggressive subtype with no targeted therapies available. The cancer cells that compose TNBC tumours tend to display highly variable features, which likely underlies the resistance to monotherapies. Hence, a shift away from the single-drug paradigm for new therapeutic strategies is becoming increasingly necessary. We address this issue by finding efficient drug combinations targeting multiple clones of tumour simultaneously by analysing single-cell RNA-seq data of the TNBC samples from patient derived xenografts (PDX) model. **Models:** We develop new predictive models of drug response trained with large amount of preclinical pharmacogenomic data from high-throughput screening of immortalized cancer cell lines against single-agent therapies. These models are built upon the gene expression features with minimum redundancy and maximum relevance to the drug sensitivity. We use variational autoencoder to cluster the high dimensional scRNA-seq and to impute the unexpected dropouts due to technical effects in this data. Model learnt from training phase is applied on single cells to predict the most effective monotherapy for each cluster. An optimization approach is used to find drug combinations targeting different clusters with maximum synergistic effects and minimum toxicity. To reduce the complexity of the cost function it is limited to the combinations of only two drugs. **Results:** We trained our prediction model on a pharmacogenomic dataset of 90 breast human immortal cell lines which have been treated with 80 chemical compounds. The model has been applied on the scRNA-seq data of a collection of four thousand cells. One the most promising combinations candidates reported by our pipeline is Sorafenib + Bortezomib which we aim to validate it in vivo to treat PDX models. **Discussion:** With successful validation, our novel therapeutic strategies and companion tests could reduce overtreatment and unnecessary toxicity by selecting the best drug combinations for each individual TNBC patient. .

858W

Identification of genetic variants associated with risk of familial endometrial cancer. N.J. Shamsani¹, T.A. O'Mara¹, S.H. Kazakoff¹, K. Nones¹, F. Newell¹, A.M. Reed², J. Harraway², J. Pearson¹, A-M. Patch¹, N. Waddell¹, A.B. Spurdle¹. 1) QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia; 2) UQ Centre for Clinical Research; 3) Sullivan Nicolaides Pathology/Mater Pathology.

Endometrial cancer, a neoplasm of the uterine epithelial lining, is the most commonly diagnosed gynecologic malignancy in developed countries. Approximately 3-5% of endometrial cancer at the population level is caused by inherited pathogenic variants in one of the mismatch repair (MMR) genes, *MLH1*, *MSH2*, *PMS2* or *MSH6*, and this proportion is increased in cases reporting a family history of endometrial or other cancers. However, many suspected familial endometrial cancer patients do not exhibit tumour MMR deficiency. Inheritance of rare high-penetrance germline variants in other genes is a possible explanation for the underlying genetic cause of some of these familial cases. Whole exome or whole genome sequencing of 35 familial endometrial cancer cases with an MMR proficient tumor profile was performed to explore the possible genetic cause of disease in these patients. A sophisticated pipeline was developed to implement international-grade variant classification guidelines. This included development and validation of MES-SWA and MES-NCSS, two plugins for The Ensembl Variant Effect Predictor (VEP) that use MaxEntScan scores to assist the prediction of variants that cause aberrant splicing through splice site loss or gain. A combination of frequency and bioinformatic filters were used to identify potentially pathogenic variants in 32 known or suspected endometrial cancer predisposition genes. Candidate variants were validated by Sanger sequencing. Three pathogenic/likely pathogenic variants (*ATM* c.7638_7646del, *MUTYH* c.494A>G, *PALB2* c.3116del) and two variants of uncertain significance (*BRIP1* c.1207C>T, *NBN* c.511A>G) were identified. One potentially spliceogenic variant in *RINT1* c.2067+6T>A was classified as likely benign based on other evidence in the public domain. We have undertaken whole genome sequencing of tumor DNA from the germline *MUTYH* variant carrier, to support *MUTYH* as a probable endometrial cancer predisposing gene. The sequencing of the tumor sample revealed a high proportion of C>A somatic mutations in a pattern which most resembles the COSMIC mutation signature 18, a signature similar to signature 36 that has been associated with *MUTYH* mutations in colorectal cancer and pancreatic neuroendocrine tumors. In conclusion, we have highlighted three known cancer genes (*ATM*, *MUTYH*, and *PALB2*) as relevant for further research studies of familial endometrial cancer.

859T

A minigene platform to validate novel immunogenic peptides arising from somatic mutations as therapeutic cancer vaccines. X. Shi, V. Kode, K. Lee, A. Chaudhuri, P. Chakroborty. MedGenome Inc, Foster City, CA.

Cancer vaccines have shown promise in treating late stage cancers (with failed traditional therapies). Somatic mutations provide a rich source of potential cancer vaccines with minimal T cell tolerance. We have built a robust screening assay to test somatic mutations for CD8⁺ T cell activation in a T cell – dendritic cell co-culture assay with synthetic peptides added from outside. We have screened more than 125 peptides by this assay and identified cancer vaccines candidates from a large number of somatic mutations. However, our screening assay also revealed that non-mutated wild-type peptides in many instances induced T cell activation, although their HLA binding affinity was weak. We suspect that addition of high concentration of the peptides from outside may overcome their weak HLA-binding property allowing them to bind HLA and activate T cells. To reduce the bias of artificial HLA presentation, we have built a minigene platform to screen pairs of wild-type and mutant peptides to assess their immunogenicity. Predicted mutant immunogenic peptides and their wild-type counterparts were cloned as minigenes and expressed in HLA-restricted antigen presenting cells (APCs). The minigene expressing APCs were co-cultured with allogenic CD8⁺ T cells and activation was monitored by the combined expression of IFN-gamma, TNF-alpha and granzyme-A (CTL phenotype of T cells) by flow cytometry. Our screening assay reveals that a significant number of indels are not presented when they are expressed as a longer peptide may be because of the reduced stability of the transcripts. However, their presentation and immunogenicity are preserved when they are expressed as short 9-mer peptides with engineered proteasomal cleavage sites. Our study also identified a range of HLA-binding affinities that correlate with presentation of the peptide on the surface of antigen presenting cells. In conclusion, our analysis demonstrates that the two approaches for investigating immunogenicity of peptides – minigene approach and external addition of peptide approach - have differential utilities for testing and validating the immunogenicity of indels and SNVs. .

860F

Allelic imbalance profiling of premalignant lesions of the lung. S. Sivakumar^{1,2}, F.A. San Lucas¹, T.L. McDowell³, Y. Jakubek¹, W. Lang³, L. Xu³, J. Fujimoto³, J. Fukuoka⁴, Y. Yatabe⁵, J. Fowler⁶, G. Davies⁶, E. Ehli⁶, I.I. Wistuba³, H. Kadara^{1,7}, P. Scheet^{1,2,3}. 1) Epidemiology, MD Anderson Cancer Center, Houston, TX; 2) MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences, Houston, TX; 3) Translational Molecular Pathology, MD Anderson Cancer Center, Houston, TX; 4) Graduate School of Biomedical Sciences, Nagasaki University, Nagasaki, Japan; 5) Department of Pathology and Molecular Diagnostics, Aichi Cancer Center, Nagoya, Japan; 6) Avera Institute for Human Genetics, Sioux Falls, SD; 7) Department of Biochemistry and Molecular Genetics, The American University of Beirut, Beirut, Lebanon.

Lung adenocarcinoma (LUAD) is the largest histological subtype of lung cancer and continues to increase in prevalence in both smokers and non-smokers. Early detection and better understanding of its multi-step progression will be crucial for developing effective prevention measures to thereby reduce the mortality from lung cancer. This can be addressed by investigating atypical adenomatous hyperplasia (AAH), the only known precursor lesion to LUAD. Yet, premalignant genomic studies are challenging, largely due to difficulties in acquiring and assessing aberrations in samples of low cellular fraction samples, features typical of these lesions. Here, we explore chromosomal instability, by virtue of acquired megabase-scale chromosomal aberrations leading to allelic imbalance (AI), in matched AAH, LUAD and normal lung parenchyma samples from 16 stage-I LUAD patients (n=48). We applied hapLOH, a sensitive haplotype-based AI detector, to identify large copy number alterations (e.g., gain, loss) as well as copy-neutral loss of heterozygosity across the genome. Overall, AAHs exhibited significantly lower AI burden than LUADs, with lifetime smokers carrying a higher AI burden than non-smokers. Loss of 17p was the most recurrent event in AAHs of our cohort (6/16) followed by other recurrent events including 1q, 8p, 13q, 16q, 18q, 19p and 19q in three or more cases. Phylogenetic analysis revealed eight cases that showed evidence for shared AI events between matched AAH and LUADs, of which, four cases showed similar overall AI burdens between the tissues. These shared AI events included loss of 17p, 13q, 19p, 19q and 9q, thus providing evidence for their role as truncal events in the pathogenesis of AAH and subsequent transformation to LUAD. Further, by integrating our findings with somatic single nucleotide variants cataloged in these samples, we identified three cases that exhibited a first shared (AAH and LUAD) AI hit, accompanied by a LUAD-specific mutation hit in known LUAD drivers such as KRAS, TP53, NOTCH1, STK11 and KEAP1. Our novel results suggest an early role for large chromosomal aberrations (e.g., 17p loss) as well as acquired secondary hits (e.g., smaller mutations) in the pathogenesis of these premalignant lesions. Further, the similar burden and presence of large truncal events in matched AAH and LUADs might be suggestive of a field of cancerization effect aiding in the clonal selection or expansion of these premalignant lesions.

861W

Leveraging unique molecular indices to detect chimeric gene fusions.

B. Vilhjalmsón¹, F. Lescai¹, T. Rask¹, Z. Etches¹, V. Gregersen¹, R. Madsen¹, S. Tian², D. Kamaev¹, M. Johansen¹, L. Husted¹, J. Caffrey², A.-M. Hein¹, E. Lader², L. Schausser¹, M. Simonsen¹. 1) QIAGEN Aarhus, Aarhus, Denmark; 2) QIAGEN Sciences Inc., Frederick, Md, USA.

Fusion genes are driver mutations in several diseases, particularly cancer. Because they produce chimeric proteins, fusion genes are useful biomarkers for diagnosis, prognosis and treatment of in numerous types of neoplasia. Detection of certain fusion genes in a cancer patient therefore becomes an important task in bioinformatics analyses. For this reason, there are many fusion gene detection methods available. However, most of them are not compatible with reads tagged with Unique Molecular Indices (UMIs), and thus unable to make use of their advantages for estimating expression levels and detecting variants (Xu *et al.*, BMC Genomics 2017). To address this, we developed a novel fusion detection method that leverages UMIs to identify gene fusions. We developed a workflow for analysing datasets obtained with the QIAseq Targeted RNAseq panels. The method first identifies fusion candidates by re-mapping unaligned ends of reads that map partially to one gene to all gene models of the genome, taking broken paired-end read information into account. In addition, the method considers exon boundaries, fusion redundancy, and other features when aggregating the evidence for the candidate fusion(s). It then re-maps all of the reads to a transcript reference database that includes the identified candidate fusions using the QIAGEN CLC RNA-seq aligner. The method along with user-friendly analysis workflow is implemented in the QIAseq Targeted Panel plugin v. 1.1 or higher, which is available in the CLC Biomedical Genomics Workbench. We benchmarked the workflow by applying it to 7 test datasets, and demonstrate that the method is able to detect gene fusions with high accuracy.

862T

Analysis of next-generation sequencing data on African American hereditary prostate cancer cases identifies CNVs in multiple genes. K. Wood Termini¹, T. Mamidi¹, M. Baddoo², M. DeRycke³, S. Riska³, S. McDonnell³, D. Schaid³, S. Thibodeau⁴, J. Carpten⁴, C. Cropp^{5,6}, D. Mandal¹. 1) Department of Genetics, LSU Health Sciences Center, New Orleans, LA; 2) Tulane Cancer Center, New Orleans, LA; 3) Mayo Clinic, Rochester, MN; 4) Keck School of Medicine of the University of Southern California, Los Angeles, CA; 5) Integrated Cancer Genomics Division, Translational Genomics Research Institute, Phoenix, AZ; 6) Department of Pharmaceutical, Social and Administrative Sciences, McWhorter School of Pharmacy, Samford University, Birmingham, AL.

Prostate cancer (PCa) is the second leading cause of cancer-specific deaths among American men, and there is a significant racial disparity between African American (AA) and Caucasian (CAU) affected men, with AA men having a 60% higher incidence rate than their CAU counterparts. Due to limited genomic data on AA men with PCa, no ancestry specific genetic variants have been detected to date. Copy number variations (CNVs) have been documented in PCa tumors, but changes in CNV have not been studied in germline DNA from hereditary PCa (HPC) cases, which accounts for approximately 15% of PCa cases. The African American Hereditary Prostate Cancer (AAHPC) Study, established in 1997, enrolled 77 AA families from seven clinical sites across the United States (Ann Epidemiol, 2000). The goal of the present study is to identify germline CNVs in whole exome sequencing (WES) data on cases from the AAHPC cohort (early onset and ≥4 PCa affected/family). We hypothesize that rare germline CNVs in actionable genes contribute to higher incidence of PCa in HPC families with African ancestry. We have analyzed WES data on 14 AAHPC cases from 11 families included in the International Consortium for Prostate Cancer Genetics sequencing study. The CNV calling algorithms CANOES (Nucleic Acids Research, 2014) and XHMM (AJHG, 2012) were used to detect germline CNVs by modeling read counts using a precise distribution and the Hidden Markov Model, and CNVs were visualized using Integrative Genomics Viewer (IGV). CANOES and XHMM called 313 germline CNVs, 127 of which were IGV-confirmed; 33 CNVs were located within 25 cancer-related genes and 12 were specifically within six PCa-related genes. CANOES called germline CNVs in eight cancer-related genes, including two PCa-related genes *DEC1* and *MTUS1*. XHMM called germline CNVs in 12 cancer-related genes, including four PCa-related genes *CHEK2*, *AGR3*, *UGT2B17* and *FAM115C*. Both programs called CNVs in the five cancer-related genes *BTNL8*, *DMBT1*, *RHD*, *CTSB* and *GSTT1*. WES analysis on additional affected family members and other AAHPC families is ongoing. In conclusion, we identified germline CNVs from AAHPC WES data using CNV calling algorithms. The results from this study are consistent with somatic findings, suggesting that the germlines of AAHPC and other AA patients may hold valuable information necessary for discovering clinical biomarkers for PCa, which will allow us to facilitate early diagnosis and disease prevention in AAHPC families.

863F

Fast deconvolution tool for separating subtype specific signals from mixed tumor genomic data. *L. Yang, H. Zhu, JS. Marron.* University of North Carolina at Chapel Hill, Chapel Hill, NC.

With the advance of deep sequencing techniques, the heterogeneity within a tumor tissue, referred as intra-tumor heterogeneity, becomes a prevalent confounding factor to tumor genomic studies. Analysis on genomic profilings from heterogeneous tumor samples can potentially lead to false positive differential expression conclusions, and even influence patients' clinical outcomes and therapeutic responses. To address the intra-tumor heterogeneity issue, we develop a Fast Tumor Deconvolution (FasTD) tool to separate the pure tumor signals from tumor-nontumor mixtures in an efficient way. Assuming a linear combination of the abundance of the mixing components and availability of some reference information for the non-tumor part, our semi-parametric regression-based model can quickly provide estimates for the tumor proportion in a mixture, as well as output the tumor specific genomic profile. We demonstrate FasTD is a competitive tumor deconvolution tool for both simulated data and The Cancer Genome Atlas RNA-seq datasets, with no requirement for pre-selected signature genes.

864W

Comprehensive transcriptomic analysis of cell lines as models of primary tumor samples across 22 tumor types. *K. Yu^{1,2}, B. Chen^{1,2}, D. Aran¹, T. Goldstein¹, M. Sirota^{1,2}.* 1) Institute for Computational Health Sciences, University of California, San Francisco, CA; 2) Department of Pediatrics, University of California, San Francisco, CA; 3) Department of Pediatrics and Human Development, Department of Pharmacology and Toxicology, Michigan State University, Grand Rapids, MI.

Cancer cell lines are commonly used as models for cancer biology. While they are limited in their ability to capture complex interactions between tumors and their surrounding environment, they are a cornerstone of cancer research and many important findings have been discovered utilizing cell line models. Not all cell lines are appropriate models of primary tumors, however, which may contribute to the difficulty in translating *in vitro* findings to patients. Previous studies have leveraged public datasets to evaluate cell lines as models of primary tumors, but they have been limited in scope to specific tumor types. We present here a comprehensive pan-cancer analysis utilizing nearly 10,000 transcriptomic profiles from The Cancer Genome Atlas and the Cell Line Encyclopedia to evaluate cell lines as models of primary tumors across 22 different tumor types. The correlations between cell lines and primary tumor samples varied widely across the 22 tumor types and we identify tumor types that may benefit from the generation of additional cell line models. Additionally, we found that cell-cycle pathways are consistently upregulated in cell lines, while immune related pathways are consistently upregulated in primary tumor samples across all the tumor types. In a case study, we compared colorectal cancer cell lines with primary tumor samples across the colorectal subtypes and identified two colorectal cell lines that were derived from a different cell type than the primary tumor samples. Lastly, we propose a new set of cell lines panel, the TCGA-110, which contains the most representative cell lines from 22 different tumor types as a more comprehensive and informative alternative to the NCI-60 panel. Our analysis of the other tumor types are available in our web app as a valuable resource to the cancer research community, and we hope it will allow researchers to select more appropriate cell line models and increase the translatability of *in vitro* findings.

865T

Pan-cancer biomarker identification: Survival distinctions based on transcriptome-wide RNAseq expression. S. Zaman¹, B.I. Chobrutskiy¹, D.T. Segarra¹, G. Blanck^{1,2}. 1) University of South Florida Morsani College of Medicine, Tampa, FL; 2) Immunology Program, H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida.

Biomarkers are disease characteristics that represent an outcome associated with a specific disease. Currently, there are few cancer biomarkers that are used in clinical practice. As such, we aim to identify RNA expression levels as biomarkers for disease-specific cancer outcomes, i.e., overall and disease-free patient survival, using available RNAseq and survival data from The Cancer Genome Atlas (TCGA) databases for 28 different cancers (n = 11,049). We used a bioinformatics approach to assess the association of the expression of every gene in the tumor transcriptomes with survival rates. The original algorithm developed for this study first organized the transcriptomes based on differential RNAseq expression data and then matched significant differences in the RNAseq data to survival data for Kaplan-Meier survival analyses. Results indicated that the expression of several genes associated with greater than 15 years survival difference, when comparing the top and bottom quintile of expressers. As an example, we detected an association of greater than 15 years of survival difference with different expression levels of the following genes, for Lower Grade Glioma: *NFE2L3* (p < 0.0001), *ZDHHC22* (p < 0.0001), and *EIF3L* (p < 0.0001). Moreover, our results indicate several genes with greater than 15 years difference in recurrence rate when comparing the top quintile of expressers of a specific gene and the bottom quintile of expressers for that same gene for several cancers. Again using Lower Grade Glioma as an example, for disease-free survival: *TGIF1* (p < 0.0001), *IGF2BP3* (p < 0.0001), and *CRY2* (p < 0.0001). Together, these results indicate that differential gene expression patterns have drastic effects on patient survival and cancer recurrence. Moreover, these results provide a resource for identifying candidate genes whose expression provides profound survival distinctions that may be worth further investigation as biomarkers for improved patient outcomes.

866F

Comprehensive characterization of alternative polyadenylation in human cancer. Z. Zhang¹, Y. Xiang¹, G.B. Mills², L. Han¹. 1) Department of Biochemistry and Molecular Biology, McGovern Medical School at The University of Texas Health Science Center at Houston, Houston, TX; 2) Department of Systems Biology, The University of Texas MD Anderson Cancer Center, Houston, TX.

Alternative polyadenylation (APA) is emerging as a major post-transcriptional mechanism for gene regulation, and dysregulation of APA contributes to several human diseases. However, the functional consequences of APA in human cancer are not fully understood. Particularly, there is no large-scale analysis in cancer cell lines. Our study characterized the global APA profiles of 6398 patient samples across 17 cancer types from The Cancer Genome Atlas and 739 cancer cell lines from the Cancer Cell Line Encyclopedia. We built a linear regression model to explore the correlation between APA factors and APA events across different cancer types. We used Spearman correlation to assess the effects of APA events on drug sensitivity and the Wilcoxon rank-sum test or Cox proportional hazards model to identify clinically relevant APA events. We revealed a striking global 3'UTR shortening in cancer cell lines compared with tumor samples. Our analysis further suggested PABPN1 as the master regulator in regulating APA profile across different cancer types. Furthermore, we showed that APA events could affect drug sensitivity, especially of drugs targeting chromatin modifiers. Finally, we identified 1971 clinically relevant APA events, as well as alterations of APA in clinically actionable genes, suggesting that analysis of the complexity of APA profiles could have clinical utility. Our study highlights important roles for APA in human cancer, including reshaping cellular pathways and regulating specific gene expression, exemplifying the complex interplay between APA and other biological processes and yielding new insights into the action mechanism of cancer drugs. .

867W

Model-based analysis of positive selection significantly expands the list of cancer driver genes. S. Zhao, M. Stephens, X. He. University of Chicago, Chicago, IL.

Identifying driver genes is a central problem in cancer biology and has received great attentions from researchers. However, existing methods for detecting driver genes from somatic mutation data struggle to distinguish positive selection signals from highly heterogeneous background mutational processes. Here, we present a powerful statistical approach, driverMAPS (Model-based Analysis of Positive Selection) for driver gene identification. The key feature of driverMAPS is its modeling of mutation rates at the base-level, reflecting both background mutational processes and positive selection. Its selection model captures elevated mutation rates in functionally important sites using multiple external annotations, as well as spatial clustering of mutations. Its background mutation model accounts for both known covariates and local, gene-specific, variation caused by unknown factors. Applying driverMAPS to TCGA data across 20 tumor types identified 159 new potential driver genes. Cross-referencing this list with data from external sources strongly supports these findings.

868T

Deep learning for extrachromosomal DNA detection in metaphase images. U.C. Rajkumar¹, K. Turner², J. Ko², W. Zhang², M. Chandraker¹, P. Mischel², V. Bafna¹. 1) Department of Computer Science and Engineering, University of California at San Diego; 2) Ludwig Institute for Cancer Research, University of California at San Diego.

Recent studies have shown that extrachromosomal DNA (ecDNA) - circular DNA molecules - occur in ~40% of all human cancers but rarely in normal samples (Turner, Nature 2017). The ecDNA replicate and segregate independently and explain many complex rearrangements of focal amplifications. Moreover, they are enriched in oncogenes, and show tremendous cell-to-cell heterogeneity, suggesting a proliferative and evolutionary mechanism for the cancer cell via modulation of ecDNA counts, as well as a possible marker for cancer progression. As sequence based methods cannot easily distinguish between extra- and intra-chromosomal events, automated cytogenetic analysis of ecDNA containing cells presents an important challenge. We developed a method, ecDetect2, for the automated segmentation of DAPI stained cell metaphase images into nuclei, chromosomes, and ecDNA elements. ecDetect2 is based on U-nets, a variation of fully convolutional networks for semantic segmentation. Specifically, it contains 4 modules: the first 3 use U-nets for binary segmentation of nuclei, chromosomes, and ecDNA, respectively. The final module uses a supervised learning approach to combine the outputs of the deep learning modules to produce a multi-class segmentation and statistically quantify the results in cytogenetic terms. Post segmentation, ecDetect2 computes statistics on cell to cell ecDNA counts. To train ecDetect2, we started with 483 manually annotated metaphase images (1392x1040 pixels) and cropped them into 256x256 sub-images. This final training and cross-validation data-set consisted of 5949 images with chromosomes, 6063 with nuclei, and 587 with ecDNA. To the best of our knowledge, this dataset, which will be continually augmented, is the first of its kind containing ecDNA elements. ecDetect2 was tested on a separate dataset containing 180 images of dimensions 256x256. It achieved an Intersection over Union (IoU) score of 70.1% for ecDNA. In cytogenetic terms, it predicted 489 ecDNA, containing 462 (94% precision) manually marked ecDNAs, and missing only 14 (96% recall). An earlier version, ecDetect (Turner, 2017), based on classical computer vision methods, achieved 36% recall and 97% precision. A majority of the false positives observed for ecDetect2 were due to spurs from intact nuclei that one could mitigate by fine-tuning the nuclei prediction on additional data.

869F

Automated classification of *BRCA1* and *BRCA2* variants shows near perfect concordance with ENIGMA expert panel assessments. J.L. Poitras, D. Richards, H. Su, T. Love, R. Yip. QIAGEN Bioinformatics, Redwood City, CA.

As the compendium of putatively disease causing variants expands, gathering the most current and accurate information is critical to computing variant classifications. Manual curation is laborious, time consuming, and error prone, as information critical to variant interpretation may be missed. The QIAGEN knowledgebase includes the most extensively curated database of variant specific publications, as well as data from third party sources such as ClinVar, HGMD, CentoMD, and OMIM. This knowledgebase functions as the cornerstone of our clinical decision support tool, QIAGEN Clinical Insight (QCI), facilitating rapid variant filtering and prioritization, automated ACMG classification, and reporting. Here, we compare the concordance of QCI's variant classification with the ENIGMA expert panel assessments of *BRCA1* and *BRCA2* variants. As of December 2017, ClinVar contained 6154 expert reviewed *BRCA1* and *BRCA2* variants. These variants were exported, and the resulting VCF was uploaded to QCI. QIAGEN curated content within the tool was used to automatically compute ACMG variant classifications and provide underlying evidence. QCI computed classifications were compared with ENIGMA assessments, and demonstrated an extremely high rate of concordance, with 81.2% of classifications perfectly matching ENIGMA's assessments. Moreover, the majority of differences observed would not affect clinical management, occurring either within Pathogenic/Likely Pathogenic or Benign/Likely Benign (11.1%) or Benign/Likely Benign vs VUS (7.2%), bringing the degree of concordance at the level of clinical actionability to 99.6%. Only 3 variants were significantly discordant (Benign/Likely Benign by ENIGMA vs. Likely Pathogenic by QCI). These included synonymous variants with functional data supporting splicing defects not considered by ENIGMA. In summary, with respect to clinical actionability, QCI automated ACMG classification was 99.6% concordant with ENIGMA expert panel variant assessments. This level of accuracy speaks to the quality of the clinical, functional, and population level data in the knowledgebase, as well as the robustness of the underlying algorithm used to apply the ACMG guidelines.

870W

Integrating DeepVariant into a clinical bioinformatics pipeline. L. Hon¹, J. Van den Akker¹, Z. Mahkovec¹, G. Mishne¹, A. Ondov¹, L. Dorfman², P.-C. Chang², R. Poplin², M. DePristo², J. Ginsberg¹. 1) Color Genomics, Inc., Burlingame, CA; 2) Google, Inc., Mountain View, CA.

BACKGROUND: Bioinformatics pipelines are an integral component in clinical genetic testing based on next-generation sequencing. To maximize sensitivity, add redundancy, and increase system robustness in a clinical setting, ensemble approaches to variant calling can be utilized. Here, we describe a workflow to evaluate the added value of additional callers of Single Nucleotide Variants (SNVs) and insertion/deletions (indels, <50 bp), which was applied to assess DeepVariant. **METHODOLOGY:** To bolster the performance for difficult-to-call variants and regions, Color has previously incorporated multiple variant callers. Given the strong performance of Google's variant caller DeepVariant [1], we wanted to assess the utility of DeepVariant (v0.6) to strengthen variant detection capabilities. We first evaluated sensitivity and specificity using several well-characterized NIST cell lines. However, NIST datasets for comparison are typically limited to high confidence genomic regions where multiple workflows converged to a single consensus call [2]. The core of our evaluation therefore focused on the ability to detect variants that are considered difficult to call. These variants were selected from publicly available cell lines, synthetic DNA, and clinical specimens for which our current set of callers had trouble identifying the correct variant. Thus, this set was strongly enriched for clinically relevant variants in regions of low/high GC content, variants in the vicinity of homopolymers (e.g. *MSH2*, c.942+3A>T), and long/complex indels (e.g. *ATM*, c.7849_7857delGTCAGAAGTinsAGGTCTCCTTTGAGTGCCTCA). **RESULTS:** DeepVariant performed well in a number of areas such as GC-rich regions, including calling the pathogenic variant *MEN1* c.206_207insGCCCC in Coriell cell line NA11630. We detected a small number of novel variant calls over our existing pipeline, which suggests a limited impact on our downstream workload. Based on this positive assessment, DeepVariant has been added to our clinical bioinformatics pipeline. Additional analyses are currently underway to more broadly assess the performance of DeepVariant. **CONCLUSION:** To ensure high sensitivity, it is critical to include callers optimized for different types of variants. Here we show that DeepVariant can be used to provide redundancy in variant detection and potentially increase sensitivity, especially in regions of low/high GC content. [1] Poplin et al. Biorxiv 2018 [2] Krusche et al. Biorxiv 2018.

871T

A study of the new sentieon and edico genomics sequence data processing methods. *J.M. Bell¹, S.M. Grimes¹, H.P. Ji^{1,2}.* 1) Stanford Genome Technology Center, Palo Alto, CA; 2) Division of Oncology, Department of Medicine, Stanford University School of Medicine, Stanford, CA,.

Since the beginning of high-throughput genomics, there has been an exponential increase in the amount of data produced by sequencing. One of the largest challenges has been to produce bioinformatics tools that can keep pace with this increase while performing accurate analysis. We studied the speed and accuracy of two new tools, the Sentieon suite available from Sentieon, and the DRAGEN system available from Edico, compared to the current standard, GATK and MuTect2. The Sentieon program instantiates the GATK algorithms, while the DRAGEN device uses a field-programmable gate array (FPGA) chip to accelerate processing. In order to test these analysis methods, we compared a variety of metrics from an established control genome (NA12878) as well as a series of archival clinical samples that were preserved with formalin-fixation and paraffin-embedding (FFPE), investigating both the relative speed of the analysis methods and the concordance of results. Because DRAGEN also has a copy number variant (CNV) analysis module, we used the standard program CNVkit to generate comparison CNV calls. Sentieon uses a version of GATK recoded in C; it is generally about 20x faster than GATK itself. We have found that germline fresh-frozen analysis of NA12878 is several times faster with DRAGEN than with Sentieon. The rate of analysis can be higher with larger inputs, perhaps owing to lower overhead. Conversely, there is a non-linear loss of speed as FFPE data inputs grow larger. As sequenced FFPE samples produce noisier data than do fresh-frozen samples, we hypothesize that the DRAGEN system is disproportionately affected by the increase in noise as depth increases and is thus slowed. The DRAGEN system includes an integrated variant caller. Results have a high concordance with MuTect2, and certain edge cases filtered by DRAGEN but not by MuTect2 are likely to be false positive calls in MuTect2. In analyzing FFPE data, however, DRAGEN allows a large number of calls that are characteristic of deamination changes (C/G->T/A) but that are filtered out by MuTect2. Edico's CNV caller shows significant overlap with the CNVkit caller but also a number of significant differences that could affect assessment of a tumor sample. In summary, both Sentieon and Edico's analysis methods show significant speed increases over GATK, while the relative quality of analysis varies depending on the characteristics of the input.

872F

High-state-count negative binomial hidden Markov model improves detection of copy number variation in cell-free DNA for cancer detection and response prediction. *E.S. Gafni, A.D. Adam M. Drake, K.N. Niehaus, Y.L. Liu, I.H. Haque, J.J. John.* Freenome, San Francisco, CA.

Many cancers harbor large scale copy number variants (CNVs) which can be detected by changes in genome-wide read coverage in sequencing data. Additionally, some tumors release detectable amounts of DNA into the blood as circulating cell free DNA (cfDNA), in which case sequencing of total cfDNA provides information on a mixture of tumor and normal genome composition. However, plasma tumor fraction (the fraction of cfDNA derived from the tumor) is a critical parameter for the ability to detect these CNVs, as apparent changes in read coverage become smaller and more difficult to distinguish from noise as tumor fraction declines. Current cfDNA CNV detection methods [1] have limited sensitivity to these events particularly when tumor fraction is low (<8%). Improvements in the ability to detect CNVs at low tumor fraction may be valuable to assist in the detection of cancer or to monitor its response to treatment. Here we propose a method which uses a negative binomial regression against a panel of normal samples as well as a Hidden Markov Model (HMM) with an optimized implementation of the Viterbi algorithm with a highly discretized state space (>500 states). The negative binomial regression against a panel of normals allows us to correct for technical batch effects assuming consistency in sample processing. The large HMM state space allows us to avoid the complexities of jointly estimating tumor fraction and CNV tumor clonality, where there may not be a single unique solution. We utilize expectation maximization in order to jointly correct for batch effects as well as infer the negative binomial dispersion parameter of each region of the genome. We validate the performance of the method in both simulation and real-world plasma cfDNA data, demonstrating robustness to CNV region size, CNV amplification/deletion level, tumor fraction, and technical artifacts such as GC bias. At genome-wide coverage of 10-15x, we are able to achieve sensitivity to mean shifts of less than 1%, allowing superior detection of CNVs at low tumor fraction and for small relative copy number change, which have utility for the early detection of cancer and response prediction for immune checkpoint therapy, respectively. [1] Adalsteinsson VA et al., Nature Communications 2017

873W

DEFOR: Depth and frequency based somatic copy number alternation detector. *H. Zhang¹, X. Zhan¹, J. Brugarolas², Y. Xie^{1,2}.* 1) Department of Clinical Sciences, University of Texas Southwestern Medical Center, Dallas, Texas 75390; 2) 2. Department of Internal Medicine and Kidney Cancer Program, Simmons Comprehensive Cancer Center, University of Texas Southwestern Medical Center, Dallas, Texas 75390.

Somatic copy number alteration (SCNA) is the change in copy number that has arisen in somatic cells. SCNAs were observed frequently in tumors and recurrent SCNAs were also identified in several types of tumors. SCNA profiling plays an important role to study the mechanism of tumor and guide the therapeutics. In recent years, SCNA detection using high-throughput sequencing is becoming popular that was because of quickly sequencing technology development. Although several bioinformatics methods have been developed for SCNA detection using exome-sequencing data, these methods rely a strong assumption that the median or mean value of the copy numbers of the genome is two (normal status). That assumption was usually used to give an estimation of the depth or depth ratio that represents the normal status. That assumption usually holds for germline copy number variations, because most parts of the genome don't have copy number changes. However, that assumption may not hold in tumor cells where large scale copy number alterations were observed frequently. In such situation, the mean or median value of copy numbers across the genome may be not two, and so the copy number estimated based on that assumption may be unreliable any more. Therefore, the existing methods may not perform well to call SCNAs for the unstable tumor genomes. Samples from tumor tissue are usually contaminated with normal cells from nearby tissue, and the purity of different tumor samples varies obviously. Because the purity of tumor sample affects the observed copy number, it adds another layer of variations in SCNA detection of tumor samples. We developed a new method DEFOR to detect SCNAs in tumor samples from exome-sequencing data. DEFOR supports the estimation of copy numbers in six different statuses: 1) normal status, 2) loss of both alleles, 3) loss of one allele, 4) loss of one allele and followed by the amplification of the remained allele, 5) gain of one allele and 6) gain of both alleles. Allele frequency, depth and purity of tumor samples were considered in the model. Based on the evaluation using real dataset, DEFOR has a good accuracy to detect SCNAs from tumor samples and outperformed other methods especially for unstable tumor cells in which SCNAs occurred in a large proportion of the genome.

874T

The transcriptome as a diagnostic aid: a pan cancer method for identifying the site of origin of complex metastases. *J.K. Grewal¹, S. Gakkhar¹, Y. Ma¹, Y.J. Zhao¹, A. Mungall¹, R. Moore¹, H. Lim², D. Renouf², K. Gelmon², S. Yip³, J. Laskin², M. Marra¹, S.J.M. Jones¹.* 1) Michael Smith Genome Sciences Centre, Vancouver, British Columbia, Canada; 2) Department of Medical Oncology, BC Cancer, Vancouver, British Columbia, Canada; 3) Pathology and Laboratory Medicine, University of British Columbia, Vancouver, British Columbia, Canada.

Next Generation Sequencing (NGS) data can provide biological insights into cancer biology, and also potentially be used to minimize time costs associated with patient work-up and management. However, there is a lack of demonstration of their utility in traditionally manual analytical tasks, particularly for cancer diagnostics. Traditional diagnostics in cancer require pathologist input, and cancer diagnosis is important for providing biological context to genomic analysis findings. Several pilot precision oncology projects are scaling up and need to be able to process large amounts of NGS data quickly and efficiently. The concomitant scale-up in the throughput for diagnostic assessments is a challenging manual task. An automated diagnostic method that utilizes solely NGS data can serve as a powerful orthogonal diagnostic aid to pathologists in such scenarios, or when traditional pathology is inconclusive. We investigated the utility of high throughput RNA-sequencing as a fast and automated approach for cancer diagnostics. We trained an ensemble of neural networks to distinguish 40 primary cancer types and 26 adjacent normal tissues. By adapting previously published weight analysis approaches for neural networks, we identified cancer-specific features that were important in classification. The classifier has high precision and recall for primary cancers and subtypes with mixed histology. Cross-calling patterns reflect known biological similarities within certain gastrointestinal cancers and within cancers with shared cell type of origin. Furthermore, we demonstrate the utility of this method in resolving diagnoses for treatment resistant metastatic cancers and cancers of unknown origin. A subset of these re-diagnoses have been validated by follow-up immunohistochemistry. Lastly, within this body of work we propose solutions to common computational challenges arising with using categorically imbalanced biological datasets. Our method uses the whole transcriptome as input, and is not confounded by tumour content. Its immediate application is as an orthogonal diagnostic in cases where the site of origin of a cancer is unknown, or when standard pathology assessment is inconclusive.

875F

Regulatory mechanisms of chronic lymphocytic leukemia (CLL) risk variants in mediating target gene expression. *H. Yan, S. Tian, G. Kleinstern, S. Slager.* Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, MN 55905, USA.

The disease-associated single-nucleotide polymorphisms (SNPs) identified by genome-wide association studies (GWASs) mostly reside in noncoding regions of the genome. It has been established that many of these noncoding variants have regulatory roles in mediating target gene expression, often involving the alterations of local chromatin accessibility, histone modifications, transcription factor (TF) binding affinity, and promoter-enhancer interactions. In CLL, we and others have identified 41 GWAS risk variants for which the mechanism of casualty remains largely unknown. Herein, we evaluated whether these SNPs affect transcriptional program in CLL. We focused on the epigenetic data for marks mapping to the gene regulatory regions using B cells and/or lymphoblastoid cell lines. Data included ChIP-seq data for three histone modifications (H3K4me3, H3K4me1, H3K27ac) and 73 TFs, as well as chromatin accessibility data generated by ATAC-seq, DNase-seq, and FAIRE-seq. We also used histone modification ChIP-seq and DNase-seq data generated by the Roadmap Epigenomics Project in primary human tissues and cells. Comparing the 41 index SNPs and over 4,000 SNPs in linkage disequilibrium ($R^2 \geq 0.5$ in the EUR ethnic group in 1000 Genomes Project) with these index SNP to the above epigenetic data revealed their enrichment in regulatory regions, particularly in the enhancer regions, with strong evidence for the allele imbalance and alteration of TF binding motifs. We identified the potential target genes for the casual SNPs using GTEx eQTL data and public Hi-C, capture Hi-C and ChIA-PET data in lymphoblastoid cell lines and/or primary blood cell types. Many of the target genes, such as *BCL2L11*, *ID3* and *IRF2*, showed expression changes in CLL and are known to be involved in hematologic malignancies including CLL. Together, our analysis highlights the role of epigenetic mechanisms in CLL pathogenesis.

876W

Ranking on cis-eQTL selected transcription factors produces focused and accurate predictions. *B.P. Kellman¹, N.E. Lewis^{2,3}.*

1) Bioinformatics and Systems Biology, University of California, San Diego, San Diego, CA; 2) Department of Pediatrics, University of California, San Diego, San Diego, CA; 3) Department of Bioengineering, University of California, San Diego, San Diego, CA.

The bioinformatic toolbox is large and growing yet actionable transcription factor (TF) predictions remain an open challenge. Tools such as oPossum, PASTAA use frequently occurring sequence motifs to deduce active promoter regions. These tools produce essential predictions but can be overwhelmed by the presence of similar DNA in epigenetically masked regions. The lack of context specificity can result in a large number of false positive predictions and ultimately limiting the utility of the prediction overall. Recent work in TF prediction using cis-eQTL analysis in GTEx to identify tissue-specific active regions of the promoter. DNA motifs in those active promoter regions were queried in CIS-BP to obtain TFs associated with tissue-specific active promoter sequences (Sonawane, 2017). We developed an approach to rank the predicted TFs based on the frequency with which it appears in a background of random gene lists. The probability that a TF is a valid regulator is calculated along a simple cumulative distribution relative to the random gene lists background. We have found this approach to be successful in recovering known regulators of the lysosomal pathway, cholesterol metabolism, and canonical cancer-related pathways. We hope that our ranking these empirically informed TF predictions can sufficiently mitigate the false positive rate of these methods thereby making these tools more useful to the broader bioinformatic and biological community. An implementation of our TF ranking method is available in R (<https://github.com/bkellman/teQTLparse>).

877T

A reference database of transcription factor cascades in the human genome (TFCascades.info): Implications in cancer biology.

S. Piyawitwanich¹, G. Ventyana¹, M. Sudarshan¹, O.K. Mathew², S.K. Belman³, K.W. Johnson⁴, R. Hodos⁴, B.S. Glicksberg⁴, R. Bhattacharya⁵, J. Chelico³, P. Mukherjee⁵, J.T. Dudley⁴, L. Subramanian¹, K. Shameer⁶. 1) Courant Institute of Mathematical Sciences, New York University, New York, NY, USA; 2) B3, JKR, Aluva, Kerala, India; 3) Center for Research Informatics and Innovation, Northwell Health, New Hyde Park, NY, USA; 4) Institute for Next Generation Healthcare, Mount Sinai Health System, New York, NY, USA; 5) University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA; 6) Department of Information Services, Northwell Health, New Hyde Park, NY, USA.

Transcription, a central biochemical control system that optimizes gene expression, is mediated by transcription factors (TFs). TFs perform their gene regulatory function singularly or in a cascade mode. Many transcription factors activate not only a gene but also other transcription factors, creating cascades. TFs are primary drug targets of different anti-cancer therapies. Several TF cascades with two or three transcription factors (for example: IRF8-KLF4) have been reported with functional evidence for implications in human diseases. To date, there is no molecular class-specific reference dataset on TF cascades in humans. We built a reference dataset using a predictive framework that combines human TFs compiled using biocuration coupled with protein-protein interaction (177,016 pairs and 8,627 distinct proteins) and biochemical activity data aggregated from STRING database. We identified a total of 37,879 transcriptional cascades with maximum depth of 16 TFs and compiled the data in a centralized reference resource (www.tfcascades.info). We also explored The Cancer Genome Atlas data on ovarian cancer ($n=316$) and pancreatic cancer ($n=258$) to assess the impact of transcription cascade on various oncophenotypes including tumor grade, treatment response, relapse rate, survival, and metastasis using graph analysis with a connected component algorithm. By analyzing TF cascade-target aberrations and correlations with oncophenotypes as a computable graph, we built a TFCascade-Oncophenome map, which can be used as a predictive framework to find novel regulatory modules, drug targets, and molecular interventions that could improve the outcomes for cancer patients. Pan-cancer analyses of cascades would help to understand novel transcriptional regulatory cascade networks driving human cancers. Assessment of the regulatory network properties of such systems across various types of cancers and human disease would help to build computable models of disease severity.

878F

Contrasting whole genome and whole exome sequencing as competing modalities for obtaining perpetual genomic data. *B.G. Kermani.* Crystal Genetics, Inc., Los Altos, CA.

From its inception, next-generation sequencing (NGS) has brought the promise of rendering perpetual genomic data (PGD) --where the sequencing is done once, and is utilized *ad infinitum*, serving different applications. While germline applications are often viewed as the main use of PGD, somatic applications are valid subjects as well. In cancer, somatic mutations can be traced longitudinally for monitoring patients, e.g., to evaluate their response to drugs and other treatments. By definition, whole genome sequencing (WGS) can produce perpetual data. As WGS has not yet become economical, whole exome sequencing (WES) has been a major contender in producing PGD. In this work, we show that while WES has found successful applications in many areas, it lacks certain properties to make it fully suitable for producing PGD, particularly for health-related applications. We will show that beyond the obvious concerns on WES, i.e., limited targeted regions, and despite higher coverage, it could be subject to higher false positive and false negative rates, as compared to WGS. Although some of these concerns are partially addressed by extending the exomes to include some non-exonic bases, the economy of exome sequencing still forces many of the available solutions to cover only a limited set of bases, as compared to all possible bases of interest, for a multi-application scheme. In this study, a set of anonymized tumor-normal patient samples is used to contrast the quality of the variant calls in WGS and WES modalities, for the same patients. The study suggests that the quality difference between these two modalities is non-trivial, for both germline and somatic applications. Also, it is shown that, in many cases, even popular variant callers are not able to recover from the quality degradations associated with WES. It is also shown that a hybrid *de novo* genome analysis pipeline coupled with an *in silico* verification (ISV) tool can provide significant resilience to these data quality challenges. In summary, we propose that only WGS, in tandem with a high-quality genome analysis pipeline and verification, can provide a sustainable model for addressing multiple applications, in both germline and somatic domains. In a PGD scheme using WGS, however, the analysis pipeline and verification module play major roles. Without proper analysis and verification, the outcomes may not be of sufficient quality to cover a wide range of applications, especially those related to health.

879W

High-resolution mapping of cancer cell networks by co-functional gene detection. *E.A. Boyle¹, W.J. Greenleaf¹, J.K. Pritchard^{1,2,3}*. 1) Genetics Dept, Stanford University, Stanford, CA; 2) Biology Dept, Stanford University, Stanford, CA; 3) Howard Hughes Medical Institute.

Genome-wide CRISPR screens have rapidly accelerated the unbiased and high-throughput investigation of diverse biological systems. Recently, multiple groups have demonstrated mapping of genetic interactions by screening genetic knockouts across several cell line backgrounds in parallel and detecting correlated gene essentiality profiles. Here, we adapt this approach to a broader set of >300 copy-number corrected whole genome CRISPR screens published by the Project Achilles group. We learn a cell line signature of nonspecific sgRNA toxicity and with it eliminate a source of confounding that foils similar correlation-based methods. Our corrected gene profiles confirm known signaling pathways that include cancer drug targets and suggest that the sensitivity of a cell line to a gene knockout is prognostic for cancer drug activity in a surprisingly large number of cases: on average 500 genes per drug. Amongst these profiles, we report than 1 million correlated genes genome-wide that we deem “functional neighbors”. Functional neighbors broadly reflect shared properties of cell networks such as cellular compartment, biological pathway, and protein complex with a granularity that exceeds traditional gene sets. Beyond protein function, we leverage the diversity of the cell line panel to conclusively show that cell type and mutational status illuminate cell signaling paradigms that would be inaccessible with smaller panels of cell lines. Lastly, we employ a gene community detection approach to implicate core genes for cancer growth and summarize functionally redundant signal in whole-genome CRISPR screens. By integrating information from multiple cancer databases, we observe that genes known to be important for cancer frequently associate closely together in the cell network and that topology and function are inextricably linked. This work establishes new algorithms for probing cancer cell networks and motivates the acquisition of further CRISPR screen data across diverse genotypes and cell types to further resolve the complexity of cell signaling processes.

880T

Establishment of reference samples for the detection of somatic variants in cancer. *L.T. Fang¹, W. Xiao²*. 1) Bioinformatics Research & Early Development, Roche Sequencing Solutions, Belmont, CA; 2) FDA/NCTR, Jefferson, AR.

The FDA is well aware of next generation sequencing (NGS)'s crucial role in precision medicine. However, development and evaluation of oncology NGS applications is challenging due to the lack of community-validated cancer reference samples. To address this challenge, over 180+ scientists from 60+ institutions have come together to form the Somatic Mutation Working Group of the SEQC2 consortium. We present here a gold set against which new tools and platforms can be benchmarked. We selected a pair of tumor-normal cell lines as the initial reference samples: a triple-negative breast cancer cell line HCC1395 and its B lymphocyte derived normal HCC1395BL. We have sequenced the genome on multiple short-read and long-read sequencing platforms at multiple institutions, and have established high-confidence call sets of germline variants and somatic mutations. In total, we have performed whole genome sequencing (WGS) for each of the genomes to combined coverage depths of 1700X from 12 replicates on Illumina HiSeq, 700X from 9 replicates on Illumina NovaSeq, 600X from 11 replicates on 10X Genomics, and 50X on PacBio. Germline and somatic SNVs/INDELs were obtained primarily based on the 12 Illumina HiSeq and 9 NovaSeq data sets. We used the SomaticSeq framework with six mutation callers and three aligners on each of the 21 WGS replicates to obtain the first release of somatic mutation gold set. The confidence-level of each somatic mutation call was obtained based on cross-institution and cross-aligner reproducibility of each call. There are about 37,000 somatic SNVs and 1,500 somatic INDELs with variant allele frequencies (VAF) ranging from 2.5% to 100% in the gold set. In addition, we also report likely false positives that have been misclassified by certain algorithm(s) in certain sequencing replicate(s). We used four germline callers to obtain the high-confidence germline variants. High-confidence regions were defined based on the limitation of current technologies. Genomic coordinates in high-confidence regions but not specified in the gold set are assumed to be non-variant. We will continuously accept input from the community to advance accuracy and completeness of our reference samples in order to serve as a community standard well into the future.

881F

Association of coding germline variants with relapse of prostate cancer: A PPCG study. E. Anokian¹, D.S. Brewer^{2,3}, D. Wedge⁴, A. Butler⁵, A. Lynch⁶, C. Massie⁷, S. Merson¹, P. Boutros⁸, M. Fraser⁹, T. Yamaguchi⁸, C.Q. Yao⁸, J. Korbels⁹, J. Weischenfeldt¹⁰, E. Pitkänen⁹, S. Waszak⁹, C. Hovens¹¹, P. Georgeson¹¹, G. Cancel-Tassin¹², O. Cussenot¹², G.S. Bova¹³, T. Schlomm¹⁴, R. Bristow¹⁵, A. Kibel¹⁶, M. Loda¹⁷, C. Sander¹⁷, B. Brors¹⁸, D. Neal^{18,19}, C.S. Cooper^{1,2}, Z. Kote-Jarai¹, R.A. Eeles^{1,20}, *The Pan Prostate Cancer Group*. 1) Institute of Cancer Research, London, United Kingdom; 2) Norwich Medical School, University of East Anglia, Norwich, UK; 3) Earlham Institute, Norwich, UK; 4) Big Data Institute, University of Oxford, Oxford, UK; 5) Cancer, Ageing and Somatic Mutation Programme, Wellcome Sanger Institute, Hinxton, UK; 6) School of Medicine/School of Mathematics and Statistics, University of St Andrews, UK; 7) Cancer Research UK Cambridge Institute, University of Cambridge, UK; 8) Ontario Institute for Cancer Research, Toronto, Ontario, Canada; 9) European Molecular Biology Laboratory (EMBL), Heidelberg, Germany; 10) Biotech Research & Innovation Centre (BRIC), Copenhagen, Denmark; 11) University of Melbourne, Melbourne, Australia; 12) Centre de Recherche sur les Pathologies Prostatiques et Urologiques (CeRePP), Paris, France; 13) University of Tampere, Tampere, Finland; 14) Department of Urology, Charité – Universitätsmedizin Berlin, Berlin Germany; 15) Manchester Cancer Research Centre, University of Manchester, Manchester, UK; 16) Brigham & Womens Hospital, Boston, USA; 17) Dana-Farber Cancer Institute, Boston, USA; 18) DKFZ - German Cancer Research Centre, Heidelberg, Germany; 19) Uro-Oncology Research Group, Cancer Research UK, Cambridge Institute, Cambridge, UK; 20) The Royal Marsden Hospital NHS Trust, London, UK.

* Joint last authors Prostate Cancer (PrCa) is the fourth most common type of cancer worldwide, and the most commonly diagnosed cancer in men in the UK. Few heritable factors predictive of aggressive prostate cancer have been identified to date. Although common variants identified to date explain over a quarter of the familial relative risk of PrCa, GWAS subset analysis of aggressive disease has failed to find loci specifically associated only with the aggressive phenotype. The aim of this study was to find coding germline variants that associate with clinical outcome of PrCa. We sequenced 226 whole genomes from blood DNA from UK samples at a target depth of 30x. We used a GATK pipeline to identify germline mutations. After sample and variant QC, 188 samples were included in the downstream analysis, and 6,532 likely deleterious coding variants were identified. Pathway analysis was applied to all 50 Hallmark pathways from GSEA MSigDB and an additional three gene sets: a curated DNA Repair Gene panel of 175 genes from our previous work, the BROCA extended panel of 66 genes, and a panel of 115 genes comprising candidate PrCa genes we have curated for a study of treatment of DNA repair gene mutation carriers. We used 2 different methods to assess impact of gene sets on poor prognosis using relapse as outcome: (a) multivariate Cox model, with covariates for Gleason, PSA at tumour collection and pathological T score; (b) and a regularized (Lasso) multivariate Cox model, with the same covariates. Both models showed that alterations in the angiogenesis pathway are significantly associated with poorer prognosis ($p = 0.015$, no multiple testing correction applied). For angiogenesis, only 3/36 genes (*CXCL6*, *JAG2* and *POSTN*) had likely deleterious mutations, and the Cox model with Lasso penalisation indicated that *CXCL6* is the gene with highest impact (hazard ratio = 2.73). This study is part of an international effort called the Pan Prostate Cancer Group (PPCG), which aims to analyse whole genome sequences from >2000 PrCas, with matched transcriptome and methylome data. PPCG is the largest worldwide single cancer organisation with such datasets, and builds upon the CRUK-UK International Cancer Genome Consortium (ICGC) Prostate Cancer Project. We will use further samples from other PPCG cohorts to validate our results.

882W

Enhanced breast cancer risk prediction from imputed gene expression enhanced breast cancer risk prediction from imputed gene expression. J. Zhou¹, G.P. McVicker², H.L. Carter¹. 1) University of California San Diego, San Diego, CA; 2) Salk Institute for Biological Studies.

Early detection for breast cancer is significantly associated with long-term patient survival. Although most individual variants identified in breast cancer contribute weakly to breast cancer risk, they can be combined into genetic risk scores (GRS) to significantly improve risk prediction for some cancer types. An existing method for calculating GRSs uses germline genome-wide significant single nucleotide variants (SNVs); however, this approach is limited to modeling individual contributions of highly statistically significant risk SNVs, which have small effect sizes and account for only a small fraction of heritable disease risk. Furthermore, it is difficult to determine the underlying biological mechanisms by which these SNVs drive breast cancer susceptibility. In this project, we aim to model how genetic variants contribute to disease risk by determining how an individual's SNVs affect gene expression within relevant cell types and then associating these changes with cancer risk. To do this, we performed a transcriptome-wide association study (TWAS) to identify expression-trait associations from a breast cancer GWAS meta-analysis using FUSION (Gusev et al. 2016). Using this approach, we recovered a set of 47 transcriptome-wide significant genes for further study. In conjunction with this analysis, we are training linear models on reference panels of individuals with genotype and expression data in breast-cancer specific tissues to model the functional weights of genes, which can then be used to predict expression in individuals for which expression data is not available. We will map the TWAS results to biological networks and perform network analyses, such as random walk with restart and Google PageRank, to recover core network modules associated with breast cancer risk. Because SNVs often affect disease susceptibility via perturbed gene expression, these modules will be used to develop functionally-accurate GRS with improved predictive power for breast cancer.

883T

Gene prioritization using HiChIP and eQTL in prostate cancer. C.

Giambartolomei¹, J. Seo², X. Qui², M.K. Freund¹, N. Mancuso¹, A. Gusev², M. Freedman², B. Pasaniuc¹. 1) UCLA, Los Angeles, CA; 2) Dana-Farber Cancer Institute, Boston, MA.

Genome-wide association studies (GWAS) have led to the identification of >100 prostate cancer risk loci explaining ~33% of familial risk, which usually reside in intergenic enhancers or other noncoding DNA features. While typically risk SNPs are annotated to the nearest gene, these SNPs are likely to influence expression of target genes located kilobases away, in a cell-type and time specific manner. The integration of expression quantitative trait loci (eQTLs) in normal and tumor tissues with GWAS of prostate cancer has proven to be a useful strategy to prioritize genes beyond the nearest gene in the linear genome, and formal statistical approaches have been developed to integrate GWAS and eQTLs (e.g. coloc, TWAS). However, these approaches are still limited in identifying the correct causal genes, with multiple genes normally identified within a region because of statistical correlations between genetic variants. Recently, HiChIP, an efficient novel protein-mediated chromatin-conformation assay, has been used to generate contact maps of active enhancers-to-target genes. Due to the role of enhancer-promoter interactions in the regulation of gene expression, interaction data can be used together with GWAS and eQTL to prioritize target genes for prostate cancer. Herein, we analyzed H3K27ac and K4me3 HiChIP on prostate carcinoma cell lines LNCaP. Using 81,318 case and 61,074 control individuals genotyped on the Oncoarray, we find 247 enhancer-promoter loops that overlap a GWAS signal (95% credible set). We observe that GWAS from prostate cancer are enriched (~66%) within the HiChIP interactions, and that the heritability increases with number of loops anchored at each peak. We further integrate these results with eQTL data from Cancer Genome Atlas (TCGA) to explore the credible SNP-contact genes. We prioritized genes using HiChIP results together with eQTLs and GWAS, and identify genes supported by physical interactions between regulatory elements and their target genes. We demonstrate the advantage of leveraging HiChIP data to facilitate the identification of target genes at GWAS loci.

884F

Combining enhancer DNA methylation and RNA-seq to map gene regulatory network changes that drive chemotherapy resistance in recurrent ovarian cancer. N. Gull¹, M. Jones¹, S. Coetzee¹, A. Reyes¹, J. Plummer¹, B. Davis¹, S. Chen¹, J.P. Govindavar², J. Rutgers², J. Lester², K. Lawrenson^{1,3}, D. Hazelett¹, B. Karlan², S. Gayther^{1,4}, B.P. Berman^{1,4}. 1) Center for Bioinformatics and Functional Genomics, Department of Biomedical Science, Cedars-Sinai Medical Center, Los Angeles, CA, 90048; 2) Dept of Pathology and Laboratory Medicine, Cedars-Sinai Medical Center, Los Angeles, CA, 90048; 3) Women's Cancer Program, Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA, 90048; 4) Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA.

Epithelial ovarian cancer (EOC) is the most lethal and predominant pathologic subtype of ovarian cancer, usually diagnosed in the late stage of disease, leading to a strong decrease in 5-year survival rate as disease severity worsens. The most common subtype of EOC is high grade serous ovarian cancer (HGSOC). There has been little improvement in patient survival or treatment for HGSOC in recent decades, despite surgical resection and chemotherapy over 70% of patients will experience disease relapse, and eventually chemotherapy resistance will progress to fatal disease. Previous studies of methylation in HGSOC primary tumors using microarrays show variations in methylation patterns that differ between chemo-resistant and chemo-responsive patients, such as *FZD10* and *MXK*. Additionally, *BRCA1* promoter hypermethylation (sometimes reported in germline or tumor from patients who do not carry a germline BRCA mutation) was linked to a reduction in patient survival and was observed in patients with recurrent tumors. Other alterations in methylation patterns of oncogenes (*PIK3CA*), tumor suppressor genes (*WT1*), and driver genes (*TP53*) have also been observed in EOC; these alterations indicate key changes in gene regulatory networks (GRNs) and could affect tumor biology. In order to identify changes in the methylome of chemotherapy resistant HGSOC tumors, we have generated a genomic landscape of the changes that occur as a result of chemoresistance in each available tumor (n=73) with whole genome sequencing, whole genome bisulfite sequencing (WGBS), and RNA-seq. Our WGBS data represents the first whole methylome analysis of HGSOC patients to date. These data will allow us to apply methods that we have pioneered to infer GRN changes by combining WGBS and transcriptome data to identify master regulator transcription factors and their downstream target genes that are disrupted as a result of chemotherapy in HGSOC. We have identified several alterations in methylation in chemo-resistant HGSOC tumors; regions of the genome where hyper and hypo-methylated regulatory regions perturb the expression of genes from cell division and DNA repair pathways. Our combinatorial approach has provided a genome-wide *cis*-regulatory epigenetic landscape in ovarian cancer as it acquires and maintains chemoresistance, and identifies potential novel therapeutic targets.

885W

Effects of structural variation on proximal lncrnas and mrnas in b-cell leukemia. C. Nodzak¹, V. Vasconcelos², J.A. Yunes², X. Shi¹. 1) Dept. of Bioinformatics and Genomics, University of North Carolina at Charlotte, Charlotte, NC; 2) Laboratorio de Biologia Molecular, Centro Infantil Boldrini, Campinas, Sao Paulo, Brazil.

Throughout cellular development the local three-dimensional structure of the genome facilitates coordinated regulation of gene expression. These conserved and cell-type specific organizational and architectural elements of the genome have been denoted topologically associated domains (TADs), and may encompass noncoding and coding genes relegating them to shared nuclear space allowing for possible interactions to occur. Antisense long noncoding RNA (lncRNA) transcripts have been shown to facilitate some aspects of chromatin remodeling at coding gene loci through R-loop formation and binding of proteins capable of modulating the epigenetic state along gene and promoter regions. In addition, lncRNAs may act as microRNA sponges or form complexes with proteins, or directly bind mRNA molecules, which can contribute to changes in the amount of available mRNA. B-cell acute lymphoblastic leukemia has been characterized by a collection of recurrent genomic rearrangements that disrupt the relationship between proximal lncRNAs and mRNAs, and may create non-cognate enhancer mediated gene expression events that contribute to irregular expression of these genes. In this study, we show that the expression levels of lncRNAs and protein coding genes located within cell type specific intrachromosomal topologically associated domains (TADs) from GM12878 are able to accurately classify B-cell acute lymphoblastic leukemia patients by the major recurrent chromosomal rearrangement using published data from St. Jude's and the TARGET initiative. These findings suggest that we are able to capture some regulatory relationships between gene pairs overlooked by simple differential expression, which is of particular significance for discernment between Ph-chromosome(+) and Ph-like samples that share global transcriptome patterns yet lack common genomic rearrangements. We further interrogate the effects of structural variants located nearby TAD boundaries to establish links between structural variation, chromatin topology, lncRNA and gene expression variation among different subtypes of leukemia based on recurrent rearrangements.

886T

Analysis of the *HSF1* transcription program using clustering frameworks to predict clinical cancer endophenotypes. L. Bang¹, T. Prince², D. Kim^{1,3}. 1) Biomedical and Translational Informatics Institute, Geisinger Health System, Danville, PA; 2) Department of Urology, Geisinger Health System, Danville, PA; 3) The Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, PA.

Heat shock factor 1 (*HSF1*) is a master transcription factor that regulates the cellular stress response along with enabling malignant cell growth. The *HSF1* transcription program includes over 2,000 genes as defined by chromatin binding, yet linking patient subgroups based on their expression profiles into clinical outcomes is understudied. A rapid iteration method for clustering gene expression profiles was carried out to predict clinical endophenotypes. In this study, we performed the clustering analysis to identify patient subgroups that are associated with cancer endophenotypes. Over and under-expression profiles of genes within the *HSF1* transcription program were attained from TCGA for 5 tumor types: prostate adenocarcinoma (PRAD), breast invasive carcinoma (BRCA), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC) and low-grade glioma (LGG). Using k-means and hierarchical clustering techniques, patients with common gene expression patterns within the *HSF1* transcription program were found to also have differential patterns of clinical variables. The k-means algorithm was iterated over multiple hyperparameters, and differential distribution of clinical variables in each cluster measured with a chi-square test and validated by comparing with a permutation test. This method allowed us to identify PRAD patients with likely lower or higher Gleason scores. For BRCA, LUAD and LUSC we were able to predict lower or higher tumor grades, while for LGG we could predict tumor histological type. Using clustering analyses may enable us to gain insight into how *HSF1* and its transcription program drives malignancy and influences cancer endophenotypes.

887F

Detecting cancer vulnerabilities through gene networks under purifying selection in 4,700 cancer genomes. A. Gupta^{1,2,3}, H. Horn^{1,2,3}, P. Razaz^{1,2,3}, M. Lawrence^{1,2,3}, G. Getz^{1,2,3}, K. Lage^{1,2,3}. 1) Harvard Medical School, Boston, MA; 2) Broad Institute of Harvard and MIT, Cambridge, MA; 3) Massachusetts General Hospital, Boston, MA.

Large-scale cancer sequencing studies have uncovered dozens of mutations critical to cancer initiation and progression. However, a significant proportion of genes linked to tumor propagation remain hidden, often due to noise in sequencing data confounding low frequency alterations. Further, genes in networks under purifying selection (NPS), or those that have undergone negative selection, may play crucial roles in sustaining cancers but have largely been overlooked. We describe here a statistical framework that identifies genes that have a first order protein interaction network significantly depleted for mutations, to elucidate key genetic contributors to cancers. Not reliant on and thus, unbiased by, the gene of interest's mutation rate, our approach has identified 685 putative genes linked to cancer development. Comparative analysis indicates statistically significant enrichment of NPS genes in previously validated cancer vulnerability gene sets. Specifically, the comparisons yield odds ratios of 5.4 ($p < 2.9 \times 10^{-5}$) for copy number alterations (CYCLOPS) and 4.7 ($p < 2.2 \times 10^{-16}$) for dependencies as determined by knockdown studies (DEMETER). Ninety-five of the 685 NPS genes are not crucial to cells in the general population (probability of loss-of-function intolerance scores < 0.9 for genes in the Exome Aggregation Consortium), and aggregate knockdown of these genes results in aberrant tumor growth, as determined by the Achilles cancer cell line gene knockout database. We are currently validating the highest priority NPS candidate cancer genes through functional assays. As more tumor genomes are sequenced, integrating systems level mutation data through this network approach should become increasingly useful in pinpointing gene targets for cancer diagnosis and treatment.

888W

Prognostic model for multiple myeloma progression integrating gene expression and clinical features. C. Sun¹, H. Li¹, R. Mills^{1,2}, Y. Guan^{1,3}. 1) Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI; 2) Department of Human Genetics, University of Michigan, Ann Arbor, MI; 3) Department of Internal Medicine, University of Michigan, Ann Arbor, MI.

Multiple Myeloma (MM) is a hematological cancer caused by abnormal accumulation of monoclonal plasma cells in bone marrow. With the increase in treatment options, risk-adapted therapy is becoming more and more important. Survival analysis is commonly applied to study progression or other events of interest and stratify the risk of patients. In this study, we present the current state-of-the-field model for MM prognosis and the molecular biomarker set for stratification: the winning algorithm in prediction accuracy in the 2017 Multiple Myeloma DREAM Challenge. Specifically, we built a non-parametric complete hazard ranking model to convert the right-censored data into a linear space, where commonplace machine learning techniques, such as Gaussian process regression and random forests, can play their roles. The model combined the gene expression profile and clinical features to predict the progression of MM. Compared with conventional models, such as Cox model and Random Survival Forests, our model achieved higher accuracy in 3 within-cohort predictions. In addition, it showed robust predictive power in cross-cohort validations and in the validation cohorts. A set of molecular signatures related to MM progression was extracted from our model, which may function as the core determinants of MM progression and of important guidance value for future research and clinical practice. Functional enrichment analysis and mammalian gene-gene interaction network revealed several key biological processes and pathways involved in MM progression..

889T

Characterization of chromosomal allelic imbalances through RNA-seq.

Z. Ozcan^{1,2}, F.A. San Lucas¹, Y. Liu^{1,2}, Z. Weber², E. Ehl², G. Davies², R.G. Fowler², P. Scheet^{1,2}. 1) Department of Epidemiology, The University of Texas M.D. Anderson Cancer Center, Houston, TX; 2) The University of Texas M.D. Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences, Houston, TX; 3) Avera Institute for Human Genetics, Sioux Falls, SD.

Transcriptome sequencing (RNA-seq) is becoming a standard aspect of tumor profiling, extending beyond transcript quantification and identification of alternative transcripts and gene fusions. For instance, recent advancements in RNA-seq data analyses, allow for computational assessment of allele-specific gene expression and generate profiles of expressed somatic mutations. Here, we identify chromosomal allelic imbalances (AI) through detection of haplotype-specific patterns in gene transcripts. To do so, we obtained RNA-seq and SNP array or whole-exome sequencing (WES) data for five cancers in the TCGA: uterine carcinosarcoma (UCS), lung (adenocarcinoma and squamous cell carcinoma), prostate and pancreatic. AI calls were made in both the exome and RNA-seq data by: (1) calling 1000 Genome Project variable sites in the sequencing data, (2) phasing haplotypes, and (3) characterizing allelic imbalances in RNA and DNA using hapLOHseq. SNP array-based AI was detected with hapLOH. RNA-seq- and WES-derived AI patterns exhibited equivalent specificity (72%) based on comparisons to AI events inferred from SNP arrays, which we consider here as a gold standard. However, RNA-seq appeared to have lower sensitivity, detecting 43% of the chromosomal AI events identified by WES. Interestingly, we observed AI patterns in RNA that were not evident in paired DNA samples and attempted to relate these features to survival in UCS. We found a suggestive negative association ($P = 0.076$), with higher RNA-specific AI load in UCS patients having a median survival of 771 days compared to 1526 days for those patients with lower loads of RNA-specific AI. Our results suggest that analysis of chromosomal AI in RNA-seq has equal specificity for detecting DNA-level AI when compared to exome sequencing, although at lower sensitivities, and thus may be a useful surrogate for DNA changes in resource-limited settings. Clinically, our analyses suggest that patients with higher RNA-specific AI load may have a worse overall prognosis; we are investigating this further in the data from lung, prostate and pancreatic. The AI we identify in the RNA-seq samples may reflect long-range dysregulation for even copy-neutral, balanced, regions of the genome. One possible cause of RNA-specific AI could be the presence of cis mutations that impact a large-region of the transcription of one of the two haplotypes.

890F

Alignment and analysis of xenograft transcriptomes. S. Tapas¹, S.

Ramachandrala², H. Kumar Adil², D. Chohan², V. Chakraborty². 1) School of BioHealth Informatics, Indiana University Purdue University, Indianapolis, IN; 2) Strand Life Sciences Private Limited, Kirlskar Business Park, Bengaluru, Karnataka, India.

Studies on Patient-Derived Xenografts (PDX) are of immense value in understanding pharmacological mechanisms. However, one of the major challenges is the involvement of stromal cells along with the tumor cells which are of interest. While the Next-Generation Sequencing (NGS) technology has greatly aided in understanding these mechanisms, normal/stromal contamination remains to be resolved to ensure that the expression profiles truly represent the tumor state. Therefore, aligning read sequences to Human Reference alone may not be accurate. Trimming of read sequences based on host reference genome can lead to the loss of reads that are common to both due to homology. It is now agreed that alignment of read sequences to a master Reference consisting of both Human and host would give the most unbiased expression profile. However, due to the memory requirements of dealing with such a large reference and the associated technical challenges makes this approach unappealing to research biologists. Here, we showcase how this could be accomplished easily in StrandNGS, a desktop software for NGS data analysis. To illustrate this approach, we have used data from GSE102370 submitted by Goyette et al. Expression analysis of GSE58644 (Affymetrix arrays) was performed using GeneSpring GX v14.9 software to understand the gene expression profiles of genes that are known to impact survival. A concatenated reference created using reference genomes of mouse (mm10) and human (hg38) was used to align the RNA-Seq samples from GSE102370 in StrandNGS v3.1.1 software. The respective gene annotations were also concatenated. All the downstream analyses were carried out as discussed by the author. A comparison was made between the expression profiles reported by the authors and those that were derived using the current approach. It was observed that the profiles were different and qualifying gene sets varied.

891W

SoVaTSiC: Somatic variant tool for single cell. A. Javed^{1,2}, J.H.J. Tan², S.L. Kong², J.A. Tai², H.M. Poh², F. Yao², Y.Y. Sia², E.K.H. Lim², T.W. Tan³, A.M.T. Pena⁴, D.S.W. Tan⁴, A.M. Hillmer^{1,5}. 1) School of Biomedical Sciences, The University of Hong Kong, Hong Kong, China; 2) Genome Institute of Singapore, A*STAR, Singapore; 3) Department of Biochemistry, National University of Singapore, Singapore; 4) Department of Medical Oncology, National University of Singapore, Singapore; 5) Institute of Pathology, University Hospital Cologne, Germany.

Single cell omics provides an opportunity to interrogate genetic and phenotypic heterogeneity at an inter-cellular level. A key application is in the cancer domain where this heterogeneity can provide an avenue for the tumor to acquire a path to therapeutic resistance. With growing popularity, there is a demand for computational tools to aid single cell analysis. The focus thus far has primarily been on transcriptomics. In particular there are no tools designed to tease apart somatic mutations in single cell genomics data. We provide SoVaTSiC the first software specifically aimed to address this task. The software bundle includes quality control measures along with practical guidelines to fine tune the tool based on the noise (allelic drop out and false negative rate) observed in the data. Previous results in a lung adenocarcinoma patient, showed high accuracy (86% validation rate, combining true positives and true negatives). They revealed insight into tumor phylogeny by identifying key events in tumor evolution. These inferences are supported by the histological evidence. We have further evaluated the software in a direct comparison with the state of the art single cell variant caller MONOVAR, SoVaTSiC outperforms in accuracy versus specificity tradeoff in both in-house as well as public datasets. SoVaTSiC somatic mutations are more consistent with the inferred phylogeny of the tumor and include mutations of direct therapeutic relevance missed by MONOVAR (which is primarily intended for germline variants).

892T

Single-cell RNA sequencing analysis of the non-small cell lung cancer cell line after DNA demethylating agent treatment. J.W. Zeng^{1,2}, S.F. Su^{1,3}, H. Ho¹, Y.H. Chen¹, H.W. Chen⁴, K.C. Li^{1,5}. 1) Institute of Statistical Science, Academia Sinica, Taipei, Taiwan, Taipei, Taiwan; 2) Data Science Degree Program, National Taiwan University, Taipei, Taiwan; 3) Graduate Institute of Oncology, National Taiwan University, College of Medicine, Taipei 10051, Taiwan; 4) Graduate Institute of Toxicology, College of Medicine, National Taiwan University, Taipei, Taiwan; 5) Department of Statistics, University of California, Los Angeles, USA.

Motivation: Lung cancer is the leading cause of cancer death worldwide with non-small cell lung cancer (NSCLC) being its most common form. However, the molecular basis of the substantial response-heterogeneity remains unclear. The CpG island methylator phenotype (CIMP) has been identified in NSCLC. CIMP tumors exhibit distinct genetic features that correlate with clinical outcome, thus highlighting the role of DNA methylation in tumorigenesis. To study the gene expression alteration induced by DNA demethylating agents at the single cell level, we applied the single-cell RNA sequencing (scRNA-Seq) on the treated and mock samples of our previously established NSCLC cell line. **Material and Method:** The NSCLC cell line was prepared for the mock and drug treatment. A DNA demethylating agent was used for consecutive three-day treatment. Both mock and treated samples were trypsinized on day seven, and loaded onto the 10x microfluidics chip for scRNA sequencing. We developed a workflow to analyze the scRNA-Seq data from data preprocessing to cell-cycle analysis. Differential expression (DE) analysis was utilized to find the genes that expressed differently between mock and treated samples.

Result: 511 mock and 797 treated cells were obtained from 10x Genomics Cell Ranger. We performed cell-cycle analysis to distinguish cells in different phases of the cell cycle: G1/S, S, G2/M, M and M/G1. We found that the number of cell cycle associated genes identified in treated cells was less than cells in the mock sample. Also, for cells in the S phase, we identified a list of significant DE genes associating with tumor suppressors. **Conclusion:** scRNA-Seq is a robust and economically practical tool for studying transcriptome at the single cell level, and it provides a better resolution for the drug's impact on the gene expression. Even though the numbers of cells loaded to scRNA-Seq did not reach the optimal number of cells, our workflow was able to identify the cell-cycle correlated genes exhibiting differential expression between the mock and drug-treated lung cancer cell lines.

893F

Role of alternative polyadenylation dynamics in acute myeloid leukemia at single-cell resolution. Q. Zhou^{1,2}, C. Ye¹, Y. Hong², Q.Q. Li^{1,2}. 1) Western University of Health Sciences, Pomona, CA; 2) Xiamen University, Xiamen, Fujian 361102, China.

Alternative polyadenylation (APA) has been discovered to play regulatory roles in the development of many cancer cells through preferential addition of a poly(A) tail at specific sites of pre-mRNA. A recent study found that APA was involved in the mediation of acute myeloid leukemia (AML). However, unlike gene expression heterogeneity, little attention has been directed toward variations in single-cell APA for different cell types during AML development. Here, we used single-cell RNA-seq data of a massive population of bone marrow mononuclear cells (BMMCs, ~17k) from healthy and AML patient samples to investigate dynamic APA usage in different cell types. Abnormalities of APA dynamics in the BMMCs from AML samples were uncovered compared to the stable APA dynamics in samples from healthy individuals, as well as lower APA diversity between eight cell types in AML. Genes with APA usage specific to the AML samples were significantly enriched in cellular signal transduction pathways that contribute to AML development. Moreover, many leukemic cell marker genes such as *NF-κB*, *STAT1*, *GATA2* and *P27* exhibited APA dynamics that specifically affected abnormal proliferation and differentiation in BMMCs of AML patients. Additionally, mature erythroid cells displayed greater APA dynamics and global 3' UTR shortening compared with other cell types, and APA diversities along with the erythropoiesis process decreased gradually. Our results revealed extensive involvement of APA regulation in leukemia development and erythropoiesis at the single-cell level, providing a high-resolution atlas to navigate cellular mRNA processing landscapes of differentiated cells in AML.

894W

Genetic heterogeneity profiling by single cell RNA sequencing. Z. Zhou¹, N. Zhang². 1) Graduate Group of Genomics and Computational Biology, University of Pennsylvania, Philadelphia, PA; 2) Wharton Statistics Department, University of Pennsylvania, Philadelphia, PA.

Single cell RNA-sequencing (scRNA-seq) technologies have made it possible to study tumor transcriptomes at single-cell resolution, allowing better cell type classification and unprecedented study of intratumor transcriptomic heterogeneity. Yet, the underlining mechanisms for tumor initiation, progression, metastasis and relapse are often driven by genetic heterogeneity, that is, subclonal evolution involving genetically distinct subpopulations of cells. Detection of these genetically distinct subclones and understanding the evolutionary dynamics of the tumor is essential for accurate prognosis and effective treatment. A tumor's genetic heterogeneity underlies its transcriptome heterogeneity, and thus, intra-tumor transcriptomic variation needs to be quantified within the context of intra-tumor clonal variation. There is yet no method that can derive the genetic mutation profile and gene expression level for the *same* cell with acceptable accuracy, high throughput, and low cost. Inference of clonal membership from scRNA-seq data is currently unreliable for all but the most dominant clones. Here we propose a computational framework, DENDRO (DNA based EvolutionNary tree preDiction by scRNA-seq technOlogy), that takes scRNA-seq and matched bulk DNA sequencing data for a tumor and accurately reconstructs its evolutionary phylogeny, assigning each single cell from the scRNA-seq data to a subclone. DENDRO utilizes information from both copy number aberrations and single nucleotide mutations in transcribed region, and is based on a likelihood model that accounts for single-cell level expression stochasticity as well as the technical noise in scRNA-seq. The accuracy of DENDRO is benchmarked on *in silico* mixture and spike-in data sets, and its effectiveness is illustrated through case studies using published scRNA-seq data from multiple cancer types. Finally, we demonstrate an online app, based on DENDRO, to guide researchers in experimental design, e.g. determining the number of cells and coverage per cell, for subclonal inference by scRNA-seq data.

895T

Next-generation sequencing with comprehensive bioinformatics analysis facilitates somatic mosaic APC gene mutation detection in patients with familial adenomatous polyposis. *E. Kim^{1,2}, B. Kim², S.T. Lee², J.R. Cho².* 1) Brain Korea 21 PLUS Project for Medical Science, Yonsei University; 2) Department of Laboratory Medicine, Yonsei University College of Medicine.

Background: It was recently recognized that patients with clinical familial adenomatous polyposis (FAP) but without detectable pathogenic mutations are associated with somatic mosaic APC mutation. **Methods:** We reanalyzed the Next-generation sequencing (NGS) gene panel testing results of such patients and tested several variant calling algorithms to identify low level mosaic variants. Twenty-eight patients with clinical suspicion of FAP but no detectable pathogenic variants of colonic polyposis associated genes were included, and endoscopic biopsy samples of patients were analyzed. Variant calling tools HaplotypeCaller, MuTect2, VarScan2, and Pindel were used. We also used 3'-Modified Oligonucleotides (MEMO)-PCR or conventional PCR for confirmation. **Results:** Among 28 patients, seven somatic mosaic pathogenic variants (five frameshift variants and two nonsense variants) were identified, with variant allele frequency ranging from 0.3 to 7.7%. These variants were mostly detected through MuTect2 and Pindel. The somatic mutations found in leukocytes were enriched in the colonic polyp from 7% to 20% in a patient. The somatic mosaic pathogenic variants were further confirmed using MEMO-PCR or conventional PCR. **Conclusions:** NGS testing with an adequate combination of bioinformatics tools is effective to detect low level somatic variants and deletions in a single assay. Because mosaic APC mutation is more frequent than previously thought, the presence of mosaic mutation must be considered when analyzing genetic tests of patients with FAP.

896F

Algorithm-driven comparison analysis of liquid biopsy samples. *X. Xu¹, D. Stetson², D.L. Gross¹, J. Hull¹, M. Kaganovich¹, J. Johnson², J.C. Barrett², B.A. Dougherty², D.A. Caplan¹.* 1) SolveBio, New York, NY; 2) Translational Sciences, Oncology, IMED Biotech Unit, AstraZeneca, Boston, MA.

Background Liquid biopsy, or the detection of tumor-derived DNA from blood samples, is increasingly used in clinical decision making and translational research. The consistency, accuracy, and detection thresholds of this new technique are not well characterized. Comparison analyses of variant and allele frequency detection from the same patient across sample types, commercially available tests, and patient timepoints are crucial to understanding technical and biological limitations of liquid biopsy and implications for individual patients. Currently, comparison analyses are custom, requiring informatics expertise and error-prone manual analysis. **Methods** We created a set of software modules for liquid biopsy comparison analyses, utilizing the SolveBio platform for indexing, querying, and normalizing genomic data. Twenty four replicate plasma samples were sent to 4 commercial circulating tumor DNA (ctDNA) Next Generation Sequencing (NGS) labs. Matched tumor-normal tissue was sent to one lab. Our novel software enabled programmatic 1) normalization of reported genomic data among multiple commercial test reporting formats and panel sizes, 2) algorithmic removal of confounders such as germline and clonal hematopoiesis variants, 3) merging of the results into an easy-to-query table and 4) visualization of patient-level and study-level results. **Results** Our software and analysis pipeline merged 698 calls from 24 patients with 6 sets of results (4 plasma, 1 tumor, 1 normal) into 465 unique variants, saving dozens of person-hours of manual analysis. Programmatic removal of variants classified as germline or clonal hematopoiesis-related reduced this set to 56 variants. We designed novel plots to visualize this complex data. Beyond the initial use case, these software modules have been used in multiple other tissue-plasma comparison studies. Substantial technical characterization of liquid biopsy techniques, labs, and the genomic output is critical for widespread clinical adoption of this technology. Our software platform allows for the comparison and analysis of ctDNA NGS results. We intend to expand our software to enable the cross-lab comparison of emerging biomarkers of therapeutic response, such as tumor mutational burden and microsatellite instability (MSI) status.

897W

Strelka2: Fast and accurate germline and somatic variant calling. S. Kim¹, K. Scheffler¹, A.L. Halpern¹, M.A. Bekritsky², E. Noh¹, M. Källberg², X. Chen¹, Y. Kim¹, D. Beyter³, P. Krusche², C.T. Saunders¹. 1) Illumina, Inc., 5200 Illumina Way, San Diego, CA, USA; 2) Illumina Cambridge Ltd., Chesterford Research Park, Little Chesterford, Essex, UK; 3) Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA, USA.

We present Strelka2 (<https://github.com/Illumina/strelka>), an open-source small variant calling method for clinical germline and somatic sequencing applications. Strelka2 introduces a novel mixture-model based estimation of indel error parameters from each sample, an efficient tiered haplotype modeling strategy and a normal sample contamination model to improve liquid tumor analysis. For both germline and somatic calling, Strelka2 substantially outperforms current leading tools on both variant calling accuracy and compute cost. Strelka2's germline calling performance was measured on data from the recent PrecisionFDA challenges. For the PCR+ and PCR-free data, Strelka2's indel F-score improved upon the best submission by 3.2% and 0.11%, respectively, while giving single nucleotide variant (SNV) results within 0.06% - 0.1% of the best submissions. Strelka2 also greatly outperformed recently released versions of GATK4 and FreeBayes in accuracy while being 7.7 and 3.5 times faster, respectively. We evaluated Strelka2's somatic variant calling accuracy by mixing sequencing data of NA12878 and NA12877 to simulate datasets with varying tumor and normal purities. On these in silico mixture data, Strelka2 showed substantially higher precision than MuTect2 and VarDict at all recall thresholds over all test datasets, with an average F-score improvement of 19-25% for SNVs and 16-20% for indels. We note Strelka2's tolerance to normal sample impurity, reflecting updates to better support liquid and late-stage solid tumor analyses. At 80% tumor sample purity, a change in normal sample purity from 100% to 90% generated a drop in F-score for MuTect2 and VarDict, respectively, of 65% and 37% for SNVs and 42% and 33% for indels. For Strelka2, the same change only caused a drop in F-score of 5.6% for SNVs and 18% for indels. As was the case for the germline analysis, Strelka2 somatic analysis was substantially faster than available alternatives, demonstrating an average runtime advantage of 8x over MuTect2 and 3x over VarDict. We reiterate that our results were generated with default method settings appropriate for factory-scale analysis, not requiring human intervention to parameterize or connect complex sequences of tools.

898T

Measuring the effect of genetic variation at splicing junctions using a cellular fluorescence reporter system. L. Wan, N. Torres, C. Yeh, C. Maclean. Ranomics, Toronto, Ontario, Canada.

Sequencing technologies have facilitated comprehensive cataloguing of disease-associated single nucleotide polymorphisms (SNPs) at intron-exon junctions. However, determining the impact of these variants on alternative splicing or protein function continues to be a significant bottleneck. Minigene assays are the current gold standard for measuring splicing efficiency, but they are tedious to implement due to RNA isolation steps, and offer a qualitative assessment of splicing. To overcome the shortcomings of the minigene assay, we have developed a simple, fluorescence-based reporter to quantify splicing efficiency of an intron-exon junction of interest in human cells. Splicing efficiency can be measured quantitatively and on a per-cell basis by fluorescence microscopy or through cell sorting. Our splicing reporter includes an intrinsic fluorescence control for reporter intake or integration efficiency, and is designed to be flexible and scalable. We have performed pilot experiments with variants of *BRCA1* intron-exon junctions at which pathogenic mutations have been observed that are known to disrupt splicing to demonstrate the clinical and biological relevance of our reporter system.

899F

Beyond the exome: Identification of novel mutational signatures and oncogenic dependencies in the non-coding genome of multiple myeloma.

C. Ashby, M.A. Bauer, R.G. Tytarenko, P. Patel, S. Deshpande, F. van Rhee, M. Zangari, S. Thanendrarajan, C.D. Schinke, F.E. Davies, G.J. Morgan, B.A. Walker. University of Arkansas for Medical Sciences 4301 W. Markham St. Little Rock, AR 72205.

In multiple myeloma (MM), over one thousand samples have undergone whole exome sequencing (WES), showing key oncogenic dependencies within molecular sub-groups. Analysis of WES data has revealed two mutational signatures: a non-specific signature and an AID/APOBEC signature associated with the t(14;16) molecular sub-group. However, most of the genome has remained unstudied as a consequence of an almost exclusive focus on the exonic regions. However, we know from other cancers that non-coding regulatory mutations play a role in disease progression, gene deregulation and prognosis. We have examined 10X Chromium whole-genome sequencing (WGS; mean depth=45x) of 98 MM patients in different progressive states. Matching RNA-seq (n=85 patients) and microarray data (n=98) were used for gene expression analysis, as were additional non-matching microarray datasets for outcome (n=1553). Significantly more mutations ($P<0.0001$) were found in non-exonic (median 7.04 / Mb) vs. the exonic space (median (3.1 / Mb). Mutational signature analysis shows that samples with an AID/APOBEC COSMIC signature have significantly increased mutation counts (median 19692 vs. 26424, $P=0.009$) in both the coding and non-coding regions compared to those without. Additional mutational signatures detected were somatic hypermutation (sig. 9), a general cancer (sig. 5), and two signatures of unknown origin, one with C>G and T>A mutations in a ApNpT/GpNpC/TpNpA context and the other with mostly C>A and C>T mutations. Mutational hotspots indicative of somatic hypermutation were detected in key B cell maturation genes, including *BCL6*, *CXCR4*, *LTB*, and *BACH2*. Additional hotspots were seen in long noncoding RNAs including *NEAT1*, which has been implicated in a variety of other cancers. Mutations in the promoter region of *NEAT1* led to higher expression, which is linked to a shorter time to progression to MM ($P=0.008$) and poorer overall survival ($P=0.0007$). Finally, we performed a mutational enrichment analysis in MM molecular sub-groups, identifying several non-coding RNA genes such as *DANCR* [t(14;20)], *MIR155HG* [t(14;16)] and *MIR663AHG* [t(11;14)] are more frequently mutated in one subgroup compared to the rest of the dataset. We show that WGS identifies additional mutational signatures affecting the non-coding genome as well as mutational hotspots that result in gene dysregulation. This provides additional avenues of investigation to identify novel therapeutic targets in MM.

900W

High-content phenotyping of human genetic mutations by functional variomics.

S. Yi¹, Y. Li², D.J. McGrail², P.K.S. Ng², K.J. Jeong², G.B. Mills², N. Sahn². 1) Dell medical school, Univ of Texas Austin, Austin, TX; 2) UT MD Anderson Cancer Center, Houston, TX.

Human genomes are highly complex with numerous somatic mutations identified across patient populations. Previous studies on genomic mutations have shed light on new means for cancer therapeutic interventions. With rapid advances in next-generation sequencing, accumulating genotypic information in the absence of efficient and systematic functional analyses of genomic aberrations will create a bottleneck in understanding genotype-phenotype relationships in cancer. To address these challenges, here we report a systems-level functional variomics approach integrating high-throughput phenotyping with robust computational analyses to investigate mutation-specific effects. This systematic functional platform consists of massively parallel mutagenesis, sensitive survival assays using growth factor-dependent cell models, and functional network perturbation profiling of mutations on signaling effects. We profile thousands of genomic aberrations, including point mutations, gene fusions and indels, and significantly expand the repertoire of characterized actionable mutations. This study represents a valuable resource and provides insights in prioritizing disease-causing mutations, and uncovering patient-specific disease mechanisms at a high resolution, a critical step towards personalized precision medicine.

901T

Magnis, a fully automated NGS sequencing-ready library preparation system.

S. Bigdeli¹, B. Buehler¹, S. Im¹, H. Hogrefe¹, K-C. Yeap², C-S. Siow², S-K. Lee², S-W. Yuen², O-C. Lim², A-J. Fang², S-F. Ching², R. Bagga¹, J. Hal-laday⁴, S. Jandle⁴, K. Poon³, P. Maunary¹, R. Allen¹. 1) Agilent Technologies, San Diego, CA, USA; 2) Agilent Technologies, Penang, Malaysia; 3) Agilent Technologies, Santa Clara, CA, USA; 4) Agilent Technologies, Cedar Creek, TX, USA.

Agilent's Next Generation Sequencing (NGS) target enrichment kits for preparation of sequencing-ready libraries are in widespread use in laboratories around the world. Automation of these complex protocols eliminates the need for training specialists, improves turnaround time, library quality, and reproducibility for NGS libraries. In a large collaborative effort among several groups within Agilent, we are developing a walk away liquid handler system Magnis capable of preparing up to eight NGS sequencing-ready libraries per run without the need of human intervention during the run using both regular and FFPE samples. We have programmed Magnis and optimized Agilent's HaloPlex HS and SureSelect XT HS chemistries to prepare automated Illumina-ready libraries. The system also comes with pre-aliquoted plated reagents for each chemistry made by Agilent's production facility. Prepared libraries can be kept sealed and refrigerated on the deck for up to 72 hours post-run. In a clinical setting, a decreased turnaround time may aid doctors with patient diagnosis. We have made modifications to the manual target enrichment protocols to optimally generate libraries on Magnis using different baits while utilizing innovative techniques for pipetting to prevent well-to-well contamination. An example we use for testing the Magnis system is one of Agilent's latest baits, the cost-effective SureSelect Human All Exon V7 exome, a hybrid-capture solution focusing on the interpretable part of the human genome. Both chemistries have been fine-tuned and optimized, and libraries have been tested for yield and reproducibility. Sequencing has confirmed reliable coverage, specificity, uniformity, and duplicate reads. We have demonstrated that the integrated germicidal UV bulb installed on the deck removes >99.5% of the amplifiable DNA at all tested deck positions in less than 30 minutes. We have also developed qPCR-based spike-in assays to confirm sample to sample contamination below the limit of detection (less than 0.001%). We have used sequencing to address potential run to run, and index contamination between samples. Protocol run time is currently under 8 hours for HaloPlex HS and under 9 hours for SureSelect XT HS. All yield and sequencing specs are targeted to meet the criteria for manually run protocols.

902F

Detection of gross rearrangements in BRCA1 by next generation mapping and long read technologies.

M. Bonin¹, U. Faust², M. Kehrer², N. Maier¹, A. Gazou², C. Schroeder², S. Kotschote¹, HG. Klein¹. 1) IMGM Laboratories GmbH, Martinsried, Bavaria, Germany; 2) Institute of Medical Genetics and Applied Genomics, University Hospital Tuebingen, Germany.

In monogenetic disorders, the functional investigation of rare, unclassified variants helps to analyze their pathogenic relevance and can improve differential diagnosis and predictive testing. Similarly to other human cancer susceptibility genes, the utility of testing in genes associated with hereditary breast and ovarian cancer, such as *BRCA1* and *BRCA2*, strongly relies on the possibility of establishing a relationship between the nature of the observed genetic alterations and their consequences on the functionality of the corresponding protein products. This can be generally inferred for mutations that introduce exonic duplications/deletions, premature stop codons (i.e. truncation) or affect mRNA integrity and/or stability and splicing, giving rise to functionally compromised proteins. However, apart from common polymorphisms, the assessment of the clinical relevance of other allelic variants may not be equally straightforward. We investigate two independent duplications within *BRCA1* gene of breast cancer patients by different molecular methods like Bionano Genomics and Oxford Nanopore (ONT) Sequencing on DNA and RNA level. In case 1 we are able to analyze a genomic duplication within *BRCA1* gene from exon 1 to 19. Initial diagnostic approaches show 31 MLPA positive markers and no relevant mutations in a 12 gene NGS Panel. With the help of Next Generation Mapping (NGM, BioNano Genomics, DNA) and ONT Long Read Sequencing (DNA & RNA) we were able to define the exact location and the orientation of the duplicated region. For the second case we investigate a smaller unpublished genomic duplication within *BRCA1* gene from exon 17 to 23. Also in this case MLPA shows only some positive markers and the NGS panel of 12 relevant genes for hereditary Breast Cancer results without any relevant variants. We performed NGM and Long Read Nanopore Sequencing for two affected *BRCA1* individuals. Using these two approaches we identified the orientation and the genomic location of the structural variants. In addition, both methods are capable of identifying both single and multiple exon deletions. NGM as well as Long Read Sequencing promises to close the diagnostic gap for exonic deletions and duplications in monogenetic disorders.

903W

Characterizing tumor heterogeneity at single cell level. C. Catalanotti, S. Maheshwari, K. Belhocine, S. Garcia, V. Kumar, L. Hepler, A. Price, T. Wheeler, M. Schnall-Levin, R. Bharadwaj, D.M. Church. 10x Genomics, Pleasanton, CA.

Cancer genomes are highly unstable with new genetic variations emerging even within a single metastatic site. Such heterogeneity is manifested in the forms of multiple genetically distinct clonal populations and is shaped by the ongoing processes of Darwinian selection and drift. To fully understand the landscape of genetic heterogeneity within a tumor is critical for predicting its evolutionary potential. Single cell (SC) genome sequencing is emerging as a powerful tool to explore it, providing a direct readout of the cellular frequency of somatic mutations and relative ploidy, while avoiding the complexities of deconvolution from bulk WGS. However, current SC DNA approaches suffer from low throughput, cumbersome workflow and remain only accessible to a few select laboratories. We have developed a microfluidic, partitioning-based solution that allows sensitive CNV detection of thousands of SCs concomitantly. In this two-step workflow, isolated cells or nuclei are treated to gain complete accessibility to genomic DNA, then co-partitioned along with functionalized gel beads in nanoliter-sized droplets. Barcoded fragments, representing genomic abundance, are then converted into Illumina-compatible libraries. Analysis of our SC whole-genome sequencing data shows excellent coverage uniformity, comparable to leading plate-based SC amplification techniques. We profiled 10,000 single nuclei from five spatially separated sections of a single primary breast tumor. Histopathology analyses reported a 75% tumor purity. We used SC sparse WGS to quantify tumor purity and ploidy across each section and observed an increment in tumor purity from 6% in the first section (S1) to >80% in the fifth section (S5). By aggregating the single-cell data, we inferred pseudo-bulk ploidy profiles ranging from 1.99 to 3.39 across the five sections. To resolve genetically distinct sub-clones in the different tumor sections, we used Euclidean distance measure to identify a group of more than 4000 non-diploid cells from all five samples as cancer cells whose average ploidies ranged from 2.6 to 5.3. Finally, Discriminant Analyses of Principal Component (DAPC) used CNV events to identify 8 major clusters. Spatial clustering of genetically similar single cells suggested a polyclonal model of tumor evolution for this breast. These data provides evidence for the power of scalable SC DNA analyses of CNV for characterizing tumor heterogeneity.

904T

Exploring the architecture of organoid genomes with PromethION technology. S. Goodwin¹, I. Lee², F. Sedlazeck³, R. Sherman², M. Kirche², G. Arun¹, J. Lihn¹, D. Spector¹, M. Schatz², W. Timp², W.R. McCombe¹. 1) Cold Spring Harbor Laboratory, Cold Spring Harbor, NY; 2) Johns Hopkins University, Baltimore, MA; 3) Baylor University, Waco, TX.

Cancers are frequently characterized by structural rearrangements that can confer novel phenotypes, include drug resistance to the cancerous tissue. There is need to accurately characterize a cancers repertoire of structural events in order to provide the most complete insight into tumor pathology. However, recent advances on long read sequencing have shown that short read approaches can miss a significant number of large structural events. Unfortunately the amount of material derived from biopsies is insufficient for comprehensive, long read sequencing. Organoids, a three dimensional cell culture derived from a primary tumor sample, can be grown in the lab and can provide sufficient material for long read sequencing. Despite growing interest in this technology it is unclear what changes to the genome may occur during the organoid generating process, particularly with respect to Structural events. Long read technology, along with short read data can be used to deeply probe the genomic architecture of organoids. The Oxford Nanopore PromethION offers a unique opportunity to explore genome architecture with long reads. The very high throughput and low cost, up to 100Gb per \$2000 chip, allows for greater depth and sample numbers than have been previously possible. This will allow for a deep understanding of the underlying structure of an organoid genome compared to the donor material while also advancing the use of long read sequencing as clinical tool.

905F

Seeing beyond the target: Uncovering germline research cohorts within targeted tumor-only sequencing data. A. Gusev¹, S. Mangul², J. Carrot-Zhang³, M. Ducar¹, L. MacConaill¹, E. Stupka¹, M. Meyerson¹, N. Zaitlen³.

1) Dana-Farber Cancer Institute, Boston, MA; 2) University of California Los Angeles (UCLA), Los Angeles, CA; 3) University of California San Francisco (UCSF), San Francisco, CA.

Tens of thousands of cancer patients have had their tumors sequenced to identify clinically actionable mutations. In addition to saving lives, this activity has produced valuable research data, empowering significant discoveries in translational domains. However, the targeted nature of typical clinical tumor sequencing has a constrained research to pre-defined genes, and the absence of matched normal specimens precludes powerful germline analyses. In this work we develop a platform (SBT: Seeing Beyond the Target) to mine discarded and off-target tumor sequences to uncover rich genome-wide data including common germline genotypes, T and B cell receptor sequences, rDNA and mtDNA copy number, and HLA types. Using data sets with both targeted and whole genome genotypes, we show that our approach can be leveraged for accurate inference of individual-level polygenic risk, genetic ancestry, immune profiles, and can empower cohort-scale germline and germline/somatic GWAS studies. Mapping to non-human genomes further empowered recovery of microbial reads, which occurred in numbers sufficiently large to study the taxonomic composition of microbial communities in the tissue type represented by the sample. Several small-scale studies have demonstrated that these features are potential prognostic indicators, and our methods now make them available in large-scale clinical cohorts. We assess performance across real sequenced tumors from diverse cancer types and sequencing platforms, comparing to gold-standard germline genotyping and whole-genome/RNA-seq. We perform SBT analysis of thousands of tumor panel sequences from the Dana-Farber Cancer Institute *PROFILE* cohort. We examine replication of known variants, germline-somatic interactions, and immune profile associations with cancer outcomes. We believe that SBT will greatly increase the research potential of clinical tumor data sets and provide a bridge between the germline and somatic research communities. The full pipeline is available in an efficient, freely available software platform at <https://github.com/smangul1/seeing.beyond.target/wiki>.

906W

Single-cell transcriptome analysis reveals the gene exchange of cancer cells with high malignant potential and immune cells. S. Hashimoto¹, T. Torigoe², H. Iwabuchi¹.

1) Graduate School of Medicine, Kanazawa University, Kanazawa, Ishikawa-ken, Japan; 2) School of Medicine, Sapporo Medical University, Sapporo, Japan.

We applied novel strategy (Nx1-seq) of single-cell transcriptome analysis to characterize complex heterogeneous samples in the cancer tissue in the myometrial infiltration side (M-side) and endometrial side (E-side) of a human endometrioid adenocarcinoma tissue. Human endometrioid adenocarcinoma tissues are comprised of varying ratios of different cell populations of infiltrating immune cells and cancer cells. Myometrial invasion is an independent prognostic parameter of endometrioid carcinomas and is correlated with the risk of metastasis to the lymph nodes. In addition, it is believed that cells expressing some malignancy-related genes increase at the myoinvasive front. However, in this study, cancer cells in the E-side were highly malignant compared with those in the M-side. Many cells on the E-side were positive for spheroid-specific markers such as SOX2, CTSV and GNG2, which are related to tumorigenesis. In addition to the increased frequency of cells with a cancer stem cell marker in the E-side, the population of EMT in the E-side was higher than in the M-side. Moreover, the population of infiltrating T cells in the tumor, which is a marker for prognosis, was smaller in the E-side compared with the M-side. These data, produced using Nx1-seq, demonstrated that cells with high malignant potential (HMP) were present in the site of the same cancer tissue. Furthermore, when the gene expressions of cancer cells from the E-side and M-side were compared, several genes were identified at sites where cancer cells had HMP. Interestingly, the differential gene expression profile of macrophages and T cells in the E-side and the M-side were similar to those of cancer cells. To confirm this phenomenon, immunostaining of the cancer tissue was carried out. Immunostaining of UCHL1, a highly expressed gene in cancer in the E-side, showed a similar staining pattern in macrophages and T cells as well as cancer cells. In addition, other differential genes were also observed in macrophages and T cells as well as cancer cells. These data showed that many immune cells have the properties (genes) of cancer cells in cancer tissues with HMP. To elucidate this phenomenon, it is very important to clarify the mechanism of new cancer development. Therefore, to achieve more effective cancer therapies, the diversity of cancer cells at each distinct site of tumor tissues must be considered.

907T

Metabolomic and transcriptomic profiling reveals distinct metabolomic patterns and key signalling pathways in tumour tissue of squamous cell carcinoma of the lung. L.T. Hoang¹, E. Starren¹, S. Willis-Owen¹, A. Mandal¹, C. Domingo¹, A. Nastase¹, Y. Zhang¹, K. Lo¹, E. Lim², A. Nicholson³, M. Moffatt^{4,5}, W. Cookson^{1,5}. 1) National Centre for Mesothelioma Research, Imperial College London, UK; 2) Department of Thoracic Surgery, Royal Brompton and Harefield NHS Trust, London, UK; 3) Department of Pathology, Royal Brompton and Harefield NHS Trust, London, UK; 4) The Brunei Cancer Centre, PJSC, Brunei; 5) Corresponding Authors.

An increase in glucose, glutamine and lipid metabolism are hallmarks of cancer cells, including the adenocarcinoma (ADC) subtype of non-small cell lung cancer (NSCLC). However, the metabolomic characteristic of the Squamous Cell Carcinoma (SCC) subtype of NSCLC and the differences between the two subtypes are poorly understood. In this study, we investigated pre-treatment profiles of 395 metabolites (Metabolon) in paired tissue samples (tumour with matched normal lung tissue) for 37 NSCLC patients. Twenty patients had ADC with the remaining seventeen having SCC. Transcriptomes were generated from the same paired samples (extracted RNA of tumour and normal lung tissue pairs) run on Affymetrix Gene 1.1 ST Arrays. The transcriptome data was used for Weighted Gene Co-expression Network Analysis (WGCNA) to identify co-expressed gene modules and to search for correlation with the key metabolites. In general, adenocarcinoma and SCC were found to have a similar metabolomic pattern with the predominance of metabolites involved in glucose, glutamine and lipid metabolism. Interestingly, in contrast to ADC, SCC showed two different metabolomic profiles among the tumours. WGCNA revealed two co-expressed gene modules that were significantly correlated with the production of 5'-CMP (Pyrimidine metabolism and DNA damage repair) and creatine metabolites. Among the canonical pathways that were significantly over-represented in these gene modules, the P53 signalling pathway was significantly up-regulated in SCC compared to ADC. This might contribute to a stronger cellular proliferation processes observed in both metabolomic and transcriptomic data from the SCC patient samples. The distinct metabolomic profiles in the tumour of SCC patients and the interaction between the metabolomic and transcriptomic profiles may contribute to the heterogeneity in clinical and treatment outcomes in NSCLC, with diagnostic and therapeutic potential.

908F

Using drug metabolizing enzymes variations in risk stratification algorithm to guide the dosing decision in Rhabdomyosarcoma pediatric patients. R.M Labib^{1,2}, E ElNadi¹, Hadeer Adel¹, N.H El-Shafei², L.M El-Sharkawy², M Morkos², S.A Abdelshafi¹, D Yassin¹. 1) Children's Cancer Hospital-57357-Egypt, Cairo, Egypt; 2) Department of Clinical Pharmacy, Faculty of Pharmacy, October University for Modern Sciences and Arts (MSA), Egypt.

Background / Objectives: Rhabdomyosarcoma (RMS) is a small round blue cell tumor; representing 7% of all childhood malignancies, and over 50% of all soft tissue sarcomas. Detection of germline variants that determine interindividual variability in pediatric cancer patients' response to chemotherapy is a pathway to more accurate personalized targeted treatment. This trial aims to assess the feasibility of using an algorithm/signature of single nucleotide polymorphism (SNP) to predict response of RMS to chemotherapy. **Design / Methods:** In a cohort of 73 pediatric RMS patients treated with cyclophosphamide based first line treatment, we genotyped multiple candidate SNPs in drug transporters (e.g ABCC3 "MRP3") and drug metabolizing enzymes using restriction fragment polymorphism (RFLP) and taqman pre-designed assays in this pilot. Clinical data on survival, demographics, pathology, chemotherapy dose and clinical response were collected. We examined the association between these genotypes and overall survival, failure free survival (FFS) and achievement of complete response. In phase 2 of this project, we will genotype an array of other less frequent SNPs in DME to assess its frequency and impact in our population. **Results** Blood samples from 73 RMS pediatric patients were collected and genotyped. Patients were enrolled and treated according to IRS-V protocol based on cyclophosphamide. The allelic frequencies of study SNPs were determined. Those SNPs were correlated with all the clinical demographic variables, overall survival and FFS. Carriers of at least one mutant allele CYP2B6 rs2279343 and ABCC3 had significantly longer event-free survival. **Conclusions:** It is possible to transfer genomic results from laboratory to clinical practice in pediatric oncology to support clinical dosing decision. Our preliminary results demonstrated that CYP2B6 c.516T and ABCC3 genotypes may predict FFS in pediatric RMS patients and be used to better stratify risk prior to treatment.

909W

Selective nanopore sequencing of human BRCA1 by Cas9-assisted targeting of chromosome segments (CATCH). Y. Michaeli¹, T. Gabrieli¹, H. Sharim¹, D. Fridman¹, N. Arbib², Y. Eberstein¹. 1) Chemistry, Tel Aviv University, Tel Aviv, Israel; 2) Department of Obstetrics and Gynecology, Meir Hospital, Kfar Saba, Israel.

Next-generation sequencing (NGS) is challenged by structural and copy number variations larger than the typical read length of several hundred bases. Third-generation sequencing platforms such as single-molecule real-time (SMRT) and nanopore sequencing provide longer reads and are able to characterize variations that are undetected in NGS data. Nevertheless, these technologies suffer from inherent low throughput which prohibits deep sequencing at a reasonable cost without target enrichment. Here, we optimized Cas9-Assisted Targeting of CHromosome segments (CATCH) for nanopore sequencing of the breast cancer gene *BRCA1*. A 200 kb target containing the 80 kb *BRCA1* gene body and its flanking regions was isolated intact from primary human peripheral blood cells, allowing long-range amplification and long-read nanopore sequencing. The target was enriched 237-fold and sequenced at up to 70× coverage on a single flow-cell. Overall performance and single-nucleotide polymorphism (SNP) calling were directly compared to Illumina sequencing of the same enriched sample, highlighting the benefits of CATCH for targeted sequencing. The CATCH enrichment scheme only requires knowledge of the target flanking sequence for Cas9 cleavage while providing contiguous data across both coding and non-coding sequence and holds promise for characterization of complex disease-related or highly variable genomic regions.

910T

Broad genomic and transcriptomic profiles of Asian breast cancer. J.W. Pan¹, M. Zabidi¹, P. Ng¹, M.Y. Meng¹, S. Sammut², B. Sandey³, C.H. Yip³, P. Rajadurai³, O.M. Rueda², C. Caldas², S.F. Chin², S.H. Teo¹. 1) Cancer Research Malaysia, Subang Jaya, Malaysia; 2) Cancer Research UK Cambridge Institute, Cambridge, UK; 3) Subang Jaya Medical Centre, Subang Jaya, Malaysia.

Background: Breast cancer is the second largest cause of cancer-related deaths among Asian women and incidence is projected to rise. Asian breast cancer patients are younger and have a higher prevalence of ER-negative disease, but few genomic analyses of somatic mutations and gene expression of Asian breast tumours have hitherto been described. **Methods:** To characterise the molecular profiles of Asian breast tumours, we conducted whole-transcriptome sequencing (RNA-Seq) and paired tumor-normal whole-exome sequencing (WES) of 576 breast tumours from a hospital-based cohort in Malaysia. WES was done on an Illumina HiSeq4000 platform to a mean depth of coverage of 80X/40X (tumor/matched normal), and RNASeq to a depth of 40X (tumor RNA). Additionally, tumor copy number variation was determined using shallow whole-genome sequencing (sWGS) at 0.1X depth of coverage. Somatic variants were identified using Mutect2 and Strelka, molecular subtypes were stratified according to PAM50 and ic10 clusters, and immune scores were generated using ESTIMATE. **Results:** We show that there is an increased prevalence of *TP53* somatic mutations, a higher proportion of aggressive molecular subtypes, and a higher immune score in Asian breast tumours compared to the predominantly Caucasian breast tumours reported in TCGA and METABRIC. **Conclusion:** Our results contribute to an improved molecular understanding of Asian breast tumours, and lay the foundations for enhancing precision medicine initiatives in Asia.

911F

Comprehensive detection of germline and somatic structural mutations in cancer genomes. A. Pang, J. Lee, T. Anantharaman, A. Hastie, H. Cao, M. Borodkin. Bionano Genomics, San Diego, CA.

In cancer genetics, the ability to identify constitutive and low-allelic fraction structural variants (SVs) is crucial. Conventional karyotype and cytogenetics approaches are manually intensive. Microarrays and short-read sequencing cannot detect calls in segmental duplications and repeats, often miss balanced variants, and have trouble finding low-frequency mutations. We describe the use of Bionano Genomics's Saphyr platform to comprehensively identify SVs for studying cancer genomes. DNA larger than 100 kbp are extracted, labelled at specific motifs, and linearized through NanoChannel arrays for visualization. Molecule images are digitized and *de novo* assembled, creating megabases long Bionano genome maps. Somatic mutations can be identified by running the variant annotation pipeline, which compares the cancer sample's assembly calls against > 600,000 SVs in Bionano's control sample SV database, and against a matched control sample's variants. In addition, two new Bionano pipelines leverage these long molecules to identify additional somatic SVs: the copy number variation pipeline and the molecule mapping pipeline. By examining the coverage-depth of molecule alignment to the public human genome reference, the pipeline can identify megabases long amplifications and deletions. Similarly, clusters of split-molecule alignments against a reference can reliably find translocations and other rearrangements. We applied this suite of discovery tools to construct a comprehensive map of SVs in a well-studied melanoma cell line, COLO829. We collected 200X of data from the tumor and the matched blood cell line, constructed contiguous assemblies (N50 > 50 Mbp), and called over 6,000 SVs in each genome. Then, we classified 52 as somatic by comparing the tumor and the blood control. Furthermore, molecule mapping identified extra mutations, bringing the total to 78. We verified a known 12 kbp deletion impacting *PTEN*, and t(1;10) and t(10;18) translocations. Furthermore, the copy number profile captured the duplications of *KIT* on chr6 and *BRAF* on chr7, as well as other chromosomal-arm gains and losses. In conclusion, with one comprehensive platform, Saphyr can discover a broad range of traditionally refractory but functionally-relevant SVs, and further improves our understanding of cancer etiology.

912W

Senescent tumor cells genes expression profiling in papillary thyroid carcinoma. T. Park¹, Y. Kim¹, S. Kang², H. Kim². 1) Biochemistry, Ajou University School of Medicine, Suwon, Korea, Republic of; 2) Pathology, Ajou University School of Medicine, Suwon, Korea, Republic of; 3) Hemato-oncology, Ajou University School of Medicine, Suwon, Korea, Republic of.

Cellular senescence has been perceived as a barrier against carcinogenesis. However, the senescence-associated secretory phenotype (SASP) of senescent cells can promote tumorigenesis. Here, we show senescent tumour cells are frequently present in the front region of collective invasion of papillary thyroid carcinoma (PTC). Although wide variations in SASP expression existed across PTC cases, the expression of interleukin (IL)-6, IL-8 and matrix metalloproteinases (MMPs) was markedly upregulated in senescent tumor cells as compared with levels observed in non-senescent tumor cells. Furthermore, RNA-sequencing analysis revealed upregulation of several chemokines. RNA expression ratios between the non-senescent tumor cells and the senescent tumor cells and found increased expression of several chemokines including CCL4, CCL19, CCL21, CXCL2 or CXCL12 in the senescent tumor cells. These findings suggest that senescent cells are actively involved in cancer microenvironment.

913T

Linking gut microbiome function to tumour molecular pathways using metatranscriptomics. *R. Purcell¹, A. Sulit², F. Frizelle¹, J. Pearson³, S. Schmeier².*

1) Department of Surgery, University of Otago, Christchurch, New Zealand; 2) Institute of Natural and Mathematical Sciences, Massey University, Auckland, New Zealand; 3) Biostatistics and Computational Biology Unit, University of Otago, Christchurch, New Zealand.

Adverse changes in the gut microbiome have been linked with colorectal cancer. However, the majority of studies to date rely on microbiome analysis techniques that purely catalogue the bacterial taxa within samples. Bacterial function is often inferred without taking account of functional redundancy between species. In order to better understand the potential mechanisms of microbial communities within tumours, we have taken a novel approach of interrogating transcriptomic data from colorectal tumours to investigate links between bacterial function and tumour molecular pathways. We accessed a retrospective cohort of ~300 colorectal cancer tumours from our institutional tissue bank and performed transcriptomics to interrogate the expressed RNA in the tumours. Sequencing reads that mapped to the human genome were used as input to a Consensus Molecular Subtype (CMS) classifier, and tumours classified into different subtypes. Tumour molecular pathway analysis was carried out using a variety of pathway and ontology databases including Reactome, Gene Ontology, PANTHER and KEGG. Non-human sequence reads were classified into microbial taxa using a metatranscriptomics approach. Through this approach, we are able to identify bacterial species enriched or depleted in certain colorectal cancer subtypes and, through mapping translated peptide sequences, we are able to investigate the bacterial functional capacity that is active in different subtypes. Through our analysis we find that particular bacterial species are overrepresented in certain subtypes but also we are able to relate tumour microbiome function to molecular mechanisms active within a given tumour sample or tumour subtype, e.g. bacterial lipoprotein expression and host toll-like receptor signaling.

914F

Arthritis initiated by immune-checkpoint inhibitor cancer therapy is a distinct entity from seropositive rheumatoid arthritis. *E.D.O. Roberson^{1,2}, D. Parks¹, A.M. Akk¹, C.T.N. Pham¹.*

1) Dept. of Medicine, Division of Rheumatology, Washington University, St. Louis, MO; 2) Dept. of Genetics, Washington University, St. Louis, MO.

Human cancers avoid apoptosis and immune surveillance by accumulating new mutations. The mutational process is random, but selective pressure purifies the most advantageous mutations. Accumulation of mutations may form neoepitopes that can stimulate T cell activation by antigen presentation. Immune-checkpoint molecules, like PD-1 and CTLA-4, and ligands, like PD-L1, inhibit T cell activation. Tumors that overexpress these molecules are therefore protected from immune attack. Monoclonal antibodies to these immune-checkpoint molecules are effective and durable treatments in a variety of cancers, including melanoma, non-small cell lung cancer, renal cell carcinoma, and Merkel-cell carcinoma. However, some patients experience immune-related adverse events (**irAEs**) that can be mild (eczematous rash) or serious (CMV activation colitis, thyroiditis, neuropathy, and arthritis). irAE arthritis resembles rheumatoid arthritis (**RA**), and can be severe. An unanswered question is whether checkpoint therapy unmasks pre-existing RA or is a separate entity. We approached used the 10X Chromium single-cell RNA-Seq system to compare PBMCs from an irAE patient and an RA patient. We obtained usable sequence from 1,172 irAE patient cells and from 1,654 RA patient cells. We converted the unique molecular identifier counts per cell to transcripts per million, and projected the data into lower dimensions using Uniform Manifold Approximation and Projection. We calculated enrichment of genes in a cluster by comparing to all other clusters, and estimated cell type using the ImmGen. Several clusters had different fraction cells in the irAE and RA samples. One cluster represented 27% of RA cells, but only 12% of irAE cells. The most similar cell type was activated T cells. Another cluster was only 0.4% of RA cells, but 2.6% of irAE cells, and was most similar to a sub-population of neutrophils. We are currently optimizing feature selection for clustering and identifying the underlying cell populations, including adding complementary CyToF data for cells stored from the same blood draw. However, the difference in fraction of cells in each cluster, some clusters only being present in one individual, and differences in expression of genes sometimes within clusters between patients, suggest that irAE arthritis is not identical to seropositive RA. Additional data from seronegative arthritis will be needed in the future to help determine if irAE arthritis is also distinct from it as well.

915W

Characterization of breast cancer samples using integrated genomics and metabolomics data. R.M. Rodrigues-Peres¹, A.M. Porcari^{2,6}, J. Zhang³, K. Garza³, J. Lin³, B.S. Carvalho⁴, I. Lopes-Cendes⁵, M.N. Eberlin⁶, L.S. Eberlin³, L.O.Z. Sarian¹. 1) CAISM – Women's Hospital, State University of Campinas - UNICAMP, SP, Brazil; 2) Laboratory of Multidisciplinary Research, São Francisco University (USF), Bragança Paulista, SP, Brazil; 3) Department of Chemistry, The University of Texas at Austin, TX, USA; 4) Department of Statistics, Institute of Mathematics, Statistics and Scientific Computing, State University of Campinas - UNICAMP, SP, Brazil; 5) Department of Medical Genetics, State University of Campinas - UNICAMP, SP, Brazil; 6) Thomson Mass Spectrometry Laboratory, Department of Chemistry, State University of Campinas - UNICAMP, SP, Brazil.

Introduction: Genomic and metabolomic studies may contribute to elucidation of pathological mechanisms related to development and progression of breast cancers, besides improving the tumor diagnosis. The aggressiveness of breast cancer is due to genetic factors, including the behavior of altered genes and post-translational mechanisms which may affect the way products of these genes are expressed. The objective of this study was to characterize breast cancer samples through the identification of Copy Number Variations (CNVs) related to a number of genes potentially linked to mammary carcinogenesis and to determine the lipid profile of these samples, in order to establish a link that could possibly explain how these features may affect tumor biology, its development and disease prognosis. **Methods:** Eight samples of ductal breast tumors and eight samples of normal breast tissue of the same women, for control, were selected by histopathological characterization of biopsy specimens. Chromosome Microarray (CMA) technique was then performed to access CNVs from these tumors and Desorption Electrospray Ionization Mass Spectrometry Imaging (DESI-MSI) technique was applied to verify the abundance of lipids on breast tissue. Relations between the data assessed through these analyses were obtained by Metaboanalyst 3.0 tools for metabolic pathways search. **Results:** After data refinement, CMA identified 780 significant regions containing CNVs, which related to 1199 genes. DESI-MSI identified 4 classes of lipids which were significantly more abundant in tumor tissues when compared to normal tissues: glycerophosphoserine, glycerophosphoinositol, glycerophosphoglycerol and glycerophosphoethanolamine. When these genes and lipids were related, the following metabolic pathways were found to be altered: arachidonic acid metabolism ($p = 0.000361$), ether lipid metabolism ($p = 0.00223$), metabolism of xenobiotics by cytochrome P450 ($p = 0.00674$), biotin metabolism ($p = 0.0445$) and one carbon pool by folate ($p = 0.0489$). **Conclusion:** These findings show that amplifications and deletions of key genes can affect the lipid profile in breast tumors and that the described lipids are sufficient to determine whether a tissue is tumoral or not. These processes should continue to be studied for well understanding of how these interactions are influencing the cellular pathways that may contribute to the onset and progress of the breast tumor, aiming to improve the management of the disease.

916T

Novel targeted long read approaches for uncovering undetected mutations in the BRCA1 and BRCA2 genes. J.A. Rosenfeld¹, F. Sedlazeck², S. Goodwin³, M. Delabastide³, Y. Chen¹, S. Iyer³, C. Boles⁴, S. Ganesan¹. 1) Cancer Institute of New Jersey, Rutgers University, New Brunswick, NJ; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 3) Cold Spring Harbor Laboratory, Cold Spring Harbor, NY; 4) Sage Science, Beverly, MA.

Mutations in the BRCA1 and BRCA2 genes are important drivers of many cancers particularly those of the Breast and the Ovary. Because of this cancer linkage, there are screening programs whereby a woman with familial cancer history has these genes sequenced to determine her cancer risk. Historically, this sequencing was done using Sanger sequencing, but more recently, short read Illumina sequencing has become the standard. While these Illumina reads are cheap and allow for the screening of mutations, their short length of 100bp has limited their efficacy in detecting translocations as well as SNPs and short indels. The reason for the lack of detection of SNPs and short indels is due to the difficulty of mapping short reads in regions of the genome that are even mildly repetitive. In order to address these gaps in our understanding of BRCA1 and BRCA2, we have used innovative approaches for targeted long read sequencing of these loci. Whole genome long-read sequencing is cost-prohibitive and traditional capture techniques are insufficient for long-read sequencing. We have used the Sage Science CATCH+ approach that couples Cas9 genome cutting and pulse-field gel electrophoresis to capture DNA for the BRCA loci, achieving ~30X enrichment of the BRCA1 loci. In addition, we have used custom enrichment methods from Swift Biosciences that target 2kb regions of the BRCA genes, achieving more than 100X enrichment of the loci. As a first test, we have used the highly characterized breast cancer cell lines including SKBR3. Once the techniques are validated, we plan to extend our approach to other cell lines and patient samples. Through these approaches, we have identified a genome translocation between chromosomes 13 and 17 and multiple SNPs that were not detected with the traditional Illumina short read approaches. These variants will lead to greater knowledge and understanding of the BRCA genes and their connection to cancers.

917F

Deep targeted duplex sequencing for low allele fraction variant detection. *M. DeFelice, B. Blumenstiel, J. Grimsby, M. Fleharty, C. Cibulskis, V. Adalsteinsson, M. Duran, J. Abreu, T. Desmet, S. Gabriel, N. Lennon.* The Broad Institute, Cambridge, MA.

Low allele fraction variant detection from liquid biopsies and other challenging biological sample types poses a unique set of challenges for a large scale research and clinical sequencing lab. With recent advances in DNA targeting and enrichment methods coupled with reduced sequencing costs, there is an increasing demand for sequencing labs to support a variety of deep targeted panels for both cancer and germline research. Specifically, there is demand for developing distinct, disease-specific custom panels and generating extremely deep coverage for detecting low allele fraction variants with high sensitivity from biological samples such as plasma derived cell-free DNA, urine, cerebrospinal fluid, and bone marrow aspirates. Often these research or clinical cohorts number in the 100's of samples rather than 1,000's requiring low per-sample project initiation costs. Historically, targeted gene panels were costly to develop and often economically feasible only when upfront costs were amortized over many thousands of samples. Today, we are able to provide a low cost, rapidly implemented panel service, compatible with a variety of sample types, targets and project size. To achieve this, we have developed a unified library construction and target capture workflow which incorporates duplex unique molecular indices (UMIs) and built a UMI-aware consensus calling pipeline. By integrating balanced array based probe synthesis, short duplex UMI adapters, and a modified single hybridization workflow, we have been able to obtain deep on-target coverage (>90% Selected) with balanced uniformity (Fold-80 <1.5) and rapid turnaround (~3 weeks) for custom panels ranging from 300kb to 2mb in size. UMI duplex adapters on all libraries allow for robust error correction and high sensitivity for low allele fractions (<1%) as well as a low false positive rate (<1FP/MB). This workflow will allow for comprehensive genomic profiling and will be the basis of research and clinically validated assays over the next years.

918W

Use of molecular identifiers and targeted NGS to enable variant detection below 1% allele frequencies in circulating cell-free DNA. *A. Wood, S. Sandhu, J. RoseFigura, L. Kurihara, V. Makarov.* Swift Biosciences Ann Arbor, MI.

The growing use of circulating cell-free (cfDNA) necessitates accurate variant detection at less than 1% allele frequencies due to a low population of DNA of interest within total cfDNA. This level of detection is critical for a variety of cfDNA applications including detection of circulating tumor DNA (ctDNA), detection of graft-derived cfDNA (GcfDNA) within normal plasma as a marker of graft integrity after organ transplantation, and for noninvasive prenatal testing (NIPT) where fetal cfDNA is detected at low frequencies in maternal plasma. However, reliable low-frequency variant detection by next-generation sequencing (NGS) is challenging due to background noise from PCR and sequencing errors. Additionally, while NGS is well suited for high throughput analysis, whole-genome sequencing can be cost prohibitive. To overcome these challenges, we employed a targeted NGS assay that uses molecular identifiers (MIDs) to uniquely label individual DNA molecules prior to amplification. MID families consisting of PCR duplicates can be used to distinguish true variants from PCR and sequencing errors and an error-corrected consensus sequence can be generated. We incorporated MIDs into a 17 amplicon EGFR pathway panel that can be used for ctDNA detection in lung cancer and a 101 amplicon SNP panel that contains 92 amplicons targeting SNPs with high minor allele frequency and 9 amplicons that identify gender ideal for GcfDNA or fetal cfDNA detection. These assays generate an NGS library using a single-tube multiplex PCR that is compatible with cfDNA. To test assay sensitivity, we performed low frequency spike-in experiments with control DNA or patient derived cfDNA. We prepared MID libraries from 10-50ng of input DNA and deep sequencing to >20,000x was done to maximize MID family size and optimize generation of a consensus sequence. This analysis accurately and reproducibly identified known variants as low as 0.1%. Importantly, the use of MIDs drastically reduced the number of false positive variants called at less than 1% increasing confidence in low frequency variant calls. Accurate variant calling at these levels is critical to track biologically relevant levels of ctDNA for early detection and monitoring of cancer, GcfDNA to screen for organ rejection after transplantation, and fetal DNA at early stages of pregnancy. This study highlights the power of MID technology in amplicon NGS library preparation to enable low frequency variant detection for cfDNA analysis.

919T

Linked read technology enables disambiguation of structural variants to *PMS2* and *PMS2CL*. J. Van den Akker, J. Lock, A. Ondov, A. Khalif, Y. Xu, L. Servais, Z. Mahkovec, A.Y. Zhou, S. Topper. Color Genomics, Inc., Burlingame, CA.

BACKGROUND: Pathogenic variants in *PMS2* account for 5-15% of Lynch syndrome, increasing the risk for colon, endometrial and other cancers. However, variant detection in the 3' region of *PMS2* is complicated by the presence of a pseudogene (*PMS2CL*), which shares high sequence homology with *PMS2* exons 12-15. Moreover, sequence exchange between *PMS2* and *PMS2CL*, which prevents analysis relying on minor sequence differences, has been observed in 33-69% of unselected cohorts (PMID 20186688, 20186689). Traditionally, sequencing the 3' end of *PMS2* has relied on a long-range PCR primed from a binding site outside the homologous region. Successful amplification of this >17kb amplicon requires high molecular weight gDNA, which is challenging for biobanked specimens and common sample types such as buccal or saliva of variable quality. In addition, this approach fails if the breakpoints of a structural variant (SV) are unknown, and for complex rearrangements such as gene conversion or reciprocal crossover. Here, we present how these limitations can be overcome with linked-read technology, which reconstructs long DNA sequences from short barcoded fragments. **METHODS:** Genomic DNA was extracted from saliva or blood, and molecular barcodes were incorporated into individual DNA molecules via isothermal amplification using Gel-Beads in Emulsion technology (10x Genomics). These fragments were used as input into the Kapa HyperPrep protocol, followed by target enrichment using a custom SureSelect XT bait set targeting the *PMS2* and *PMS2CL* regions. After sequencing on the NextSeq 550, FASTQ files were analyzed with Long Ranger v2.2.2, supplemented with CNVkit for detection of deletions/duplications by differential coverage. **RESULTS:** A set of >20 SNVs and indels, as well as >10 SVs contained within exons 12-15, were correctly identified in *PMS2* or *PMS2CL*, congruent with analysis by Ir-PCR. Additionally, breakpoints were resolved for 5 novel SVs overlapping exons 12-15. Furthermore, this platform allowed identification of complex rearrangements such as insertion of an SVA element, an inversion including exon 1, a processed pseudogene, and gene conversion. **CONCLUSION:** The data presented here validates the use of 10x linked-read technology to reliably identify variants in *PMS2*. This approach is superior to the traditional long-range PCR approach because it is effective on DNA of lower quantity (<2ng) and lower quality, and can resolve complex rearrangements impacting the 3' end of *PMS2*.

920F

A multi-enzyme DNA repair mix improves library quality and sequencing accuracy. M.R. Heider, L. Chen, C. Song, L. Sun, P. Liu, L. Ettwiller, L. Higgins, E.T. Dimalanta, T.B. Davis, T.C. Evans. New England Biolabs, Inc., Ipswich, MA.

High-throughput sequencing is a key tool for clinical diagnostic applications such as mutation profiling in diseased tissues and long read, genome assembly methods for identifying pathogens with high taxonomic specificity. Preparation of high quality next-generation sequencing (NGS) libraries is critical for obtaining accurate patient data, but clinical scientists are often limited by the input DNA quality. In cancer genomics, a common source of DNA is formalin-fixed, paraffin-embedded (FFPE) tissue from patient biopsy. FFPE DNA poses many challenges for preparing NGS libraries, including low input amounts and extensive fixation- and storage-induced DNA damage. Additionally, these damage-induced sequencing artifacts raise the background level of mutations, making it difficult to discern true, low frequency, disease-related variants from noise. Even relatively high-quality DNA inputs contain nicks and gaps limiting library yields and read lengths for long read sequencing platforms including Oxford Nanopore Technologies (ONT) and Pacific Biosciences (PacBio). We developed a second-generation DNA repair enzyme mix (V2) that efficiently repairs the most prevalent damage types found in FFPE DNA and further improves the quality and yield of NGS libraries through efficient nick and gap repair. We tested the application of the V2 repair mix in both Illumina and ONT library preparation. We prepared Illumina libraries with and without FFPE DNA Repair V2 for target enrichment studies. Following this, we performed deep sequencing, variant calling analyses, and validated the mutation frequencies using droplet digital PCR (ddPCR). For ONT libraries, we evaluated the benefit of repairing different DNA types prior to library preparation and sequencing. Illumina library yield and quality were improved when DNA samples were treated with FFPE DNA Repair V2 before library preparation. Target enrichment and whole exome enrichment experiments showed that the V2 repair mix did not alter the overall frequency of variants identified, thus it did not introduce bias, but significantly improved the sequencing accuracy by reducing the number of false variant calls. Repaired ONT libraries also showed improved library quality as well as greater read lengths. Enzymatic DNA repair prior to NGS library preparation is a critical first step both for eliminating damage-induced sequencing artifacts in Illumina sequencing as well as improving library quality and read lengths for ONT and PacBio.

921W

Identification and correction of sample specific sequencing bias for improved copy number variant detection in fixed-cell pellet samples.

J.B. Smadbeck¹, B.A. Piteł², S.A. Smoley², S.H. Johnson¹, N.L. Hoppman², G. Vasmataz¹. 1) BioMarker Discovery Group, Mayo Clinic, Rochester, MN; 2) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN.

As whole-genome next-generation sequencing (NGS) methods are used more widely in clinical practice, there will be a continued push to improve the analytical methods to account for sample quality. The quality of a sample significantly affects analytical results and is often correlated to DNA degradation resulting from the method of sample preservation. Part of this degradation manifests in sequencing bias, whereby different regions of the genome may be over or under sequenced in comparison to the expected baseline level. This makes even a basic whole-genome calculation like copy number variant (CNV) detection prone to false positives. False positives are particularly pernicious in the use of these tests in the clinical space. For this reason it is important to understand the sources of the sequencing bias and have the ability to correct it. One type of preservation method of interest for use in whole-genome NGS analysis is fixed-cell pellet (FCP) preservation, in which cells are preserved in a solution of methanol and glacial acetic acid after being used to prepare slides for conventional chromosome studies or fluorescent in situ hybridization (FISH). We have found that these samples can be effectively used for the detection of breakpoint junctions using the whole-genome mate pair sequencing (MPseq) technology. However, in copy number variant detection it was found that these samples produced a high false positive detection when compared to fresh and frozen samples. In this study we aimed to identify and correct the sources of sequencing bias in FCP samples. We took 12 normal blood and bone marrow samples preserved as fixed-cell pellets and compared their sequencing bias to a set of 40 fresh or frozen normal blood and bone marrow samples. While a typical fresh or frozen sample will have sequencing bias correlated almost entirely with the GC-content, we found that the FCP samples were influenced by properties independent of GC-content. We demonstrate that this bias may be linked to the presence of short interspersed nuclear elements (SINEs). This source and significance of this bias is unique to the FCP samples. We developed an algorithm to correct this SINE bias and showed how this correction aids in the reduction of false positives for FCP samples. Through a full understanding of a sample's bias, a researcher can predict false positive results and make corrections to more accurately analyze these samples with NGS technologies.

922T

The NantOmics Pharmacogenomics Test: Multi-omic screening of 2,489 oncology patients. *C. Schwartz¹, J. Little¹, C.J. Vaske¹, C.W. Szeto¹, S.C. Benz¹, S. Rabizadeh², J.Z. Sanborn¹.* 1) NantOmics, Santa Cruz, CA; 2) NantOmics, Culver City, CA.

Numerous oncology drugs have pharmacogenomics warnings on their FDA labels, yet pharmacogenomics screening is not routinely applied in clinical practice. Pharmacogenomics testing is used to reduce the chance of drug-induced toxicities, improve patient outcomes, and reduce treatment costs by tailoring therapies to the patient's genotype. Here we present the NantOmics Pharmacogenomics test, which screens for pharmacogenomics variants related to 19 gene-drug pairs with CPIC guidelines and FDA label indications. Pharmacogenomics screening was performed on whole genome and whole exome sequencing (WES) data of FFPE tumors and matched normals from 2,489 oncology patients. Patients were screened using a panel of 31 germline markers in 10 genes linked to toxicities from 15 cancer therapies. The test has been validated on 10 cell lines from the CDC GeT-RM, on a set of synthetic data, as well as on a cohort of patients previously genotyped by an independent CLIA-validated PCR-based panel. For variants not detectable by WES (intronic and intergenic variants), we investigate exonic variants in linkage disequilibrium to act as proxies to predict toxicity in these patients. We also investigate gene expression variants implicated in toxicity of oncology drugs in patients with RNA sequencing data available. Of the 2,489 patients screened, 96.3% contained a variant with a pharmacogenomics recommendation. Furthermore, 7.4% of patients had genomic variants associated with severe or life-threatening drug toxicities. For all alleles in our clinical panel, we observed similar allele frequencies to those reported in the ExAC database. In all validation studies, we were able to demonstrate that the test detects each variant in our panel, and correctly determines patient genotype in all studied cases. Given the high percentage of patients with potentially treatment-altering genomic variants, these results underscore the need for more routine pharmacogenomics screening in the oncological setting.

923F

High throughput, reduction transcriptomics, using ultraplex RNA sequencing. *M. Kreutz, J. Shaffer, Q. Jiang, S. Rulli, E. Lader.* QIAGEN Sciences, 6951 Executive Way, Frederick MD.

Elucidating molecular mechanisms and understanding genetic heterogeneity in cancer biology can be achieved through high throughput gene expression analysis and pathway assignment. Recent advances in RNAseq methodologies have enabled accurate gene expression profiling from single cells, but analyzing hundreds or thousands of cells is an arduous exercise in library prep. Here we describe a high throughput 3' RNAseq library prep methodology termed Ultrplex. With QIAseq Ultrplex, Reverse transcription is performed directly on lysed cells, while simultaneously assigning a unique 'cell index' to all cDNA synthesized from a cell. All subsequent transcriptome or targeted panel library steps that follow occur in a single pooled sample; with up to 384 samples assigned per sequencing index. Each 384-plex library is assigned a standard sample index, such that up to 384 pooled 384X libraries can be sequenced together. With this methodology, thousands of transcriptomes or targeted RNA panels can be prepared and sequenced together. Here, we analyze the heterogeneity of an ostensibly uniform population of cancer cells, identifying unique signatures of genes that drive cellular clustering. With QIAseq UPX, high throughput transcriptomic analysis enables the identification and characterization of gene signatures which divide cells into discrete sub-populations based on vectorial gene expression components.

924W

A comprehensive cross-platform comparison of single-cell RNA sequencing using reference samples. *C. Wang¹, W. Xiao², W. Chen¹, Sequencing Quality Control (SEQC)-2 Consortium Single-cell Project Team.* 1) Center for Genomics, Loma Linda University, Loma Linda, CA; 2) National Center for Toxicological Research, FDA, Jefferson, AR.

Rapidly developing single-cell RNA sequencing (scRNA-seq) technologies allow interrogation of the transcriptome in unprecedented detail, but questions arise as to how accurate and reproducible different platforms are; and there is lack of a benchmarking and validation using the standard reference samples. So far, there is no study that has carried out a systemic cross-platform comparison and evaluation on different scRNA-seq technologies using a standard reference sample, in which a large amount of whole genome, whole exome and whole epigenome data are also available to validate the scRNA-seq data. **Under the umbrella of the 2nd phase of the Sequencing Quality Control (SEQC-2) Consortium, we exploited a comprehensive study design** including four scRNA-seq platforms, i.e., 10X Genomics Chromium, Fluidigm C1, Fluidigm C1 HT and WaferGen, across five testing sites, using two well-characterized reference cell lines, i.e., human breast cancer cell line with a matched normal control cell line derived from B lymphocytes of the same patient. In addition to the scRNA-seq transcriptomic studies, we also investigate the epigenome including DNA methylome obtained using Illumina TruSeq Epic DNA-seq and chromatin accessibility by ATAC-seq at population cell level from the reference samples. We further interrogate the whole-genome and whole-exome of the reference samples at population cell level with deep sequencing, which would allow us to evaluate the accuracy of scRNA-seq. **Our objectives are:** 1) to establish a benchmarking and QC metrics for scRNA-seq; 2) to evaluate the accuracy and reproducibility of scRNA-seq across different platforms; 3) to evaluate the inter-lab reproducibility of scRNA-seq; 4) interrogate the epigenomic regulations on single-cell transcriptome in terms of the consistency and the biological discovery across different scRNA-seq technologies. **Conclusions:** Our analyses indicate that both the technical and biological batch effects exist and the reproducibility of scRNA-seq across platforms and sites is dictated by both genes selected and bioinformatics pipelines used. We optimized a bioinformatics normalization method and pipeline to correct the batch effects to achieve the best practice for scRNA-seq analyses. Furthermore, validated by both DNA methylome and ATAC-seq results, we identified a list of genes for benchmarking scRNA-seq performances across different platforms and testing site.

925T

Enzymatic DNA fragmentation for next generation DNA sequencing sample preparation from FFPE tissue biopsy samples. *M. Borns¹, B. Rogers¹, K. Chen¹, G. Amparo², C. Pabon², L. Niu², B. Arezi¹, H. Hogrefe¹.* 1) Agilent Technologies, Genomics Solutions Division, La Jolla, CA; 2) Agilent Technologies, Genomics Solutions Division, Santa Clara, CA.

Mechanical DNA shearing is the most common method for fragmenting DNA for library construction. Current methods for mechanical shearing are difficult, labor intensive, time consuming, expensive, and therefore unsuitable for clinical customers. Because of these drawbacks, enzymatic shearing methods are becoming more popular. However, the enzymatic shearing methods available so far, require additional optimization for different input amounts or when used with samples of varying quality. Here, we address these challenges by developing an endonuclease-based enzymatic double stranded-DNA fragmentation method which generates random DNA fragments from a variety of fresh frozen tissue, blood, and FFPE samples of varying quality and input (in the range of 10-200ng) with a single protocol (i.e., no optimization other than adjusting the PCR cycles). The DNA fragments generated have 5'-phosphate and 3'-hydroxyl termini and are directly used in end-repair/dA tailing without further purification. It is a simple, rapid, streamlined single-tube process. Unlike transposase shearing/tagmentation, it is suitable for library preparation from any FFPE tissue derived DNA including highly-damaged FFPE samples. As we will demonstrate, fragmentation and library preparation experiments using three bacteria of varying GC content (30%-67%) and human DNA shows high coverage uniformity and minimal bias across all GC content. Furthermore, we observe more complex libraries (~50%) across samples of medium to low quality (TapeStation DIN <3) compared to mechanical shearing. This is likely due to lower loss of DNA during the transfer step (i.e., single tube library prep) and/or the presence of more DNA ends that are repairable (as opposed to mechanical shearing). Higher library complexity increases coverage at higher sequencing depth which helps improve the detection of low allele frequency variants (as low as 1%).

926F

AmpliSeq HD: A rapid, custom NGS library prep method that enables detection of mutations present at 0.1% allelic frequency. *G. Luo, X. Peng, D. Mazur, S. Chen, K. Park, K. Heinemann, B. Ding, A. Hatch, Y. Fu, C. Chin, A. Kraltcheva, M. Andersen.* Clinical Next-Generation Sequencing Division, ThermoFisher Scientific, Carlsbad, CA.

Analysis of circulating cell-free DNA (cfDNA) in the blood has great potential in many clinical applications such as early cancer detection and peripheral monitoring. However, only a small fraction, as low as 0.1%, of the cfDNA sample will carry mutations from cancer cells of the patient. Very high sequencing depth focused on certain regions in the genome is required to capture such rare mutations. For this purpose, a highly sensitive, customizable targeted-sequencing approach will provide a powerful tool for clinical cancer studies to detect different biomarkers. We've developed a customizable library prep method (AmpliSeq HD) that incorporates molecular tags into multiplex PCR. The molecular tags are critical in removing random sequencing errors that would otherwise increase noise and decrease the ability to identify low frequency mutations. An engineered high fidelity polymerase is used here to ensure faithful copy of the original DNA sequence. We introduced a new cocktail of enzymes to enable an AmpliSeq like clean up step to remove primer-dimers produced in multiplex PCR. This digestion is a crucial step in the workflow to enable custom designed panel. We also optimized amplification condition to minimize false positive from cytosine deamination. With all these improvements, the AmpliSeq HD technique will allow rapid targeted library preparation in less than 3hrs with sensitivity to detect variances at 0.1% level in cfDNA or 0.5% level in FFPE DNA samples.

927W

Genomic analysis of somatic and germline mutations in malignant pleural mesothelioma.

A. Nastase^{1,8}, A. Mandal^{1,8}, S.K. Lu¹, H. Anbunathan¹, S. Gennatas^{1,5}, D. Morris-Rosendahl^{1,2}, M. Edwards^{1,2}, R.C. Rintoul³, E. Lim⁴, S. Popat⁵, M. Lathrop⁶, A. Nicholson⁷, A.M. Bowcock^{1,8}, M.F. Moffatt^{1,8}, WOC. Cookson^{1,9}. 1) National Centre for Mesothelioma Research, Imperial College London, United Kingdom; 2) Clinical Genetics and Genomics Royal Brompton and Harefield NHS Foundation Trust, London, United Kingdom; 3) Department of Thoracic Oncology, Papworth Hospital, Cambridge, United Kingdom; 4) Department of Thoracic Surgery, Royal Brompton and Harefield NHS Trust, London, United Kingdom; 5) Department of Medicine, Royal Marsden Hospital, London, UK; 6) Department of Human Genetics, McGill University-Génome Québec Innovation Centre, Québec, Canada; 7) Department of Pathology, Royal Brompton and Harefield NHS Trust, London, United Kingdom; 8) Contributed equally to the study; 9) Corresponding authors.

The genomic landscape of malignant pleural mesothelioma (MPM) is characterised by low mutational burden. Known driver events are in tumour suppressor genes but not in known oncogenes. A thorough interrogation of somatic mutations and germline risk factors may identify new genetic drivers that lead to onset and progression of the disease. This could lead to new therapeutic options as opposed to the current standard of care, chemotherapy, which is of limited benefit in relation to survival. We characterised the mutational profile of MPM with whole exome sequencing (WES) followed by target capture sequencing of DNA isolated from fresh snap frozen tissues. Samples were obtained from Mesobank UK, the BLF funded Mick Knighton Mesothelioma Tissue Bank, Respiratory BRU Biobank and Diagnostic Archive, Royal Brompton Hospital and from an Imperial College London sponsored, prospective study on thoracic cancers undergoing biopsy at St George's Hospital. Samples used for WES, comprised 21 paired (blood and tumour tissues) and 29 un-paired samples from MPM patients. Follow up sequencing was performed on a panel of 57 genes interrogated in samples from an additional 69 patients and the original cohort used for WES. The mean age of the patients was 72.1 (±8.5) years old and the ratio of male to female patients was 5.4:1. The majority of the patients (64.7%) had the epithelioid sub-type of MPM. The somatic mutation spectrum showed recurrent alterations in well known drivers of MPM: *BAP1* (37.7%), *NF2* (23.4%), *SETD2* (9.9%), *TP53* (9.1%), *SETDB1* (7.8%), *SYNE1* (6.5%), as well as genes with rare mutations in MPM: *NCOR2* (5.2%) and *MSH6* (2.6%). In addition, hotspot *NRAS* somatic mutations were identified in two patients with the sarcomatoid sub-type (G12V and Q61H respectively). We also observe germline risk variants in *BRCA2* (frameshift and stop-gained) and one case of *BAP1*. We report identification of mutations in genes that are candidate therapeutic targets. Observation of *NRAS* hotspot mutations in patients with no other known driver mutations hypothetically indicate RAF-MEK-ERK pathway being targeted in at least a fraction of MPM cases. Similarly the *BRCA2* and *MSH6* loss-of-function mutations could indicate a tumour genome with distinct vulnerabilities in the DNA damage response pathway, with potential for neoantigen formation. The implications of these alterations are being investigated further with cell line model systems queried through high-throughput genomic platforms.

928T

Discovering diverse genetic biomarkers using cancer-specific targeted NGS.

J. RoseFigura¹, S. Sandhu¹, K. Nakama¹, B. Reed¹, J. Irish¹, P. Dottino², O. Camacho-Vanegas², L. Kurihara¹, J. Martignetti², V. Makarov¹. 1) Swift Biosciences, Ann Arbor, MI; 2) Icahn School of Medicine at Mt Sinai, Departments of Genetics and Genomic Sciences, Pediatrics, Obstetrics/Gynecology & Reproductive Sciences and Oncological Sciences 1425 Madison Avenue New York, NY.

In the rapidly evolving cancer field, focused assays for different biomarkers are increasingly important tools. We have developed targeted NGS panels consisting of disease-specific, predesigned content that allow the flexibility to assess SNVs, small indels, and CNVs in a multiplexed single tube assay. Herein, we demonstrate the use of these panels to assess these variant types in validated standards and combine multiple pre-existing panels to evaluate clinical samples from patients with a known disease. The panels used in this study consist of a Pan-cancer panel, a BRCA1/BRCA2 panel, and panels targeting specific cancers, including Lung, Colorectal (CRC), Lynch, Myeloid, and Breast/Ovarian. These panels are based on a Swift's Accel-Amplicon technology, which generates overlapping amplicons for exon-level coverage of contiguous regions in a single tube assay. CNV standards were acquired for *ERBB2* amplification and for *EGFR* and *MET* amplification. The program CNVkit (Talevich et al., 2014) was used to assess copy number. Cell-free DNA (cfDNA) and peripheral blood DNA was acquired from healthy donors. Clinical samples from known Lynch patients were provided by the Martignetti lab at the Icahn School of Medicine at Mount Sinai for proof of concept experiments. The targeted NGS panels demonstrated consistent performance with a wide variety of DNA types, including cfDNA and FFPE DNA. All panels exhibited >90% on target and >90% coverage uniformity. CNV results agreed with DNA standards at copy numbers ratios (>2.9). Six clinical samples, with two corresponding tumors, were initially evaluated using the Pan-Can panel. These results were indicative of Lynch syndrome, so a combined CRC/Lynch panel was also run on these samples to further probe the disease state of the patient. Targeted NGS panels are a powerful tool for assessing the genetic etiology of multiple disease states. We have demonstrated the ability of this technology to identify important disease biomarkers in clinical samples, and to assess CNVs in pre-validated standards. The ability to assess SNV, indels, and CNV in a single tube from 10 ng of sample represents an important tool for assessing diseases in the clinic, as well as in research and discovery settings.

929F

Ultra-multiplexed CleanPlex® NGS panel: Comprehensive tumor profiling using a fast, single-tube protocol. *L.Y. Liu, Z. Liu, L. Lee, G. Liu, J. Liu, L.J. Zhang, L. Lin.* Paragon Genomics, Hayward, CA.

Next-generation sequencing (NGS) has become a key tool for screening large numbers of genes and cancer mutations because of its ability to simultaneously sequence a large number of targets. Tumor profiling requires the sequencing of a large number of genes. To most effectively use the paralleling power of NGS, multiplexed targeted sequencing can provide highly comprehensive diagnostic methods for tumor profiling and patient treatment identification. Unique challenges for tumor profiling call for a highly efficient, accurate and robust method for the unbiased enrichment of tens of thousands of target regions. However, an increase in multiplexity can also increase the risk of non-specific primer-primer interactions, decrease uniformity, and increase GC bias. Here we present a highly-multiplexed comprehensive tumor profiling panel containing 19,700+ regions, covering 364 genes. Using the CleanPlex® PCR-based, one-tube method, and an average read depth of ~1000 reads per amplicon, this highly multiplexed amplicon panel exhibits 92% uniformity at 0.2X mean bias, little GC bias, and mutation calling at as low as 1% allele frequency. This panel is compatible with genomics DNA, FFPE DNA or cfDNA, and can be used for monitoring gene mutations or tumor burden. CleanPlex® technology demonstrates that it is capable of creating high quality amplicon libraries with high uniformity, low GC bias, and sensitive variant calling even in ultra-multiplexed libraries with ~20,000 amplicons with a workflow under 3-hours.

930W

Genetic and epigenetic evidence for clonal tumor growth in sporadic parathyroid adenomas. *K. Brewer¹, R. Jenkins¹, J. Costa-Guda², A. Arnold¹.*

1) Center for Molecular Oncology, University of Connecticut School of Medicine, Farmington, CT; 2) Center for Regenerative Medicine and Skeletal Development, University of Connecticut School of Dental Medicine, Farmington, CT.

Introduction: Parathyroid adenomas (PTAs) are benign lesions responsible for the large majority of non-familial cases of primary hyperparathyroidism, and most have long been understood to be clonal outgrowths of a single tumor progenitor cell. However, some recent studies of PTA clonality using X-chromosome inactivation-based assays concluded that a substantial proportion of PTAs are not clonal and may derive from multiple cells. Because of well-recognized pitfalls in interpreting non-clonal (polyclonal) results in X-inactivation assays, we reexamined the clonality of a cohort of sporadic PTAs emphasizing methods that directly detect clonal DNA alterations. **Methods:** Among an original cohort of 83 sporadic, typical, solitary PTAs, 57 (69%) had previously proved to be unequivocally clonal using methods such as loss of heterozygosity at selected chromosome markers, classic comparative genomic hybridization, single nucleotide polymorphism array, and/or candidate gene sequencing. In the present study, we were able to interrogate 24 of the remaining 26 PTAs with still-undetermined clonal status, seeking (1) direct genetic evidence of clonal copy number variation (CNV) using genome-wide high-resolution CytoScan HD array; and (2) indirect evidence of clonality by assaying the consistency of allelic X-inactivation via human androgen receptor assay (HUMARA) on female samples. **Results:** In all 24 tumor samples of unknown clonal status, we identified tumor-specific CNVs compared with the CytoScan pooled reference sample, demonstrating that each of the tumors assayed is clonal. This result was confirmed in a validation subset of individually matched tumor and germline samples. By HUMARA, 14 of 19 (74%) female tumor samples were informative and exhibited a clonal methylation pattern; the remaining 5 yielded inconclusive results in which there were no clear non-clonal (polyclonal) patterns. **Conclusions:** The use of high-resolution methodology to identify changes in the tumor genome offers compelling evidence of the clonality of solitary sporadic PTAs and is informative in cases where X-inactivation studies are nondefinitive, such as in male or uninformative female cases. All 24 tumors interrogated in this study exhibited genetic and, in most cases, epigenetic patterns consistent with clonal outgrowth. In combination with previous data, we now have evidence that a minimum of 81 tumors of the original 83-tumor cohort (98%) are clonal.

931T

Functional mechanisms underlying the association between novel variants, their target genes and prostate cancer risk. C. Cieza-Borrella¹, S. Wakerell¹, Q. Karlsson¹, S. Merson¹, R. Rageevakumar¹, I. Whitmore¹, E. Anokian¹, E. Saunders¹, R.A. Eeles^{1,2}, Z. Kote-Jarai¹. 1) The Institute of Cancer Research, London, United Kingdom; 2) The Royal Marsden Hospital NHS Trust, London, UK.

Prostate Cancer (PCa) is the most frequently diagnosed cancer among men in developed countries and one of the most heritable solid tumours. The effectiveness of PSA screening remains controversial and new molecular biomarkers are needed to improve screening and clinical management of disease. Genome-Wide Association Studies (GWAS) and consequent fine-mapping analysis have identified over 170 low-penetrance PCa susceptibility variants associated with PCa risk; however the mechanisms through which these affect disease risk are still not fully understood. The aim of this study is to functionally characterise candidate SNPs at three finemapped GWAS loci in the DNA repair genes *RAD51B* and *RAD23B* as well as the novel PCa candidate gene *NGEP-L* (*ANO7*), in order to elucidate the biological mechanisms behind their association with PCa risk. Circular Chromosome Conformation Capture (4C) was used in five prostate human cell lines (LNCaP, DuCaP, PC3, PNT1a and RWPE-1) to investigate whole genome interactions between the candidate functional risk SNPs rs7141529 and rs767127 in *RAD51B*, and four candidate European variants in *RAD23B*; rs1771718, rs1771723, rs1746831 and rs1746828. All barcoded libraries were sequenced on a HiSeq 2500 (Illumina) and bioinformatics analysis was performed using the R program 4C-ker. We observed significant interactions between rs767127 viewpoint and *RAD51B* promoter and rs1771723 with *AR* and *E2F3* genes. Results are being validated using 3C. In our latest fine-mapping study we identified three independent PCa susceptibility signals at chr2q37, within the *NGEP-L* gene. This gene encodes a cell membrane protein involved in cell-cell signalling and despite being predominantly expressed in prostate epithelial cells, the long transcript is not expressed in most of the human prostate cell lines analysed by RNA sequencing. We used CRISPR/Cas9 technology and pCMV6-AC-GFP TrueORF vector to knock out and overexpress the gene in order to investigate its role in three human prostate cell lines: RWPE-1, DU145 and DU145 MN1. Positive clones have been selected for proliferation, viability and apoptosis assays in 3D and results of these will be presented. In conclusion, we present here two different approaches used for the functional characterisation of candidate SNPs and novel genes which provide insights into the understanding of their association with prostate cancer risk.

932F

Germline multi-gene testing in unselected endometrial cancer patients.

M. Janatova¹, J. Soukupova¹, K. Lhotova¹, P. Zemankova¹, M. Borecka¹, L. Stolarova¹, M. Vocka², P. Kleiblova^{1,3}, F. Lhotova¹, M. Koudova⁴, S. Tavandzis⁵, L. Foretova⁶, E. Machackova⁶, Z. Kleibl¹. 1) Institute of Biochemistry and Experimental Oncology, First Faculty of Medicine, Charles University, U Nemocnice 5, 12000 Prague 2, Czech Republic; 2) Department of Oncology, General Faculty Hospital and First faculty of Medicine, Charles University, U Nemocnice 2, 12000 Prague 2, Czech Republic; 3) Institute of Biology and Medical Genetics, General Faculty Hospital and First faculty of Medicine, Charles University, Albertov 4, 12800 Prague 2, Czech Republic; 4) GENNET laboratories Prague, Kostelní 9, 17000 Prague 7, Czech republic; 5) AGEL laboratories Novy Jicin, Purkynova 2138/16, 74101 Novy Jicin, Czech Republic; 6) Department of Epidemiology and Tumor Genetics, Masaryk Memorial Cancer Institute, Zlutý kopec 7, 65653 Brno, Czech Republic.

Endometrial cancer is the most common gynecological cancer, the incidence in the Czech Republic was 37/100000 in 2015. Compared to ovarian cancer, it was supposed that only 2 – 3% of all endometrial cancers are caused by inherited mutations. The majority of hereditary endometrial cancers is associated with Lynch syndrome. Nevertheless, the role of other cancer predisposition genes remains unclear. The aim of our study was to explore the prevalence of various cancer susceptibility genes mutations as well as identification of new candidate endometrial cancer predisposition genes in our set of unselected patients. We analyzed germline DNA of 107 unselected endometrial cancer patients (aged 27-91). We used a custom-designed CZECA panel for the targeted next-generation sequencing (NGS) analysis of cancer-predisposition and candidate genes. The panel targets 219 selected genes with a known predisposition to hereditary cancer syndromes and other genes that code for proteins involved in the DNA repair and/or DNA damage response with uncertain clinical relevance. In total, we identified 12 clearly pathogenic mutations in hereditary cancer predisposition genes in 11 (10,3%) endometrial cancer patients. We found four mutations in *BRCA1*, four mutations in Lynch genes (2x *MLH1*, 1x *MSH2*, and 1x *MSH6*); one patient harbored both *BRCA1* and *MLH1* mutation. In addition we found 2 mutations in *CHEK2* and one mutation in *ATM* and *NBN* each. Four of the mutations' carriers (2x *BRCA1*, *CHEK2*, and *MSH6*) were diagnosed with endometrial cancer only, seven carriers had personal history of multiple tumors. We also found 27 truncating variants of other genes in 26 patients and 38 missense alterations in 31 patients predicted to be pathogenic, whose impact on protein function and clinical significance needs to be evaluated. Our data indicate that the frequency of germline mutations in unselected endometrial cancer patients might be higher than previously expected. It is also showed that endometrial tumors can develop either within Lynch and HBOC syndromes or as a unique cancer diagnose. The mutation spectrum also confirms that wider range of phenotypes associated with germ-line mutations of cancer-susceptibility genes is becoming increasingly apparent. Acknowledgement: This work was supported by grants AZV MZCR NV15-27695A, NV16-29959A and the grants of Charles University PROGRES Q28/LF1 and SVV2018/260367.

933W

Using massively parallel reporter assay (MPRA) to simultaneously screen for functional variants at 22 common pancreatic cancer risk loci identified by genome wide association studies (GWAS). A. Jermusyk¹, L. Gu¹, L. Rost¹, J. Choi¹, L. Colli², J. Zhong¹, J. Hoskins¹, I. Collins¹, S. Chanock³, K. Brown¹, K. Yu², L. Amundadottir¹. 1) Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Gaithersburg, MD; 2) Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Gaithersburg, MD; 3) Laboratory of Genetic Susceptibility, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Gaithersburg, MD.

GWAS have identified over 20 pancreatic cancer risk loci in individuals of European descent. After GWAS, fine-mapping and functional analysis must be carried out to identify functional variants from the large number of highly correlated variants at each locus, a time intensive process. We used a high-throughput sequencing approach, Massively Parallel Reporter Assay (MPRA), to simultaneously test correlated variants from multiple pancreatic cancer risk loci for allele specific enhancer activity. We tested 950 germline variants at 22 established and suggestive risk loci in two pancreatic cancer cell lines (PANC-1 and MiaPaCa-2) and a human embryonic kidney cell line (HEK293T). Significant allelic differences were noted for one or more variants at 21 risk loci, indicating that some risk loci may contain multiple functional variants. We validated several of these functional variants by individual luciferase assay and performed electrophoretic mobility shift assay (EMSA) to identify allele specific transcription factor binding. One of these loci we are validating is 22q12 which includes XBP1, a transcription factor important for pancreatic homeostasis and recovery after inflammation. Three of the top variants at this locus from our MPRA analysis also exhibit expression quantitative trait loci (eQTL) in pancreatic tissue datasets (GTEx set, n=115 samples in histologically normal tissues; TCGA pancreatic cancer tissues, n=115) with XBP1. Since XBP1 is an important component of the unfolded protein response (UPR), this may result in an abrogated response to endoplasmic reticulum (ER) stress. We anticipate that our high throughput MPRA approach will shorten the time from GWAS discovery to biological understanding for multiple pancreatic cancer risk loci. This will help us unravel the underpinnings of the mechanisms that influence pancreatic cancer susceptibility.

934T

High-density genotyping and whole-genome sequencing-based investigation of malignant pleural mesothelioma primary cells. A.K. Mandal¹, A. Nastase¹, T. Chernova², D. Morris-Rosendahl^{1,3}, M. Edwards^{1,3}, R.C. Rintoul⁴, E. Lim⁵, S. Popat⁶, M. Lathrop⁷, A.E. Willis², A. Nicholson⁸, W. Cookson^{1,9}, M. MacFarlane^{2,9}, M.F. Moffatt^{1,9}. 1) National Centre for Mesothelioma Research, Imperial College London, United Kingdom; 2) Medical Research Council Toxicology Unit - University of Cambridge, Hodgkin Building, Leicester, United Kingdom; 3) Clinical Genetics and Genomics, Royal Brompton and Harefield NHS Foundation Trust, United Kingdom; 4) Department of Thoracic Oncology, Papworth Hospital, Cambridge, United Kingdom; 5) Department of Thoracic Surgery, Royal Brompton and Harefield NHS Trust, London, United Kingdom; 6) Department of Medicine, Royal Marsden Hospital, London, United Kingdom; 7) Department of Human Genetics, McGill University-Génome Québec Innovation Centre, Québec, Canada; 8) Department of Pathology, Royal Brompton and Harefield NHS Trust, London, United Kingdom; 9) Corresponding author.

Malignant pleural mesothelioma (MPM) is a rare and aggressive disease with poor prognosis on chemotherapy standard of care. Research into genomic aberrations that can help to stratify patients are urgently needed. As a model system, primary cells are more representative of the tumour tissue and in this study, using genotyping and whole-genome sequencing (WGS), we aimed to characterise more than 20 MPM primary cells. Twenty-three primary cells (obtained from the MRC Toxicology Unit, Leicester, n=20 and Mesobank UK, the BLF funded Mick Knighton Mesothelioma Tissue Bank, n=3), all derived from patients' tumours and one normal omental mesothelial primary cell were included in the study. All primary cells were analysed by genotyping and 20 of them further through WGS. Regions of significant focal copy-number aberrations (CNA) were identified using a combination of ASCAT (v.2.4.4), DNACopy (v.1.52.0) and GISTIC (v2.0.23) softwares. For WGS, alignment, quality-checking and variant calling were done using BWA, Picard and GATK softwares respectively. Variants were annotated using Ensembl Variant Effect Predictor (VEP; v92). Twenty-one of the primary cells were derived from male patients, with an average age of 68.2 years. Histologically, epithelioid was the major subtype. ASCAT software called eight of the primary tumour cell and the normal omental as 'non-aberrant'. For the aberrant primary cells, locus 9p21.3 containing *CDKN2A* and *MTAP* genes, was the most common deletion, followed by locus 16p13.3 (containing the *RBFOX1* gene). Loss of whole chromosome 4 was also seen in 4/15 (26.7%) of the analysed samples. Among the ASCAT classified 'aberrant' cell that were whole genome sequenced, analysis showed *BAP1* to be the most frequently mutated with 5/11 cases. Also observed were *NF2* in 4/11 cases, *SYNE1* in 3/11 cases and *TP53* in 2/11 cases. These mutations are primarily either frameshift, stop-gain or splice-site altering which indicates the tumour suppressor role of these genes. On the other hand, among ASCAT classified 'non-aberrant', all but one had no mutations detected in hallmark genes. That one primary cell has a splice-site altering *MSH6* mutation. MPM primary cells show similar mutation and CNA profile as tumour tissue, supporting their use as a pre-clinical model for understanding tumorigenesis and drug resistance. Further investigations using DNA methylation and transcriptome profiling are ongoing to make this model system more comprehensive.

935F

The identification of rare variants in Tasmanian prostate cancer pedigrees using whole-genome sequencing. K. Raspin¹, LM. FitzGerald¹, JR. Marthick¹, MA. Field^{2,3}, RC. Malley^{4,5}, RJ. Thomson⁶, A. Banks⁷, S. Donovan⁷, JL. Dickinson¹. 1) Menzies Institute for Medical Research, University of Tasmania, Hobart, Tasmania, Australia; 2) Australian Institute of Tropical Health and Medicine, James Cook University, Cairns, Queensland, Australia; 3) Genome Informatics, John Curtin School of Medical Research, Australian National University, Canberra, ACT, Australia; 4) Royal Hobart Hospital, Hobart, Tasmania, Australia; 5) School of Medicine, University of Tasmania, Hobart, Tasmania, Australia; 6) Western Sydney University, Sydney, New South Wales, Australia; 7) Hobart Pathology, Hobart, Tasmania, Australia.

Prostate cancer (PCa) is a highly heterogeneous disease with studies suggesting a degree of heritability greater than other cancers. It has been found that as much as 58% of disease risk can be explained by heritable factors. While more than 150 common genetic risk variants have been identified, these variants still only explain a minor portion of risk, are largely of low to moderate effect size, and are functionally ambiguous. There has recently been significant success in the discovery of rare genetic variants contributing to complex disease, through next-generation sequencing of large families, where rare variants are enriched and there is reduced genetic complexity. Here, we have applied whole-genome sequencing (WGS) to several large Tasmanian PCa pedigrees with the aim of identifying rare genetic variants contributing to the development of PCa. Thirty-seven individuals from six PCa pedigrees were WGS on the Illumina HiSeq XTM Ten platform. Variants were prioritised on a per-family basis by frequency (<1% in 1KGP, UK10K, ExAC and ESP), segregation with disease, mutation type (missense, nonsense or splice) and predicted functional consequence (CADD, PolyPhen and SIFT). Unaffected older male relatives and population controls were also WGS and used to prioritise variants based on non-sharing. After additional genotyping in our larger familial and population datasets, familial-based association testing found two variants to be significantly associated with PCa risk. A rare variant in the non-coding region of the *EZH2* gene was identified in two Tasmanian families, and has the potential to effect splicing. *EZH2* is constitutively overexpressed in many cancer, including PCa, and therefore, functional studies are currently underway to determine the effect of the mutation on gene and protein expression. This study was designed to identify new biological pathways involved in the pathogenesis of familial PCa and may lead to novel diagnostic and therapeutic targets for PCa patients.

936W

Integrating exome and transcriptome sequencing data revealed genes of TLR-activated MAP-Kinase signaling to be clinically important in oral cancer. R. Singh¹, M. Pal², B. Roy¹. 1) Human Genetics Unit, Indian Statistical Institute, Kolkata, India; 2) Department of Oral Pathology, Guru Nanak Institute of Dental Science & Research, Kolkata, India.

Background: Oral cancer constitutes about 95% of head and neck cancers and is one of the most prevalent types of cancers in India. Although whole exome studies in oral cancer have reported somatic mutational driver genes, studies on integrating whole exome and transcriptome sequencing data in the same tumors to identify potential therapeutic targets have been rare. **Objective:** Aim of the study is to integrate mutation and expression data to identify potential therapeutic targets in oral cancer. It further intends to explore the impact of somatic mutations on allelic expression imbalance. **Methods:** Whole exome and transcriptome data was generated from 11 pairs of tumor-normal oral cancer samples to uncover somatic mutations and expression deregulation. Allelic expression imbalance was also calculated by comparing the mutated allele frequency in genes and corresponding transcripts in tumor tissues. **Results:** Majority of mutated genes were found to be involved in TLR-activated MAPK signaling. Upon analyzing expression and mutation data, it was realized that genes altered due to mutation were mostly different from those which showed deregulated expression. Genes of receptor nodes (*TLR1*, *TLR2* and *TLR8* and *IFNAR2*) and downstream effector nodes interleukins (*IL8* and *IL1B*) were upregulated while mutated genes mostly accumulated in intermediate MAPK signaling components. Mutations in MAPK signaling and upregulated expression of *TLR2* (most deregulated receptor) and *IL8* (most deregulated effector) were confirmed from TCGA HNSCC cohort. *IL8* was also upregulated in a set of trios (n=14) with oral tumors, leukoplakia and normal samples from same individuals. Few genes of MAPK signaling pathway such as *TP53*, *KRAS*, *FOS*, *GNG12* and *HSPA2* further showed allelic expression imbalance with a preferential expression of mutated alleles than wild type allele. Apart from mutation in *TP53* and *CASP8*, 5-year disease free survival rate was calculated using TCGA HNSCC cohort data and found to be lower than 50% of the patients with mutations in other MAPK signaling genes, such as *CACNA1D*, *CACNA2D4* and *PRKCG*. **Conclusion:** Overall, the results showed importance of *IL8* in progression of leukoplakia to oral cancer, which is reported to be clinically important in other cancers. It also suggests other mutated components of TLR-activated MAPK pathway, *KRAS* and *HSPA2* (with allelic expression imbalance) and *CACNA1D* (associated with survival rate) as potential drug targets, similar to other cancers. .

937T

Truncating and deleterious missense hereditary mutations in *CHEK2* gene confer a significant risk for development of female and male breast cancer.

L. Stolarova¹, P. Kleiblova¹, F. Lhotá¹, J. Hojny¹, M. Janatova¹, P. Zemankova¹, M. Cerna¹, K. Lhotova¹, M. Borecka¹, J. Kotlas², M. Vocka⁴, V. Stranecky⁵, K. Krizova³, K. Burdova³, J. Soukupova¹, L. Macurek³, Z. Kleibl¹.

1) Institute of Biochemistry and Experimental Oncology, First Faculty of Medicine, Charles University, Prague, Czech Republic; 2) Institute of Biology and Medical Genetics, First Faculty of Medicine, Charles University and General University Hospital, Prague, Czech Republic; 3) Laboratory of Cancer Cell Biology, Institute of Molecular Genetics of the ASCR, Prague, Czech Republic; 4) Department of Oncology, First Faculty of Medicine, Charles University and General University Hospital, Prague, Czech Republic; 5) Research Unit for Rare Diseases, Department of Pediatrics and Adolescent Medicine, First Faculty of Medicine, Charles University and General University Hospital, Prague, Czech Republic.

Hereditary mutations in *CHEK2* gene predispose to various types of tumors, including breast cancer. Data from next gene sequencing (NGS) revealed a spectrum of other *CHEK2* mutations with a strong population variability. While truncating variants affecting kinase domain are considered as deleterious, the clinical significance of rare missense variants is unknown. The entire *CHEK2* coding sequence was analyzed in 1928 high-risk breast (BC) and/or ovarian cancer (OC) patients negatively tested for BC/OC-susceptibility genes mutations (by gene-by-gene analysis/ multigene NGS) previously and in 2184 geographically-matched controls. We identified 8 truncations in 39 patients (including two recurrent large genomic rearrangements) and 25 non-frame shift variants (in-frame deletion or SNV) in 96 patients. Using CRISPR/Cas9, we developed *CHEK2* knock-out cell line enabling functional classification of VUS revealing 12/25 VUS as deleterious. Deleterious mutations were significantly enriched (over controls) in female BC patients (51/1526; 3.34% vs. 8/2184; 0.37%; OR=3.87; p=5.25E-13), male BC patients (4/48; 8.3%; OR=10.17; p=8.22E-5), OC patients (6/354; 1.69%; OR=2.64; p=0.008). The analysis of histopathological data in female BC patients showed that carriers of deleterious *CHEK2* variants develop more frequently luminal A subtype (19/37; 51% vs. 269/898; 30% in non-carriers; p=0.006) and G2 tumors were detected in 27/40 (67.5%) mutation carriers vs. in 399/898 (44.4%) of non-carriers; p=0.0005). Our results demonstrated that functional analyses of *CHEK2* VUS enable sorting of neutral from deleterious mutations that contribute to the BC/OC risk, thus help to identify clinically significant hereditary *CHEK2* variants. Supported by grants AZV NV15-28830A, NV-16-29959A, SVV2018/260367 and GAUK 762216.

938F

Detection of somatic L1 insertions in single nuclei from cancer tissues.

K. Yamaguchi, A.O. Soares, J.A. Choi, H.H. Kazazian Jr. Johns Hopkins University, Baltimore, MD.

Long interspersed element-1s (L1s) are transposable elements that proliferate in human genomes. L1 retrotranspositions are endogenous mutagens that cause insertional mutations in chromosomes and have caused over 150 cases of genetic diseases. Recently, our group found somatic insertions in Barrett's esophagus-, pancreatic-, colon- and gastric cancer. These insertions are present in both pre-cancer and cancer, suggesting that L1 insertion events occur before or very early during cancer development. However, from these findings using bulk gDNA from cancer patients, we could not know how the number and varieties of L1 retrotransposition events vary among individual cells and how the cells which have de novo somatic insertions are distributed in pre-cancerous and cancer tissues. To further understand somatic L1 retrotransposition which contributes to cancer induction, we established a new method to detect L1 insertions from single nuclei based on capturing circularized L1 fragments containing 3' flanking gDNA. Using this method, we investigated L1 insertions in single nuclei from colon cancer, liver metastasis, and normal liver. By PCR and Sanger sequencing, we found nine new L1 insertions that were not observed in bulk gDNA from the same patient. From further investigation in single nuclei, we found five other insertions that had been previously observed in bulk gDNA from the same patient. Of these 14 insertions, 4 were also present in normal liver. Therefore using this new technique we can find new somatic L1 insertions in single nuclei that are not found in bulk gDNA study. In addition, many of these insertions in cancer are already present in normal cells. Based on this and future data, we should be able to determine how the new L1 insertions spread in cancerous cells.

939W

Concordance of *ERBB2* amplification identification in breast primary tumors and cfDNA. Y. Zheng¹, T.F. Yoshimatsu¹, S. Dentre^{2,3}, P. Polak⁴, O.I. Olopade¹. 1) Department of Medicine, The University of Chicago, Chicago, Illinois, United States of America; 2) Wellcome Trust Sanger Institute, Cambridge, United Kingdom; 3) The Francis Crick Institute, London, United Kingdom; 4) Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai Hospital, New York City, New York, United States of America.

Background: Breast cancer subtyping in resource poor setting is challenging due to lack of resources including access to molecular pathology. The amplification of the *ERBB2* gene results in overexpression of HER2 protein (human epidermal growth factor receptor 2), and HER2 targeted therapy has been shown to be effective in patients with HER2+ breast tumors. Recent development of robust cell-free tumor DNA (cfDNA) sequencing could be used to determine the HER2 status of tumor before resection. **Methods:** The cfDNA were extracted from plasma samples collected from eight Nigerian invasive breast cancer patients prior to surgery. Ultra-low-pass whole genome sequencing (ULP-WGS, 0.1X) and ichorCNA analyses were performed to quantify the percentage of tumor DNA in the plasma. In contrast, high-coverage WGS (90X) of untreated primary tumors from the same patients has been done and somatic copy number alterations (SCNAs) were called using the Battenberg algorithm. For each patient, we compared the breast cancer SCNAs identified by ichorCNA and Battenberg and particularly in the *ERBB2* region. **Results:** The mean age of the eight breast cancer patients was 51 (SD±9.7) years old. Tumors included in the study were 1 HR (hormone receptor)+/HER2-, 3 HR-/HER2-, and 4 HR-/HER2+. Among five patients (3 HER2- and 2 HER2+) with tumor fraction >0.05 (average 0.19) in cfDNA, the genome-wide and *ERBB2* region-wide SCNA predictions were comparable between the two methods. In addition, two *ERBB2* amplification events were identified by both NGS-based methods, which was in line with the HER2+ status of the patients. In the remaining three patients (1 HER2- and 2 HER2+), Battenberg detected *ERBB2* amplifications in two HER2+ tumors correctly, whereas ichorCNA was underpowered to detect apparent SCNAs including the *ERBB2* amplification, likely due to low tumor fraction in their cfDNA (0.031, 0.04, and 0.041 respectively). **Conclusion:** Our study show that *ERBB2* amplification can be called confidently at time of diagnosis and before tumor resection using cost-effective ULP-WGS from as low as 0.05 tumor fraction, allowing for earlier detection of gene amplification and strategic planning of targeted treatment. Furthermore, the ichorCNA approach accurately captured the SCNA profile of primary tumors from cfDNA, which suggests it can be useful on other clinically actionable SCNA segments in an effort to provide effective treatments earlier.

940T

Characterisation of renal cell carcinoma associated constitutional translocations by whole genome sequencing-derived structural variation analysis and literature review. P.S. Smith¹, H. West¹, J. Whitworth¹, J. Cook², M.D. Tischkowitz¹, E.R. Maher¹. 1) Department of Medical Genetics, University of Cambridge, Cambridge, United Kingdom; 2) Department of Clinical Genetics, Northern General Hospital, Sheffield, United Kingdom.

Renal cell carcinoma (RCC) is a group of genetically heterogeneous neoplasms originating in the kidney and though most cases occur sporadically, a proportion of cases are heritable. Constitutional translocations, particularly involving chromosome 3p, have been previously reported as a cause of inherited RCC and to date 17 individual cases/families have been described. We have identified a series of rare individuals with constitutional translocations ascertained following a diagnosis of RCC and cytogenetic analysis. To investigate the potential mechanisms of RCC predisposition, we performed whole genome sequencing in three cases (t(2;17)(q21;q11.2), t(3;6)(p14.1;p12.1), t(3;14)(q13.3;q22)). Sequencing was aligned to the GRCh38 human reference build and structural variants were called by Manta structural variation and indel caller. Sanger sequencing was performed to confirm candidate breakpoint regions. Most breakpoints did not disrupt known coding regions suggesting that chromosomal instability rather than disruption of a specific gene may be the cause of RCC susceptibility. However a t(2;17)(q21;q11.2) breakpoint disrupted a region 12Kb upstream of *NLK*, the protein product of which interacts indirectly with F-box/WD repeat-containing protein 7 (FBXW7). Interestingly, a RCC-associated constitutional t(3;4)(q21;q31) was reported previously to disrupt *FBXW7* (Kuiper RP et al. Cancer Genet Cytogenet. 2009;195:105-11).

941F

Using capture-C to map chromatin architecture of 24 renal cancer susceptibility regions in human cells. L. Jessop¹, L.M. Colli¹, K. Jones², S.J. Chanock³. 1) LGS, DCEG, NCI, Bethesda, MD; 2) CGR, DCEG, NCI, Bethesda, MD; 3) DCEG, NCI, Bethesda, MD.

Cancer susceptibility loci identified by genome-wide association studies (GWAS) are frequently located in intergenic regions and are expected to be involved in regulating gene expression. Because regulatory regions can be located distant from their target promoter, it is not straightforward to identify the gene targets of GWAS hits. The chromosome conformation capture methodologies (such as 3C, HiC) have been used to gain understanding about the 3-dimensional architecture of chromosomes and enhancer-promoter interactions; however, the resolution of Hi-C experiments depends in large part on the depth of sequencing, making it difficult to probe specific promoter-enhancer interactions. Others have addressed this limitation by adding a sequence-capture step (Capture-C) to focus on interactions with regions of interest. Here we apply the Capture-C method to 24 regions identified by GWAS as achieving genome-wide threshold for significance, or approaching it, for association with renal cancer susceptibility. HiC libraries were prepared from HEK293T or ACHN cell lines using the 4-cutter, DpnII. The pool of biotinylated-RNA oligos used for sequence capture was designed to cover all SNPs with an LD of $r^2 \geq 0.2$ to the lead SNPs from each of the 24 GWAS regions. Resulting Capture-C libraries were paired-end sequenced on Illumina HiSeq4000 or NextSeq. After trimming and filtering there were 595M PE reads for HEK293T, and 681.5M for ACHN. Reads were mapped to GRCh37/hg19 using the HiC-Pro pipeline, and the resulting BAM files were processed using CHiCago (settings minFragLen 25, maxFragLen 2500, binsize 1250, all others default) to detect statistically significant interactions. We found 445,773 interactions in HEK293T, and 809,142 in ACHN with a CHiCago signal cutoff of at least 5. We previously reported that rs7132434 regulates expression of *BHLHE41*, at the 12p21 renal cancer susceptibility locus. Our Capture-C results confirm this by showing a physical interaction between this SNP and gene. This new resource will be used to further explore the chromatin interactions of these 24 regions.

942W

Comparison of intra-tumor heterogeneity and clonal evolution across lung cancer subtypes by multi-region whole genome sequencing. T. Zhang¹, J. Sampson¹, P. Bertazzini², A. Pesatori², D. Consonni², B. Zhu¹, B. Hicks¹, X. Hua¹, J. Shi¹, K. Brown¹, S. Chanock¹, M. Landi¹. 1) National Cancer Institute, Bethesda, MD; 2) University of Milan, Milan, Italy.

Both lung squamous cell carcinoma (LUSC) and lung cancer with neuroendocrine features (LUNE; i.e., small cell carcinoma (SCLC), large cell carcinoma (LCNEC) and other subtypes) are associated with tobacco smoking and have a very poor prognosis. We sought to characterize their clonal architecture and illustrate how the evolutionary patterns differ across these lung cancer subtypes. We performed multi-region whole genome sequencing of 83 samples from 36 LUSC and 67 samples from 23 LUNE patients that had been resected before systemic therapy. Among somatic copy number alterations, we observed genome doubling intra-tumor heterogeneity (ITH) and massive genomic instability in both subtypes. For somatic nucleotide variants (SNVs), the total mutation burden was higher in LUSC (16.7 average number of mutations/Mb) than LUNE (12.7/Mb), but for SNVs ITH was higher in LUNE. Among driver genes, *TP53*, *CDKN2A*, *CR1*, and *NRK* in LUSC and *TP53*, *KRAS*, *KMT2D*, and *RB1* in LUNE were frequently mutated and were always clonal. Evidence of positive selection was found for *TP53*, *PTEN*, *KMT2D* and *KEAP1* in LUSC, and for *TP53* and *KRAS* in LUNE. As expected, the dominant mutation signatures in LUSC clones included signatures 4, 1, and 5, while APOBEC signatures (2 and 13) were predominantly found in sub-clones. In contrast, very little evidence of APOBEC signatures were found in LUNE. Among structural variants, LUNE had more events (339.25 average number of events/tumor) than LUSC (206.7/tumor), with considerable ITH in both subtypes, particularly in LUNE. In LUNE, the dominant mechanism responsible for SVs included fork stalling and template switching within the microhomology-mediated break-induced repair (MMBIR). We found recurrent fusion events with hotspots including those previously identified near *TTC28/CHEK2* (42% LUSC and 35% LUNE) and *LRP1B* (42% LUSC). Additional hotspots were found in chromosome bands 20p12.1 (22% LUSC), 1p31.1 (19% LUSC) and 12p11.22 (22% LUNE). We found >10-fold higher transposon insertions (LINE1) in LUSC (71.9 average number of events/tumor) than LUNE (6.3/tumor). Our findings improve our evolutionary understanding of these lung cancer subtypes and warrant further investigations on the mechanisms involved in clonal genomic instability.

943T

Genome-wide analysis of G4 sequences overlapped with SP1/MAZ binding sites in the human genome and their roles in gene regulation of the proto-oncogene Bcl3. V. Bansal¹, S. Cho², A. Ahmed¹, K. Kim^{1,2}. 1) Department of Molecular Cell Biology, Sungkyunkwan University School of Medicine, Suwon 440-746, Korea; 2) Department of Health Sciences and Technology, SAIHST, Sungkyunkwan University, Seoul, Republic of Korea.

G-rich sequences in the genome of any organisms can fold into a special type of non-canonical structure known as a G-quadruplex (G4). It is believed that G4 structures not only act as physical barrier for transcription or translational machinery but also work as an adaptor for recruiting associated proteins. MYC-associated Zinc finger Protein (MAZ) and Specificity Protein 1 (SP1), transcription factors recognizing GC rich sequences, are well known G4 associated proteins that binds to G4s present in the promoters regions of several oncogenes such as MET, VEGF, HRAS and KRAS for the regulation of their gene expression. Although extensive genome-wide studies have been performed to identify the location and function of G4s, characterization of G4s that contains the binding site for the G4-associated proteins, name it "G4 with associated protein binding site (G4-APB)" has less investigated. In this study, we studied G4 with MAZ and SP1 binding sites (G4-MS) present in the promoter regions of all human genes including 2000 cancer genes to investigate their roles in gene expression as well as in cancer. Among the cancer-associated G4-MSs, we further characterized G4-MS present in the promoter region of Bcl3, a well-known proto-oncogene, to investigate the role of G4 with SP/MAZ site and its effect on Bcl3 promoter in cancer progression. In-vitro studies like CD spectroscopy, Thioflavin T assay, Polymerase stop assay, NATIVE PAGE etc, was used to identify and characterized a G-quadruplex (named PU24WT) present in proto-oncogenic candidate Bcl3 gene near the transcription start site (TSS). Direct Chip binding analysis confirmed the binding of both SP1 and MAZ to PU24WT Bcl3 promoter region. Experimental studies reveal that TMPyP4 treatment decreases the Bcl3 luciferase promoter expression. Accordingly, treatment of TMPyP4 decreased Bcl3 RNA level in human breast cancer cell line. Subsequently, Chip data analysis in the presence of Tmpy4 shows decrease in binding of SP1 and MAZ at the PU24WT Bcl3 promoter region. This is the first study done on G-quadruplex structure in human Bcl3 promoter. In addition, since it was confirmed that TMPyP4 treatment can effectively suppress the Bcl3 expression by blocking SP1/MAZ binding to G4-MSs, it is proposed that targeting G4-MS_{BCL3} can be used as a novel strategy for breast cancer treatment. **Keywords: G4, BCL3 Promoter, TMPyP4.**

944F

Methylation quantitative trait loci analysis of human primary melanocytes informs melanoma susceptibility. J. Choi¹, T. Zhang¹, M. Kovacs¹, M. Xu¹, A. Vu¹, S. Loftus², W. Pavan², G. Jonsson², K. Brown¹, NISC Comparative Sequencing Program, Melanoma Meta-Analysis Consortium. 1) Division of Cancer Epidemiology and Genetics, NCI, Bethesda, MD; 2) NHGRI, Bethesda, MD; 3) Division of Oncology and Pathology, Department of Clinical Sciences, Faculty of Medicine, Lund University, Sweden.

Genome wide association studies (GWAS) have identified a multitude of loci correlated with melanoma risk. While the number of GWAS loci is continuously increasing due to large meta-analyses and rapidly growing cohort-based datasets, functional annotation of these loci still lags behind. To this end, expression quantitative trait loci (eQTL) analysis has been critical in identifying target susceptibility genes from GWAS loci, with abundantly available tissue-level eQTL datasets accounting for highly tissue-specific nature of eQTL. However, eQTL alone is far from complete in GWAS annotation, and alternative QTLs including DNA methylation QTL (meQTL) have not been explored as well. DNA methylation is an interesting molecular feature, where both genetic and epigenetic control contribute to gene expression regulation leading to cellular differentiation, tissue development, and tumorigenesis. To assess genetic contribution to melanocyte-specific DNA methylation, we established meQTL dataset using cultured primary melanocytes from 106 individuals mainly of European descent. To enable multi-QTL integration in the same samples, we also established RNA-based datasets including eQTL, splice QTL, and microRNA QTL. For meQTL, we measured global DNA methylation using Illumina Human Methylation 450K BeadChip, and also obtained >10 million genotypes by genotyping and imputation. In all, we identified *cis*-meQTL for 13,243 methylation probes at a genome-wide significant level (FDR < 0.05). Significant *cis*-meQTLs were mainly observed in CpG probes displaying high methylation variabilities (> 50%), and clustered in close proximity to the CpG sites (< 500kb). meQTLs were significantly enriched in melanoma GWAS loci (4.98E-11; Wilcoxon rank sum test). While formal colocalization analyses are ongoing, 9 of 20 published melanoma GWAS loci including previously characterized *PARP1* locus displayed immediate overlap with significant meQTL. Subsequent integration of melanocyte meQTL with our RNA multi-QTL data will elucidate the interplay of SNP, methylation, and gene expression, which underlies melanoma susceptibility. Further, we are comparing melanocyte meQTL to those from TCGA melanomas (n = 444) and additional melanoma samples (n = 184) to understand the tumor-normal dynamics of meQTL. Our dataset is one of the first and largest cell-type specific meQTL datasets using primary melanocytes, and will benefit melanoma genetics and pigmentation biology community.

945W

CD44 and EGFR genes expression and response to Cetuximab and Paclitaxel in oral and laryngeal cancer stem cells. L.A.M. Ferreira¹, G.M.M Fernandes¹, A.L.S. Galbiatti-Dias¹, M.M.U. Castanhole-Nunes¹, L.F. Hidalgo¹, R.S. Kawasaki-Oyama¹, P.M. Biselli-Chicote¹, J.V. Maniglia², E.C. Pavarino¹, E.M. Goloni-Bertollo¹. 1) Genetics and Molecular Biology Research Unit - UPGEM, Molecular Biology Department FAMERP- São José do Rio Preto Medical School – Brazil; 2) Otorrinolaringology e Surgery of Head and Neck Cancer Departament - FAMERP- São José do Rio Preto Medical School – Brazil.

Introduction: Head and neck cancer stem cells (HNCSCs) correspond to a small population of cells inside the tumor that exhibit self renewal ability and are responsible for tumor initiation, maintenance, origin of metastases and resistance to treatments. **Objectives:** To identify HNCSCs and non-HNCSCs; to investigate the influence of Cetuximab and Paclitaxel treatments in the cell viability, in both cell subpopulations; to evaluate *CD44* and *EGFR* gene expression in HNCSCs. **Material and Methods:** HNCSCs were isolated from two cell lines (HN13-oral cancer and HEP2-laryngeal cancer) by fluorescence activated cell sorting flow cytometry, which identify cells stained with fluorescence-conjugated CD44, CD117 and CD133. CD44-/CD117-/CD133+ cells were characterized as HNCSCs and CD44-/CD117-/CD133- as non-HNCSCs. The scratch assay was performed to evaluate the potential cell migration. *CD44* and *EGFR* mRNA expression levels in non-treated HNCSCs were determined by quantitative real time PCR (qPCR) using TaqMan® Assay, using non-HNCSCs as relative expression control. After, the cell subpopulations were treated with Cetuximab (0.06mg/ml) and Paclitaxel (0.05mg/ml) for 24h, separately. To assess the response to treatments the MTS assay was performed. Statistical analyses were carried out using Chi-square test and p<0.05 was considered significant. **Results:** HNCSCs presented increased rate of migration compared to non-HNCSCs. We found that HNCSCs from HN13 cell line demonstrated more resistance to Paclitaxel treatment, when compared with non-HNCSCs (p<0.0001). In laryngeal cancer the results were not statistically significant (Table 1). HNCSCs from HN13 cell line presented upregulation of *CD44* gene and downregulation of *EGFR* gene. In HEP2 cell line, the results were opposite. **Conclusion:** The CD44-/CD117-/CD133+ cells have stem cell properties. Paclitaxel treatment is less efficient in HNCSCs from oral cancer cell line. Differential expression of *CD44* and *EGFR* genes observed in HNCSCs contribute to a molecular characterization of the different anatomical sites, which is fundamental for the development of specific therapeutic strategies.

Table 1. Cell viability of HN13 and HEP2 subpopulations after Cetuximab and Paclitaxel treatments. *statistically significant (p<0.005).					
Cell lines	Subpopulations	Cetuximab %Cell viability	p-value	Paclitaxel %Cell viability	p-value
HN13	HNCSCs	97	0.074	88	<0.0001*
	non-HNCSCs	91		44	
HEP2	HNCSCs	87	0.5061	50	0.321
	non-HNCSCs	90		57	

Funding: FAPESP 2015/04403-8, CNPq, CAPES.

946T

Analysis of enhancer/promoter chromatin interactions in normal and cancer endometrial cell lines reveals candidate gene targets at endometrial cancer risk loci. D.M. Glubb¹, P.F. Kho¹, D.J. Thompson², A.B. Spurdle¹, T.A. O'Mara¹, Endometrial Cancer Association Consortium. 1) QIMR Berghofer Medical Research Institute, Brisbane, Australia; 2) University of Cambridge, Cambridge, United Kingdom.

We have previously used genome-wide association studies and meta-analyses to uncover 16 genetic risk loci for endometrial cancer. However, the regulatory targets at these loci have not been systematically evaluated. To identify candidate gene targets at endometrial cancer risk loci, we have assessed chromatin looping interactions between endometrial cancer risk variants and neighboring genes using the global HiChIP technique in endometrial cell lines. As we have found enrichment of endometrial cancer risk variants in H3K27Ac histone modification regions (characteristic of enhancers and promoters), H3K27Ac-associated chromatin interactions were captured for analysis. To perform HiChIP, normal (E6/E7 hTERT immortalized endometrial glandular cells) and cancer (Ishikawa, JHUEM-14 and ARK-1) endometrial cells were cross-linked and proximity ligation performed. H3K27Ac-associated chromatin was isolated and HiChIP libraries prepared before sequencing using the Illumina HiSeq 4000 system. HiChIP paired-end reads were aligned using the HiC-Pro pipeline and filtered for valid reads. The hicchipper package was used to identify and characterise global chromatin interactions (FDR < 0.05). At endometrial cancer risk loci, we identified chromatin interactions between risk variants and more than 60 candidate target genes. Interaction distances ranged from 4 kb to over 1 Mb. Expression of several candidate genes, including *CYP19A1*, *SKAP1*, *SNX11* and *EIF2AK4* was found to associate with endometrial cancer risk in whole blood transcriptome-wide association studies (TWAS). Ingenuity Pathway Analysis revealed that candidate genes were enriched for relevant functional annotations such as morphology of the reproductive system, immortalization, and tumor cell activation, proliferation, adhesion and migration. Furthermore, analyses identified major networks which were enriched for genes involved with p53, MYC-mediated apoptosis, IL-8, IL-2, prolactin, wnt/b-catenin and endometrial cancer signalling. In summary, we have performed the first global studies of chromatin interactions in normal and cancer endometrial cell lines. This analysis has revealed biologically relevant candidate target genes at endometrial cancer risk loci, some of which were supported by TWAS analyses. Transcriptional regulation of candidate target genes by endometrial cancer risk variants and their cellular effects will be assessed in future studies.

947F

Identification of novel gene expression regulators using HyCCAPP. *H.*

Guillen¹, S. Leng², D. Perumalla¹, A. Jadhav¹, M. Olivier¹. 1) Wake Forest School of Medicine, Winston Salem, NC; 2) Lovelace Respiratory Research Institute, Albuquerque, NM.

Hybridization Capture of Chromatin-Associated Proteins for Proteomics (HyCCAPP) is a recently developed methodology aimed at identifying, in an unbiased manner, DNA-bound proteins at a region of interest. In this application, we have adapted the approach to identify proteins regulating gene expression at promoter loci of interest. Crosslinked chromatin fragments are enriched through hybridization capture and bound proteins are identified and quantified by mass spectrometry. The *MGMT* promoter has been shown to regulate gene expression in a haplotype-specific manner in lung cancer cells. In an effort to identify haplotype-driven changes in protein binding correlated to distinct hypermethylation profiles within the *MGMT* promoter, H1299 lung cancer cells carrying different haplotypes were analyzed using the HyCCAPP methodology. Novel regulatory binding proteins, specific for individual haplotypes, are being validated using HyCCAPP, while at the same time, assembling the system to identify potentially targetable proteins mediating these interactions. This methodology can be a useful tool to explore known variants associated to changes in gene expression. This work is supported by NIH/NHGRI grants P50HG004952 and R01GM109099 and the Crump foundation.

948W

Small noncoding RNAs associating with the outcome of breast cancer patients. *J.M. Hartikainen¹, S. Heikkinen^{2,3}, E. Kärkkäinen¹, M. Tengström^{4,5}, V.-M. Kosma^{1,6}, A. Mannermaa^{1,6}.*

1) School of Medicine, Institute of Clinical Medicine, Pathology and Forensic Medicine, and Translational Cancer Research Area, University of Eastern Finland, Kuopio, Finland; 2) School of Medicine, Institute of Clinical Medicine, University of Eastern Finland, Kuopio, Finland; 3) School of Medicine, Institute of Biomedicine, University of Eastern Finland, Kuopio, Finland; 4) School of Medicine, Institute of Clinical Medicine, Oncology, and Cancer Center of Eastern Finland, University of Eastern Finland, Kuopio, Finland; 5) Cancer Center, Kuopio University Hospital, Kuopio, Finland; 6) Department of Clinical Pathology, Kuopio University Hospital, Kuopio, Finland.

We have performed small RNA-sequencing (RNA-Seq) of breast cancer (BC) tumors to identify small non-coding RNA (sncRNA) expression profiles which associate with the outcome of the patients. The underlying sample set is the 195 invasive BC tumor and 20 benign breast tissue samples from the Kuopio Breast Cancer Project (KBCP) study material. KBCP material is collected from a genetically homogeneous population of Eastern Finland, and includes e.g. fresh-frozen breast tissue samples, EDTA blood samples and serum, as well as comprehensive background/lifestyle information, clinical, treatment and follow-up data extending up to 25 years. Total RNA was extracted from fresh-frozen tissues using the Ambion mirVana miRNA Isolation Kit. Small RNA-Seq was performed using the Illumina TruSeq Small RNA Library Prep kit and the Illumina MiSeq sequencing instrument. Bioinformatic analysis of non-miRNA sncRNA expression included read quality assessment (FastQC v10.1), adapter trimming (TRIMMOMATIC v0.33), and removal of e.g. ribosomal RNA reads (Bowtie2 v2.2.3). Preprocessed reads were aligned (Tophat v2.0.13) to human reference sncRNA transcriptome (subset of GENCODE v22). Gene-wise read counts were collected using R function GenomicAlignments::summarizeOverlaps (v1.6.3), and statistically significantly differentially expressed (DE) sncRNAs were identified using R package DESeq2 (v1.14.1). Survival analyses were performed using R function survival::coxph (v2.41-3). Multivariate survival analyses that took into account clinical and treatment data identified several survival-affecting sncRNAs. For example, when sncRNA expression levels were divided into two groups at the median and clinical and treatment data were included in the statistical model, we observed 11 sncRNAs independently associating with BC-specific (BCSS), 11 with overall (OS), and 4 with recurrence-free survival (RFS) ($P < 0.01$). Higher expression of 7, 8, and 2 sncRNAs associated with worse BCSS, OS, and RFS, respectively. Accordingly, lower expression of 4, 3, and 2 sncRNAs associated with worse BCSS, OS, and RFS, respectively. Furthermore, we identified DE sncRNAs ($P_{adj} < 0.05$) between benign breast tissue and invasive BC, and among BC cases with invasive, local disease, between TNBC and luminal BC. The identified survival-associating sncRNA profiles may reveal signatures that aid in the prediction of therapy responsiveness, prognosis or patient outcome, and provide targets for future drug development.

949T

Methylation alterations in developmental and differentiation genes drive resistance to immunotherapy in a melanoma model. A. Kashefi¹, S. Börmö², S. Sivalingam³, M. Renn⁴, J. Hausen⁵, T. Tüting⁶, M. Hölzel⁷, P. Krawitz², J. Landsberg¹. 1) Department of Dermatology and Allergy, University Hospital Bonn, 53127 Bonn, Germany; 2) Sequencing Core Facility, Max-Planck-Institute for Molecular Genetics, 14195 Berlin, Germany; 3) Institute of Human Genetics, University Hospital Bonn, 53127 Bonn, Germany; 4) Rigontec GmbH, 82152 Planegg, Germany; 5) Institute for Genomic Statistics and Bioinformatics, University of Bonn, Germany; 6) Department of Dermatology, University Hospital Magdeburg, 39120 Magdeburg, Germany; 7) Unit for RNA Biology, Institute for Clinical Chemistry and Clinical Pharmacology, University Hospital Bonn, 53127 Bonn, Germany.

Background: Therapy resistance limits the efficacy of immunotherapy in melanoma patients. We have recently reported that phenotypic tumor cell plasticity in a proinflammatory microenvironment can drive therapy resistance to an adoptive T cell therapy (ACT). Therefore, we hypothesize that epigenetic alterations beyond the genomic mutations might be the major driver of therapy resistance to immunotherapy. To study the co-evolution of resistant phenotype with the epigenetic landscape in an inflammatory microenvironment, we perform an integrative analysis of exome, methylome and transcriptome profiles of primary and ACT therapy resistant cell lines in HGF-CDK4(R24C) transgenic mouse model of melanoma. **Methods:** We designed different resistance systems based on the origin of resistant cell lines. To analyze if ACT selects for nonsynonymous mutations, we performed whole-exome enrichment followed by sequencing (WES) in primary and ACT therapy resistant cell lines. Methylated DNA immunoprecipitation coupled with next generation sequencing (MeDIP-Seq) is used to track the distribution of methylation alterations in target genomes. Expression profiling data are generated using poly-A enrichment and RNA sequencing. **Results:** Differential methylation (DM) analysis for all samples shows that non-resistant cell lines clustered separately from therapy resistant cell lines. We did not find high impact somatic mutations in the exome data of resistant cell lines that could drive the tumor to therapy resistance while DM analyses show significant methylation alterations in the same gene promoter regions. Differential expression (DE) analysis results are consistent with profound DM clustering in our cell lines. Gene set enrichment analyses for significant DE and DM genes show the highest enrichment scores for developmental pathways and differentiation proteins. **Outlook:** Our data supports the hypothesis that profound changes in methylation patterns drive melanoma cell plasticity and immunotherapy resistance in our experimental model. Further functional studies would shed light on the correlation of "inflammatory tumor microenvironment" and epigenetically driven "resistant phenotypes" in melanoma. **KEYWORDS:** DNA-methylation analysis, Melanoma cell plasticity, Immunotherapy resistance.

950F

Regulation of human telomerase reverse transcriptase (hTERT) in human glioblastoma LN18 cells. C. Lavanya, GK. Chetan, MM. Srinivas Bharath, MK. Sabin. Human genetics, National Institute of Mental Health and Neurosciences, Bangalore, India.

Introduction: Increased telomerase activity is a common feature of almost all types of cancers. It can be blocked by targeting the *hTERT* activity at both RNA and catalytic subunits. Synthetic transcription factors, antisense oligonucleotides and inhibitors had been used to regulate *hTERT* activity and had showed varied success rates. The present study hypothesized that the telomerase specific inhibitor BIBR1532 can effectively down-regulate the telomerase activity in LN18 glioblastoma cell line. **Methods:** LN18 glioblastoma cell line was treated with different concentrations of BIBR1532 at different time intervals. MTT assay was performed to determine cell viability after BIBR1532 treatment. *hTERT* mRNA and protein expressions were quantitated using qRT-PCR and western blotting. Flow cytometry assay and TRAP assay was performed to detect the rate of apoptosis and telomerase activity in treated and control groups. One-way ANOVA was used to compare the mean values of each variable in control and BIBR1532 treated groups. **Results:** LN18 cells showed a significant dose dependent cytotoxic effect after treatment with BIBR1532. Treatment with 200 µM BIBR1532 caused more than 70% decrease in viability of LN18 cells at 48 and 72 h. The level of *hTERT* mRNA expression in cells treated with 25 µM, 100 µM and 200 µM BIBR1532 treated groups was decreased ~ 21% and ~ 61.2% and ~ 77% respectively with a $p < 0.05$. We also observed that, the level of *hTERT* protein was decreased to 44%, 61% and 86% at concentrations of 25 µM, 100 µM and 200 µM respectively. The Flow cytometry data showed that the total percentage of apoptotic cells at 200 µM of BIBR1532 with 63.6%, 73.8% and 84.5% total apoptotic cells at 12, 24 and 48 h respectively, compared to control. TRAP assay results showed that 25 µM of BIBR1532 resulted in a partial inhibition, whereas treatment with 100 and 200 µM BIBR1532 inhibited the telomerase activity up to 42.9% and 74.4% respectively ($p < 0.05$). **Conclusion:** BIBR1532 could induce apoptosis in LN18 glioblastoma cells through the down-regulation of telomerase activity at transcriptional and translational level. BIBR1532 may be used as a therapeutic agent to suppress telomerase activity in glioblastoma cell line. Further *in vivo* studies are necessary in order to explore this as a therapeutic agent.

951W

Characterizing the chromatin accessibility landscape of glioblastoma tumors at the single cell level.

R. Raviram^{1,2}, *A. Raman*^{1,2}, *S. Preiss*³, *X. Wang*³, *J. Han*³, *A. Mischel*^{1,2}, *R. Fang*^{1,2}, *L. Hu*^{4,5}, *A. Motamed*⁶, *C. He*^{4,5,6}, *A. Shiau*², *C. Chen*⁷, *B. Ren*^{1,2}. 1) Cellular and Molecular Medicine, University of California San Diego, San Diego, CA; 2) Ludwig Institute for Cancer Research, University of California San Diego, La Jolla, CA; 3) Center for Epigenomics, University of California San Diego, La Jolla, CA; 4) Department of Chemistry and Institute for Biophysical Dynamics, the University of Chicago, Chicago, IL; 5) Department of Biochemistry, The University of Chicago, Chicago, IL; 6) Howard Hughes Medical Institute, The University of Chicago, Chicago, IL; 7) Department of Neurosurgery, University of Minnesota, Minneapolis, MN.

Glioma refers to brain tumors that develop from cells of glial origin and classified based on the type of cell that give rise to the tumor. These include astrocytic tumors of which glioblastoma multiforme (GBM) is the most prevalent among adults and also has a poor prognosis for survival. There is currently no treatment for GBM and is also one of the most difficult to cure by surgery as the tumor can spread to several regions of the brain. One of the main reasons for the difficulty in treatment is because of the heterogeneous nature of the tumor. In order to characterize the heterogeneity of GBM, single cell technologies need to be used. Here we aim to also use single nuclei RNA-Seq and single nuclei ATAC-Seq to first identify the cell populations within the tumor based on their transcriptome profile and epigenetic chromatin landscape. Using ATAC-Seq to identify regions of open chromatin can help us better understand the regulatory network driving the expression of tumor-specific genes. Using these single cell technologies, we aim to characterize the intra tumor heterogeneity by using sections of the tumor from each patient as well as inter tumor heterogeneity between different patients. ATAC-Seq can be also be used to identify copy number variations in the genome, which can allow us to characterize the genetic aberrations in each tumor at the single cell level. Together, this characterization of GBM tumors at the single cell can assist in designing better treatment options that can target the heterogeneous aspects of the tumor to prevent recurrence.

952T

Candidate non-coding driver mutations in super-enhancers and long-range chromatin interaction networks across >1,800 whole cancer genomes.

J. Reimand^{1,4}, *L. Wadi*¹, *L. Uuskula-Reimand*^{2,3}, *K. Isaev*^{1,4}, *S. Shuai*^{1,5}, *V. Huang*¹, *M. Liang*^{2,5}, *J. Thompson*¹, *Y. Li*¹, *L. Ruan*¹, *M. Paczkowska*¹, *M. Krassowski*¹, *I. Dzieladze*¹, *K. Kron*⁶, *A. Murison*⁶, *P. Mazrooe*^{1,6}, *R. Bristow*⁴, *J. Simpson*^{1,7}, *M. Lupien*^{1,4,6}, *M. Wilson*^{2,5}, *L. Stein*^{1,5}, *P. Boutros*^{1,4}. 1) Computational Biology Program, Ontario Institute for Cancer Research, Toronto, Ontario, Canada; 2) Program in Genetics and Genome Biology, SickKids Research Institute, Toronto, Ontario, Canada; 3) Department of Gene Technology, Tallinn University of Technology, Tallinn, Estonia; 4) Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada; 5) Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada; 6) Princess Margaret Cancer Centre, Toronto, Ontario, Canada; 7) Department of Computer Science, University of Toronto, Toronto, Ontario, Canada.

A catalogue of mutations that drive tumorigenesis and progression is essential to understanding tumor biology and developing therapies. Protein-coding driver mutations have been well-characterized by large exome-sequencing studies; however many tumors have no mutations in protein-coding drivers and few non-coding drivers besides the *TERT* promoter are known. To fill this gap, we analyzed 150,000 cis-regulatory regions in 1,844 whole cancer genomes from the largest uniformly processed whole genome sequencing dataset compiled by the ICGC-TCGA Pan-cancer Analysis of Whole Genomes (PCAWG) project. We developed a new cancer driver discovery method, ActiveDriverWGS, validated it on protein-coding genome and found 47 genes of which >90% are known drivers. We then applied the model to cis-regulatory regions from ENCODE and found 41 frequently mutated regulatory elements (FMREs) enriched in non-coding SNVs and indels. FMREs were characterized by aging-associated mutation signatures and frequent structural variants, highlighting their importance in cancer. FMREs were enriched in super-enhancers and long-range chromatin interactions, suggesting that mutations in these regions drive cancer by altering transcriptional programs of distally located genes. The chromatin interaction network of FMREs, chromatin loops and target genes revealed a subset of FMREs where mutations associated with differential gene expression of known and novel cancer genes, activation of immune response pathways and altered enhancer marks. Thus distal genomic regions may include additional, infrequently mutated drivers that act on target genes via long-range interactions mediated by chromatin loops. Our study is an important step towards finding such frequently mutated regulatory regions and deciphering the somatic mutation landscape and disease mechanisms of the non-coding genome.

953F

Transfer learning approaches for robust fine-mapping of putative causal regulatory variants associated with colorectal cancer. A. Shcherbina¹, S. Bien², J.R. Huyghe², A. Saiakhova³, S. Schmit⁴, D. Conti⁵, T. Harrison², F. Qu⁶, L. Hsu², M. Wainberg⁷, M. Bassik⁸, G. Casey⁸, S. Gruber⁸, U. Peters², P. Scacheri³, A. Kundaje^{7,8}, GECCO, CORECT, CCFR Consortia. 1) Biomedical Informatics, Stanford University, Stanford, CA; 2) Fred Hutchinson Cancer Research Center, Seattle, WA; 3) Case Western Reserve University, Cleveland, OH; 4) Moffitt Cancer Center, Tampa, FL; 5) University of Southern California, Los Angeles, CA; 6) University of Virginia, Charlottesville, VA; 7) Computer Science Dept, Stanford University, Stanford, CA; 8) Genetics Dept, Stanford University, Stanford, CA.

We present an integrative approach for context-specific fine mapping and interpretation of non-coding disease-associated variants based on deep learning models of regulatory DNA sequence. We used this approach to analyze over 50 genome-wide loci associated with colorectal cancer (CRC) based on a meta-analysis of 125,218 CRC cases and controls from three consortia: Colon Cancer Family Registry, Colorectal Cancer Transdisciplinary Study, and Genetics and Epidemiology of Colorectal Cancer Consortium. First, we trained novel multi-task, convolutional-recurrent neural networks (NN) to accurately map 1Kb bins of DNA sequence across the genome to corresponding DNase-seq (DHS) profiles across 193 diverse ENCODE+REMC cellular contexts. Next, we fine-tuned the reference NNs on DHS and H3K27ac profiles in 21 primary CRC tumor samples and mucosa from 6 healthy controls. This transfer learning approach resulted in significant accuracy gains relative to training only on the CRC relevant tissues. We used DeepLIFT, a feature attribution method for neural networks, to efficiently infer predictive importance scores of every nucleotide in 1Kb bins centered at 23,727 variants across all associated loci, which include 3910 variants in the 99% credible sets of causal variants. Variants with high, robust DeepLIFT scores across multiple samples, bootstrapped models and heterogeneous genomic background sets, highlighting putative functional regulatory variants, were further interrogated using in-silico mutagenesis (ISM) to estimate the signed influence (effect size) of all alleles on local chromatin state. DeepLIFT profiles of nucleotides surrounding high scoring variants further highlighted sequence features such as transcription factor (TF) motifs harboring these variants. Our approach exhibited high specificity. Within each GWAS locus, across 100s of 99% credible set variants and 1000s of background SNPs, only 1-5 SNPs exhibited high and statistically significant functional scores. Several high scoring SNPs overlapped tissue-specific enhancers in colorectal mucosa. Further, these variants were often found at flanks of DHSs directly disrupting, creating or indirectly influencing motifs of lineage-specific TFs. High confidence candidates are being tested using high-throughput CRISPR screens and reporter assays in CRC lines. Our framework can easily generalize to other disease phenotypes and to score non-coding rare variants and somatic mutations including indels and short SVs.

954W

Exploring the microRNA expression between HPV positive and negative prostate cancer cell lines. K. Tokarz, T. Vilboux, J. Deeken, R. Iyer. Inova Translational Medicine Institute, Falls Church, VA.

MicroRNAs (miRNAs), a class of small RNA molecules 18-25 nucleotides in length, maintain cellular function by regulating mRNA transcription. MiRNAs play a vital role in cell proliferation and apoptosis, and can be considered useful biomarkers of cancer progression. HPV16 is a high risk strain of the human papillomavirus, and the most common detected in tumors. Prolonged infection of a HPV strain can cause various types of cancer, such as cervical, prostate, and head and neck. It has been shown that HPV16 positive tumors have different miRNA profiles than the same tumor type without HPV16. In this study, we wanted to confirm the impact integration of HPV16 on miRNA in prostate cancer (PCa) as well as characterizing similar modification in other cancer types. In addition we wanted to show if HPV18, another high-risk HPV type, has similar effect on miRNA. For this, we performed miRNA sequencing on different cell lines derived from multiple cancer (with or without HPV16 and HPV18) such as PC-3, CRL-1740, and CRL-2876. In this study, we specifically looked at the miRNA expression effecting target genes or pathways in PCa. Using a nuclease protection probe assay, we sequenced over 2200 miRNA probes of a PCa Gleason Grade 4 HPV18+ (CRL-2220), and PCa metastatic HPV negative (CRL-2876) cell line. To analyze the data, we applied a log2 transformation expression of each probe in counts per million. Then the difference in the values of each miRNA probe of the two cell lines was found to determine differential expressions, with the five most upregulated miRNA being: miR-99a-5p, miR-7106-5p, miR-3687, miR-30b-5p, and miR-20b-5p and the five most downregulated miRNA being: miR-155-5p, miR-205-3p, miR-224-5p, miR-31-5p, and miR-205-5p. Looking at the downregulated miRNA, two of the miR-205 family and one miR-31 are present, which correlated to previous studies. In these studies, miR-99 has also been shown to be down-regulated in advanced PCa, inhibiting the growth of prostate cancer cells and decreasing the expression of prostate-specific antigen (PSA). Interestingly, this miRNA cluster has rarely been shown to be upregulated in PCa. One reason for this may be that miR-99a is located on chromosome 21q21, which has been shown to be an integration site for certain strains of HPV. Understanding the effect of HPV18 on miRNA expression would shine some light on the pathophysiology of these HPV positive cancer as well as identifying potential targets for new treatments. .

955T

Description of novel human endogenous retrovirus-protooncogen chimeric transcripts in acute lymphoblastic leukemia. *J.M. Valencia-Reyes^{1,2}, S.A. Tellez-Camacho³, J.T. Granados-Riveron¹.* 1) Laboratorio de Investigación en Genómica, Hospital Infantil de México Federico Gómez, Mexico City, Mexico; 2) Escuela Superior de Medicina, IPN; 3) Centro Universitario de Ciencias Biológicas y Agropecuarias, Universidad de Guadalajara.

Human endogenous retroviruses (HERVs) are remnants of past retroviral infections acquired during our evolution as a species. Most of the HERVs have lost the infectious capacity through loss or mutation of the *env* gene. Genome's repair mechanisms have also processed multiples copies of HERVs by removing internal sequences leaving only long terminal repeats (LTRs). In turn, these LTRs can behave as potent promoters and therefore potentiate transcription of proto-oncogenes if a HERV integrates somatically upstream one of them or alternatively if an ancestral HERV loss its inhibition by methylation. Recently, in lymphomas, two instances of HERV-proto-oncogene chimeric genes have been described as severity as well as survival factors for neoplastic cells. We have developed a novel methodology to discover new such chimeric transcripts and we have evaluated its efficacy by a proof of principle assay by amplifying selectively a constitutive gene in both commercial high-quality RNA as well as RNA derived from neoplastic ALL bone marrow. So far, we have been able to identify some examples of potential chimeric transcripts, which are being characterized.

956F

Integrative genome-wide analysis of long noncoding RNAs in diverse immune cell types of melanoma patients. *Y. Zhang, L. Wang.* University of Maryland School of Medicine, Baltimore, MD.

Genome-wide identification and characterization of long non-coding RNAs (lncRNAs) in individual immune cell lineages helps us better understand the driving mechanisms behind melanoma and advance personalized patient treatment. To elucidate the transcriptional landscape in diverse immune cell types of peripheral blood cells (PBC) in stage IV melanoma, we used whole transcriptome RNA sequencing to profile lncRNAs in CD4+, CD8+, and CD14+ PBC from 132 patient samples. Our integrative computational approach identified 27,625 expressed lncRNAs, 2,744 of which were novel. Both T cells (i.e., CD4+ and CD8+ PBC) and monocytes (i.e., CD14+ PBC) exhibited differential transcriptional expression profiles between melanoma patients and healthy subjects. *Cis*- and *trans*-level co-expression analysis suggested that lncRNAs are potentially involved in many important immune-related pathways and the programmed cell death receptor 1 (PD-1) checkpoint pathways. We also identified 9 gene co-expression modules significantly associated with melanoma status, all of which were significantly enriched for three mRNA translation processes. Age and melanoma traits closely correlated with each other, implying that melanoma contains age-associated immune changes. Our computational prediction analysis suggests that many *cis*- and *trans*-regulatory lncRNAs could interact with multiple transcriptional and post-transcriptional regulatory elements in CD4+, CD8+, and CD14+ PBC, respectively. These results provide novel insights into the regulatory mechanisms involving lncRNAs in individual immune cell types in melanoma and can help expedite cell type-specific immunotherapy treatments for such diseases.

957W

Comparison of circulating cell-free DNA isolation methods by yield and downstream Pyrosequencing analysis. *M. Poulin¹, A. Meyer¹, E. Zhou², S. Miller¹, J. Xu¹, L. Yan¹.* 1) EpigenDx Inc., Hopkinton, MA; 2) Harvard University, Cambridge, MA.

The use of circulating cell-free DNA (ccfDNA) as a diagnostic substrate for many types of diseases, conditions and genetic abnormalities, is becoming more and more sought after as a non-invasive way of early genetic detection. The current technological challenge is obtaining optimal efficiency in the isolation and quantitation of the very small quantities of ccfDNA in pre or early disease patients. Pyrosequencing is an efficient technology to take advantage of small amounts of highly fragmented DNA. Specific genomic regions can be PCR amplified with a primer pair, in which one primer is biotinylated, and sequenced to interrogate very specific DNA mutations. DNA can also first be bisulfite treated, amplified and sequenced to obtain DNA methylation information at specific CpG sites throughout the genome. We have compared five ccfDNA isolation kits on 8 to 12 samples of human blood plasma for DNA yield and Pyrosequencing sensitivity for several Pyrosequencing assays, both on genomic and bisulfite treated DNA. The genotyping assays will be within SNV hotspots of human *BRAF*, *KRAS*, *IDH1* and *IDH2*. We will also be looking at the methylation of human *MGMT*, *MLH1* and *BRCA1*. This examination will not only determine which kit will result in the greatest yield, but will also assay the potential downstream application of Pyrosequencing assays.

958T

A high-resolution chromatin interaction epigenomic map of prostate cancer cells. *S.K. Rhie, A. Perez, F. Lay, S. Schreiner, P.J. Farnham.* Department of Biochemistry and Molecular Medicine and the Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles 90089, CA, USA.

The human genome is densely compacted in three-dimensional space of the nucleus via the action of chromatin-bound histones, with different compartments being in open vs. closed states. Important cellular processes such as transcription occur in the open compartments of the genome. As cells change states, specific genomic compartments transition between open and closed and vice versa, allowing for cell type-specific gene regulation. These genomic compartments are defined by chromatin loops which bring distant regions of the linear genome together in three-dimensional space. Characterizing the chromatin loops in a given cell type is clearly critical for understanding cell type-specific gene regulation. Prostate cancer is the leading cause of new cancer cases and the second most common cancer in men and the fourth most common tumor type worldwide. To better understand the regulatory networks that contribute to prostate tumorigenesis, we have developed high-resolution chromatin interaction epigenomic maps in normal and tumor prostate cells by using in-situ Hi-C. We have compared the chromatin interaction maps with open chromatin regions defined by NOME-seq, active and repressive histone marks and CTCF binding sites defined by ChIP-seq, and transcriptome profiling to provide a comprehensive characterization of chromatin loops in normal and tumor prostate cells. We identified ~6,500 topologically associating domains (TADs) per cell type and annotated these TADs as being enriched for histone marks representing an open (H3K36me3) or closed (H3K9me3 or H3K27me3) state. By comparing TADs, we identified >100 closed (transcriptionally repressed) TADs that changed epigenomic states and >1,000 TADs that changed in size between normal and prostate tumor cells. Moreover, we identified ~20,000 promoter-enhancer loops and potential transcription factors that are enriched at regulatory elements and loop anchors. Interestingly, most of the differentially expressed genes in normal vs. tumor prostate cells had promoter-enhancer loops with similar interaction frequencies between normal and tumor cells. However, the epigenomic states of the regulatory elements located near the loop anchors changed when normal cells transitioned to tumor cells. Our studies provide a list of characterized TADs and loops in normal and prostate cancer cells as a resource to the research community, to enable a better understanding of prostate cancer biology and epigenomics.

959F

Identifying genomic sites with highly predictable DNA accessibility. *K. Wnuk¹, J. Sudol¹, S. Rabizadeh¹, C. Szeto², C. Vaske².* 1) NantOmics, Culver City, CA; 2) NantOmics, Santa Cruz, CA.

DNA accessibility, chromatin regulation, and genome methylation are key drivers of transcriptional events promoting tumor growth. However, understanding the impact of DNA sequence data on transcriptional regulation is a challenge, particularly in noncoding regions of the genome. Recently, neural nets have improved the ability to predict DNA accessibility in specific cell types. This result was further advanced by model changes, and extended to apply to novel biotypes without need to re-train by adding a RNA-seq gene expression signature as input. Here we analyzed model performance by genomic site and identified where it can be applied with high confidence. Matched RNA-seq (304 samples) and DNase-seq (220 samples) data from 74 unique biotypes from ENCODE were used for evaluation. When applied at 1.71M potential accessibility sites across the genome our convolutional net achieved ROC AUC = 0.897 and PR AUC = 0.621 on held out biotypes, and ROC AUC = 0.913 and PR AUC = 0.725 on held out samples whose biotypes were allowed to overlap training. Although compelling compared to prior results the PR AUC underscored that, when considering all sites, performance still needed to improve. We broke down performance by site type and showed that at promoter and promoter flank regions (per ENSEMBL) accessibility prediction is considerably more reliable: PR AUC = 0.839 in held out biotypes, PR AUC = 0.911 in held out samples with biotype overlap. Performance was not sensitive to whether the prediction site was in proximity to genes used in the input RNA-seq signature. We applied the model to 6 TCGA cohorts for which somatic mutations were available, focusing only on promoter and promoter flank regions. We predicted 10878 sites as constitutively accessible, 40823 as constitutively not accessible, and 52891 as facultative. We also identified 5 cases with unique accessibility patterns across several hundred sites that were specific to single patients. Exploring the impact of mutations, we showed there was higher variance in the influence of INDELs on predicted accessibility of a site than SNPs. Only 5.46% of the sites with INDELs and 1.6% of those with SNPs led to a flipped decision by the classifier. These results demonstrate a new tool to query DNA accessibility of samples at the sequence level in silico, allowing us to explore cohorts for correlations of promoter accessibility to phenotypes of interest and shed light on new links that may be worth further investigation.

960W

Epigenetic driven metastasis in osteosarcoma is suppressed through

CDK12 inhibition. *I. Bayles¹, M. Krajewska², A. Saiakhova¹, J. Morrow¹, C. Bartels¹, J. Lu², Z. Faber¹, Y. Fedorov¹, B. Rubin³, D. Adams¹, R. George², P. Scacheri¹.* 1) Genetics and Genome Sciences, Case Western Reserve University, Cleveland, OH; 2) Department of Pediatric Oncology, Dana-Farber Cancer Institute; 3) Pathology and Laboratory Medicine Institute, Cleveland Clinic.

Cancer metastasis is a complex multistep process thought to be driven by the acquisition of new mutations in primary tumor cells. However, recent whole genome sequencing studies comparing primary tumors to distant metastasis have not supported this theory. We hypothesized that the metastatic process is instead driven by epigenetic alterations at enhancer elements that endow tumor cells with metastatic traits. We tested this hypothesis in osteosarcoma (OS [MIM 259500]), an aggressive pediatric cancer that primarily metastasizes to lung. Through epigenomic profiling, we identified substantial differences in enhancer activity between primary and metastatic OS tumors, and between isogenic pairs of highly lung metastatic and nonmetastatic OS cell lines. We term these regions metastatic variant enhancer loci, or Met-VELs. Met-VELs drive coordinated waves of gene expression during metastatic colonization of the lung. Moreover, osteosarcoma lung metastasis in mice can be inhibited by knockdown of AP-1 transcription factors that occupy Met-VELs, by inhibition of individual genes activated by Met-VELs, or through TALEN-mediated genetic deletion of individual Met-VELs. Leveraging these findings, we developed an approach to screen for epigenetically-related compounds that could have therapeutic efficacy against metastatic osteosarcoma. The screening approach is based on a lung explant organ culture model called the Pulmonary Metastasis Assay (PuMA) in which metastatic OS cells are cultured directly within the lung microenvironment, thereby more closely recapitulating the natural metastatic setting in humans. Deploying this strategy to screen a library of 114 epigenetically-related compounds, we found inhibitors of *CDK12* [MIM 615514] to be most effective, reducing OS cell outgrowth in the lung by >90% at submicromolar doses. *CDK12* inhibitors impeded elongating RNA Polymerase II in a gene length- and expression-dependent manner. These effects were accompanied by displacement of *CDK12* from active enhancer elements and global changes in gene expression. Together, our studies demonstrate that the epigenome plays a critical role in equipping OS tumor cells with metastatic competence, and that this property may be targetable for therapy. Our studies further offer a framework for rapid preclinical testing of compounds with anti-metastatic activity, and highlight *CDK12* as a potential therapeutic target in osteosarcoma.

961T

Microbial metabolite butyrate affects allele-specific expression and disrupts chromatin architecture in a colon cancer cell line. M.G. Durrant¹, B. Fremin¹, S.B. Montgomery^{1,2}, A.S. Bhatt¹. 1) Department of Genetics, Stanford University, Stanford, CA; 2) Department of Pathology, Stanford University, Stanford, CA.

Recent studies suggest that the composition of the microbiome can dramatically influence host gene expression and chromatin accessibility. Butyrate is a metabolite and HDAC inhibitor produced in large quantities by bacterial species found in the human gut microbiome. It can act as a source of energy for colonocytes, as an anti-inflammatory agent, and it has been associated with a reduced risk of colon cancer. The mechanism of this anti-cancer effect is not fully understood. We used longitudinal paired ATAC-seq and RNA-seq on the HCT-116 colon cancer cell line to directly determine how the epigenomic response to physiological concentrations of butyrate mediates gene expression. Allele-specific expression (ASE) analysis identified 10 genes, including known tumor growth and metastasis modulators PSPH and PHLDB2, that shift their ASE in response to butyrate, suggesting that *cis*-regulatory variation modulates the impact of butyrate on gene expression. Distinct temporal clusters of butyrate-induced changes in chromatin accessibility emerged. Clusters of peaks that became less accessible (closed) in response to butyrate were strongly enriched for a wide-range of functional pathways, including regulation of Rap protein signal transduction, regulation of JUN Kinase activity, immune and inflammatory responses, and cellular response to trichostatin A (another HDAC inhibitor). These butyrate-induced closed peaks were also found to selectively reside at distal enhancers (Hypergeometric Test, p -value = $1.7e-9$) and were strongly enriched for AP-1 binding motifs (HOMER Motif Enrichment, p -value = $1e-30$). Our results strongly indicate that the chromatin that closes in response to butyrate treatment is enriched within regions that are actively maintained by the SWI/SNF chromatin remodeler (Hypergeometric Test, p -value = $2.3e-60$) and the AP-1 complex (Hypergeometric Test, p -value = $5.4e-107$), indicating that many of the previously identified anti-cancer effects of butyrate may be linked to the disruption of SWI/SNF-dependent chromatin maintenance. In summary, we find that physiological concentrations of the microbial metabolite butyrate 1) influences the ASE of genes previously associated with cancer, 2) dramatically affects chromatin accessibility genome-wide, and 2) specifically affects chromatin at sites that are maintained by chromatin remodelers, shedding new light on the mechanism of action of this important microbial metabolite.

962F

TET1-mediated 5-hydroxymethylcytosine alteration in the pathogenesis of medulloblastoma. H. Kim^{1,2}, J. Li^{1,3,4}, Y. Kang¹, L. Chen¹, X. Yuan³, X. Liao^{3,4}, J. Zhao^{3,4}, Z. Zhang^{3,4}, L. Lin¹, J. Liu^{3,4}, P. Jin¹. 1) Department of Human Genetics, Emory University, Atlanta, GA; 2) Cancer Biology Graduate Program, Emory University School of Medicine, Atlanta, GA; 3) Department of Neurosurgery, Xiangya Hospital, Central South University, Changsha, Hunan, China; 4) Hydrocephalus Center, Xiangya Hospital, Central South University, Changsha, Hunan, China.

Medulloblastoma (MB) is an embryonal tumor that originates from cerebellum and adjacent brain stem. Standard treatment based on surgery, radiation, and chemotherapy improved survival rate; however, it still remains the most common brain tumor in children, and survivors suffer from several side effects of the treatment. Therefore, the development of targeted therapy through a better understanding of molecular characteristics in MB is needed. Recent studies revealed substantial increase in the level of 5-hydroxymethylation (5hmC) and its significant enrichment at cerebellar enhancers during postnatal development, suggesting a critical role of 5hmC in proper cerebellar development. In this study we examined the abundance of 5hmC in MB and observed a significant depletion of 5hmC compared to age-matched normal cerebellum. Genome-wide 5hmC profiling revealed the differentially hydroxymethylated regions (DhMRs) across the genome and the DhMRs could clearly distinguish MB from age-matched normal cerebellum. Interestingly we observed the acquisition of 5hmC near the genes involved in the Notch signaling pathway. Furthermore 5hmC signature identified in patient samples displayed a substantial overlap with 5hmC signature identified in mouse tumors of SmoA1 mouse model that recapitulates human SHH MB. To determine whether the modulation of tumor-specific 5hmC marks could influence MB tumorigenesis, we manipulated the expression of *TET1*, a key enzyme involved in 5hmC production, in vitro and in vivo. We found that a known *TET1* inhibitor could inhibit the growth of primary tumors derived from SmoA1 mouse in vitro. More importantly, when we crossed SmoA1 mice with *TET1* knockout mice, we observed that the loss of *TET1* could significantly delay the onset of tumor-related symptoms. These results together suggest a potentially important role of 5hmC in MB tumorigenesis and indicate an oncogenic nature of *TET1* in this process. Thus *TET1* could serve as a therapeutic target for MB.

963W

Detection of epigenetic interactions using a weighted epigenetic distance-based method. *Y. Wang, S. Wang.* Department of Biostatistics, Columbia University, New York, NY.

DNA methylation plays an important role in cancer progression. Most EWAS (epigenome-wide association studies) performs CpG site-level analysis and discard CpGs without differential methylation when some of these CpGs might be interacting with others on disease outcomes. Exhaustive search for all pairwise interactions on CpG levels are computationally prohibited. Thus, not many studies consider epigenetic interactions, either on CpG levels or on gene levels. Here we propose a weighted epigenetic distance-based method that is able to characterize similarity between any pairs of individuals on methylation measures at multiple CpGs in a gene or a genetic region as well as all pairwise CpG-level interactions. Weights are assigned to both methylation measures and CpG level interactions to up-weight signal CpGs/interactions and down-weight noise CpGs/interactions. We demonstrated the feasibility of the proposed epigenetic distance-based method incorporating interactions in simulation studies as well as the superior performance with weights which is effective in up-weighting signal CpGs and signal interactions than the non-weighted version. Applications to DNA methylation data of The Cancer Genome Atlas (TCGA) breast cancer subtypes identified genes that are differentially methylated between breast cancer subtypes that are due to epigenetic interactions but are missed by comparing methods that ignore epigenetic interactions.

964T

Integrated genomic profiling reveals cellular epigenetic normalization of salivary dysfunction following AAV2-AQP1 treatment. *D.G. Michael, T.J. Pranzatelli, C. Zheng, C. Goldsmith, J.A. Chiorini.* Adeno-associated Virus Biology Section, National Institutes of Health, Bethesda, MD, USA.

Each year, approximately 40,000 patients develop head and neck cancers in the United States, with over 500,000 cases worldwide. Of these cases, 50% of patients experience irreversible damage to the salivary glands exposed to the irradiation field. This damage represents a complex epigenetic state in which hundreds of genes are differentially dysregulated as a response to irradiation. This dysregulation results in salivary dysfunction and cell death of acini, resulting in a loss of salivary function. Current treatments are primarily palliative in nature and focus on the replacement of lost oral fluids without significant improvement in quality of life. The use of Adeno-associated viruses for gene therapy is a promising modality for the correction of Mendelian genetic disorders, but little research has been performed on the use of gene therapy to correct complex disorders that arise from the dysfunction of multiple genes. Previous work at the NIDCR has previously demonstrated in humans that administration of the aquaporin-1 water channel to human patients using adenovirus vectors resulted in subjective improvement of xerostomia in patients. A gene therapy trial for salivary AAV2-AQP1 treatment is currently underway at the NIH for the correction of irradiation induced xerostomia. We evaluated the impact of AAV2-AQP1 treatment on mice treated with 16 Gy of gamma irradiation delivered to the salivary glands. 71 days following irradiation, 1×10^{11} AAV2-AQP1 particles was delivered to the submandibular gland of each mouse. By day 200 post-AAV, the treated group exhibited a mean recovery of 68% of salivary function compared to age and gender matched controls. RNA-seq detected 1,985 genes as differentially expressed in the irradiated control group when compared to healthy controls. AAV2-AQP1 treatment reduced differential expression, with only 99 genes detected as differentially expressed (>2 FC, $p\text{-val} < 0.05$). AAV2-AQP1 treatment reduced the Euclidean distance between the healthy and post-irradiation mice by 64%, indicating a substantial recovery of the epigenetic state following AAV treatment. Finally, AAV2-AQP1 treatment normalized the gene regulatory network connectivity of the irradiated salivary gland when measured by ATAC-seq footprinting at a depth of > 200 million reads. Collectively, our results indicate that AAV2-AQP1 treatment resolves the salivary functional defects associated with irradiation via correction of the cellular epigenetic state.

965F

Non-coding RNAs underlie genetic predisposition to breast cancer. *J.D. French¹, M. Moradi Marjaneh¹, J. Beesley¹, T. O'Mara¹, P. Mukhopadhyay¹, S. Kazakoff¹, L. Fachal², N. Waddell¹, D.F. Easton^{2,3}, A.M. Dunning², G. Chenevix-Trench¹, S.L. Edwards¹.* 1) Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, 4029, Australia; 2) Department of Oncology, Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge CB1 8RN, UK; 3) Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge CB1 8RN, UK.

To date, the focus of Genome Wide Association Studies (GWAS) follow-up studies has been the impact of regulatory variants on the expression of protein-coding genes. However, in many cases, genetic variants with stronger associations with the trait of interest are dismissed as causal because they do not fall in regions marked for enhancer or promoter activity. GWAS combined with fine-mapping have identified 196 strong signals (conditional p -value $< 10^{-6}$) associated with breast cancer. The majority of candidate causal variants (CCVs) at each signal fall outside protein-coding regions. Breast cancer CCVs are enriched in regulatory elements, but the contribution of long non-coding RNAs (lncRNAs) is still currently unknown. Using targeted RNA sequencing combined with *de novo* transcript assembly, we systematically annotated multi-exonic lncRNAs transcribed from 1.5Mb intervals surrounding breast cancer GWAS signals and assess their contribution to risk. We captured >4000 lncRNA genes (FPKM > 0.5; nearly two-thirds of which are not in GENCODE, FANTOM-CAT or NONCODE databases) and show that breast cancer CCVs are significantly enriched in the exons of the captured lncRNAs but not the promoters or introns ($P_{\text{perm}} < 10^{-3}$). Approximately 75% of the captured lncRNAs are detectable (FPKM > 1) in normal breast or breast tumour RNAseq from The Cancer Genome Atlas (TCGA). Notably, principal component analyses indicate that the expression of these lncRNAs discriminate normal breast from tumour samples and breast cancer subtypes suggesting that these transcripts may play an important role in breast cancer development. Using eQTL analyses we identify 16 lncRNAs whose expression is associated with CCVs in breast tumours (FDR < 0.05). Approximately half of the lncRNA contain the eSNP within the lncRNA exon suggesting that the variants may alter the stability of transcripts, the remainder are located more distally. We performed Capture Hi-C in 6 breast cell lines and provide evidence that these CCVs fall in distal enhancers that regulate lncRNA genes through long-range chromatin interactions. In summary, we identify 874 lncRNAs as candidate breast cancer risk genes based on CCVs that fall in either 1) lncRNA exons, 2) lncRNA promoters or 3) regions that interact with the lncRNA promoters through long-range chromatin interactions. These data indicate that modulation of lncRNAs by risk-associated variants is a predominant mechanism underlying breast cancer development.

966W

Characterizing DNA methylation patterns in pancreatic cancer patients. *R.J. Jansen¹, W.R. Bamlet², G.M. Petersen².* 1) North Dakota State University, Fargo, ND; 2) Mayo Clinic, Rochester, MN.

Understanding how methylation changes with cancer development and identifying which methylation markers (CpGs) are important in related conditions (e.g., recent onset type 2 diabetes) has extreme importance when using CpGs as biomarkers for specific diseases or health outcomes. Therefore in this pilot study, we wanted to characterize the effect pancreatic adenocarcinoma (Pca) and recent onset DM has on DNA methylation and determine if these effects vary by genomic context. We sampled 30 non-smoking patients (15 Pca and 15 recent DM) from the Mayo Clinic pancreatic cancer resource of about 3K cases/2K controls and used the Illumina EPIC methylation array to measure methylation of about 850,000 CpGs in DNA extracted from their stored whole blood. We identify disease-associated CpGs adjusted for blood cell type and characterized location in CpG islands (vs. shelf, shore, open sea), and gene regions (vs. intergenic). Overall, we classified about 30 CpGs significantly ($p < 5 \times 10^{-5}$) associated with Pca and DM each, but only 1 CpG that overlap in both the sets of top 100 (those with the smallest p -values). However, we did observe a similar hypermethylation pattern with respect to genomic region among both sets of significant CpGs. When comparing all CpGs to disease associated CpGs, open sea is enriched among Pca and there is slight under-enrichment among the shore regions. This is in contrast with the few previous genome-wide studies on pancreas tissue-based DNA methylation and Pca with respect to genomic region, where it has been observed that average Pca-associated CpGs in islands tend to be hypermethylated while open seas, shores, and shelves hypomethylated. Using a methylation age prediction model, our sample tended to have younger predicted ages compared to chronological ages with the Spearman correlations as follows: 0.90 for the Pca, no DM group, 0.86 for the Pca, with DM group, 0.81 for the no Pca, with DM group, and 0.93 for the no Pca, no DM group. The lone CpG marker observed in DM and Pca models, shows promise for discriminating Control, DM and Pca, DM patients from one another. Sample size is a major limitation and this analysis is exploratory, even so, the potential modification of LAMA5 methylation and its associated functions including roles in cell adhesion, differentiation, migration, signaling, and metastasis along with a gene mutation affecting sonic hedgehog and GLI1 expression supports a potential important role for LAMA5 in Pca development.

967T

Tumor-specific DNA methylation signatures in solid and liquid biopsies.

A.C. Lau, R.L. Goldfeder, C.L. Wei. Genome Technologies, The Jackson Laboratory for Genomic Medicine, Farmington, CT.

Liquid biopsies have the potential to revolutionize cancer diagnosis, therapy selection, and treatment response. Currently, most clinical decisions are based off of tumor extraction and genotyping, however tumor biopsies are often difficult to obtain in that it is slow and painful for the patient. Therefore, as a proxy for tumor tissue profiling, analyzing cell-free circulating tumor DNA (ctDNA) in blood samples offers a fast non-invasive way to access the genetic and epigenetic information of patients' tumor status. Epigenetic regulation is thought to contribute to tumorigenesis across all cancer types. Unlike cancer mutation profiles, which can vary within a single tumor type and are sparse in cell-free DNA (cfDNA), DNA methylation tends to be more consistent and occurs in specific DNA regions. Additionally, 5-hydroxymethylcytosine (5hmC) has a conserved tissue type/cell specific genome-wide distribution, making it an excellent mark to determine tumor origin from a liquid biopsy. Our goal is to characterize the diagnostic and prognostic potential of the cell-free DNA epigenetic modifications 5-methylcytosine (5mC) and 5hmC. In the past, the low amounts of ctDNA in blood samples have presented a challenge in sequencing and identifying different epigenetic modification. We adapted and optimized nano-scale oxidative-bisulfite sequencing and analysis to achieve single-base resolution of 5mC and 5hmC in genomic DNA and ctDNA. We used these technical and analytical methods to sequence the tumor DNA, matching normal DNA, and ctDNA of pancreatic and bladder cancer patients. When comparing tumor DNA to matching normal DNA we found distinct methylation profiles, different levels of enrichment at genomic regions, and a unique set of methylated and hydroxymethylated genes. Additionally, when comparing 5mC and 5hmC patterns found in a solid tumor biopsy to the matching ctDNA, we uncover a high correlation in methylation levels as well as intact cancer-specific methylated and hydroxymethylated genes. The cancer-specific genes with enriched 5hmC patterns were unique to the gene expression pathways altered in the ctDNA tumor cell origin. Therefore, the methylated genes and regions we identified can be used to assay ctDNA alone and act as a faithful surrogate for tumor biopsies. Overall, this method offers potential new avenues to non-invasive diagnosis, monitoring and stratification of cancer progression through the analysis of the ctDNA methylome.

968F

A pilot study of DNA methylation, epigenetic age, and breast cancer risk status.

R.M. Lucia¹, A. Alvarez¹, I. Masunaka¹, A. Ziogas¹, D. Goodman¹, A. Odegaard¹, V. Lee², S. Brain², D. Heditsian², T.M. Norden-Krichmar¹, H.L. Park¹. 1) Epidemiology, University of California, Irvine, CA; 2) Patient Advocate, Markers for Environmental Exposures Study, Irvine, CA.

Background: DNA methylation is a key mechanism of epigenetic regulation, and peripheral blood DNA methylation has been studied for potential markers for breast cancer (BC) risk. DNA methylation markers for some BC risk factors, including smoking and body mass index, have been identified. However, many factors contribute to BC susceptibility, and markers for risk status according to clinical risk models that incorporate multiple risk factors have not been described. In addition, while accelerated epigenetic age has been shown to be predictive of BC risk, the relationship between epigenetic age and BC risk status according to multifactorial risk models is unknown. We conducted a pilot study to test the methods and feasibility of identifying DNA methylation markers for BC risk status and examining its relationship with epigenetic age.

Methods: While recruitment of our full study cohort (n=400) is underway, this pilot study included 16 women. Blood samples and questionnaire data were obtained from postmenopausal women ages 45-65 without a history of BC. Lifetime BC risk was estimated for each woman using the Tyrer-Cuzick model. Participants were categorized as high-risk (lifetime risk $\geq 20\%$) or average-risk (lifetime risk $< 20\%$). DNA methylation was measured using the Illumina Infinium MethylationEPIC BeadChip. Following normalization of raw data, differentially methylated positions (DMPs) were identified using an F-test. Epigenetic age was computed according to Horvath's method. Participants were categorized as having accelerated epigenetic age if epigenetic age was greater than chronological age. **Results:** This study identified 6440 DMPs associated with high-risk status, but none remained significant after false discovery rate correction. Among the top 20 DMPs, 40% were near DNA repair genes, potential oncogenes, or potential tumor suppressors. While epigenetic age correlated closely with chronological age in this sample (Pearson's $r = 0.926$), there appeared to be a trend of accelerated epigenetic age among high-risk women, with 67% of high-risk women exhibiting accelerated epigenetic age compared to 50% of average-risk women. The median difference was 1.55 years for high-risk and -0.13 for average-risk women (Wilcoxon p-value = 0.475). **Conclusion:** These results demonstrate the feasibility of the methods involved in the study. Further analysis of our larger cohort will elucidate the relationships between DNA methylation, epigenetic age, and BC risk status.

969W

Allele-specific effects of rs12416605:C/T in miR-938 on the regulation of gastric cancer genes. Y. Espinosa-Parrilla^{1,2}, I. Torruella-Loran², MK. Ramirez Viña¹, D. Zapata-Contreras¹, N. Sala^{3,4}. 1) Universidad de Magallanes, Punta Arenas, Magallanes y de la Antártica C, Chile; 2) IBE, Institute of Evolutionary Biology (Universitat Pompeu Fabra □CSIC), Experimental and Health Sciences, Barcelona, Catalonia, Spain; 3) Molecular Epidemiology Group, Translational Research Laboratory, Catalan Institute of Oncology-IDIBELL, Barcelona, Spain; 4) Human Molecular Genetics Group, Bellvitge Biomedical Research Institute (IDIBELL), Barcelona, Spain.

MicroRNAs are small endogenous regulatory RNAs that are involved in the control of multiple biological pathways. Allele variants in microRNAs are known to contribute to disease by affecting either microRNA expression or function. We have previously identified an association of rs12416605:C/T, in the seed region of miR-938, with the diffuse subtype of gastric cancer and aimed to investigate the functional effect of this SNP. To test the influence of rs12416605:C>T in miR-938 expression, either the C or T miR-938 alleles were transfected in HeLa cells and miR-938 expression was measured by RT-qPCR. The rs12416605 C allele was 1.5 times more abundant than the T allele (Student t-test, p-value < 0.05) indicating that rs12416605 may affect the expression of miR-938. Furthermore, to investigate the effect of rs12416605:C/T on the spectrum of miR-938 target genes, transcriptome analyses on HeLa cells over-expressed with either the C or T rs12416605 alleles were performed. From approximately 500 deregulated transcripts, around 50% were exclusively regulated by one or the other rs12416605:C>T alleles. Additionally, even though both variants presented cancer as the top disease associated with their deregulated genes (Ingenuity Pathway Analysis), several genes implicated in the aetiology of gastric cancer, such as the Fibroblast Growth Factor Receptor 1 (*FGFR1*), the NK2 Homeobox 8 gene (*NKX2-8*) and the chemokine *CXCL12* genes, were specifically deregulated by only one of the rs12416605:C>T alleles and were predicted as miR-938 allele specific targets. All together suggests that rs12416605 may be involved in gastric cancer susceptibility by affecting the regulatory function of miR-938. Grant FONDECYT Regular- N°1170446.

970T

Patterns of X chromosome inactivation differ between tumor and normal ovarian tissue. S.J. Winham¹, N.B. Larson¹, M.C. Larson¹, S.M. Armasu¹, J.M. Cunningham², S. Gayther³, K. Lawrenson³, E.L. Goode¹. 1) Health Sciences Research, Mayo Clinic, Rochester, MN; 2) Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN; 3) Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles.

X chromosome inactivation (XCI) is a key epigenetic gene expression regulatory process to silence transcription of one of the two copies in females. However, some genes are known to escape this process and are expressed from both copies, and these escape patterns vary across tissues. Although XCI is specific to females, XCI patterns have not previously been studied in normal ovarian tissue. Furthermore, we have previously shown that XCI patterns may play a role in ovarian cancer (OC), particularly if tumor suppressors or oncogenes are inactivated or escaping in a dysregulated manner. In this study, we assess XCI patterns in normal tissue derived from ovarian and fallopian tube surface epithelial cells from 70 patients, and compared these patterns to those previously described for other tissue types, as well as those observed in tumor tissue samples derived from N=45 OC patients. Using allele-specific expression data, we used a two-stage method to estimate the probability of escape from XCI for over 300 genes. The XCI patterns in the normal ovarian samples were similar to previously published patterns of other tissue types. However, approximately 13% of genes showed patterns across individuals that were different from other tissue types, with these genes being inactivated in ovarian tissue. Many have been previously reported as tumor suppressors in ovarian or other cancers (e.g. *KAL1*, *USP9X*, *DDX3X*, *TRAPPC2*, *TCEANC*, and *KDM6A*). Notably, 63% of these discrepant genes also showed patterns of inactivation in the ovarian tumor samples. While XCI ovarian tumor patterns were concordant for most genes compared to normal ovarian tissue, 9% of genes that were inactivated in the normal samples were escaping XCI in the tumor samples. Furthermore, across 276 genes, 72 genes had significantly different probabilities of escape between tumor and normal samples ($P < 0.00018$), with most showing complete inactivation across normal samples but higher probability of escape in the tumor samples (including *DDX3X*, *TRAPPC2*, *KDM6A*, and *RPS4X*). Notably, *RPS4X* was previously associated with poor prognosis in OC, and *DDX3X* and *TRAPPC2* have also been identified as possible drug targets. We found evidence of a unique ovarian tumor XCI profile, suggesting that XCI may play an important role in OC biology. Additional studies to validate in large samples and to examine somatic changes with paired tumor-normal tissue are needed.

971F

Association of Androgen Receptor (AR) gene CAG repeat polymorphism in prostate cancer and benign prostate hyperplasia in men from India.

S. Daram¹, S. Zubeda¹, S. Poornima², VL. Ayyala¹, B. Srinadh¹, P. Shailaja¹, Q. Hasan¹. 1) Kamineni Hospitals, Hyderabad, Telangana, India; 2) Kamineni Life Sciences, Hyderabad, Telangana, India.

Prostate cancer (PCA) is the second most common malignancy in men worldwide with incidence and mortality are 3.7 and 2.5 per 1, 00,000 men annually. In PCA, 70% of cases are reported from developed countries because of aggressive screening with the PSA test, Digital Rectal Examination (DRE) and subsequent biopsy. PCA incidence and mortality in major cities of India is 3 to 10 per 1, 00,000 men. Prostate disease is a heterogeneous, which ranges from benign to pathogenic. Androgens acting through Androgen Receptor (AR) play an important role in physiology and etiology of PCA. CAG repeat polymorphism has been associated with PCA in different populations. Aim of the present study was to assess the Androgen Receptor gene CAG repeat polymorphism in Prostate Cancer, Benign Prostate Hyperplasia (BPH) and in comparison with age matched healthy males. Peripheral blood from 315 males was collected after obtaining ethical committee approval. Genomic DNA was isolated using salting out method followed by polymerase chain reaction (PCR) and polyacrylamide gel electrophoresis (PAGE) for CAG repeat analysis. Odds ratios (OR) and 95% Confidence Intervals (CI) were calculated using logistic regression analysis. The CAG repeat in controls range from 20 to 32 with a Mean of 23±3 which is significantly longer when compared with PCA, which showed a range from 9 to 24 with Mean 15±4 and in BPH CAG repeat range from 16 to 25 with Mean 22±2. The odds ratio of BPH and Controls (OR=3.0978 95%CI=1.7611 to 5.4493) showed a significance of P=0.0001, PCA and BPH (OR=9.6000 95%CI=3.2570 to 28.2962) with a significance of P<0.0001, PCA and Controls (OR=29.7391 95%CI=10.1702 to 86.9612) with a significance of P<0.0001. Shorter CAG repeats in AR gene polymorphism were significantly associated with PCA than controls and BPH.

972W

A novel germline *DICER1* mutation in association with thyroid and ovarian Sertoli-Leydig cell tumors. N. Settas¹, L. Michala², A. Berthon¹, F.R. Fau-
cz¹, A. Iliadi³, A. Gkika³, C. Dacou-Voutetakis⁴, A. Voutetakis⁴, C.A. Stratakis¹.

1) Section on Endocrinology and Genetics, Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), National Institutes of Health (NIH), Bethesda, MD 20892, USA; 2) Department of Reproductive Endocrinology, Alexandra Hospital, University of Athens, Athens, Greece; 3) Department of Biochemistry, Institute of Child Health, Athens, Greece; 4) Endocrine Unit, 1st Department of Pediatrics, Medical School, University of Athens, Athens, Greece.

DICER1 is an endoribonuclease that acts post-transcriptionally by processing mRNA into siRNA and microRNA, thus leading to mRNA downregulation. *DICER1* syndrome is usually caused by germline variants and is characterized by a variety of benign or malignant tumors: pleuropulmonary blastoma, ovarian Sertoli-Leydig cell tumor, cystic nephroma, pituitary blastoma and thyroid tumors including multinodular goiter. Patients with germline aberrations in the *DICER1* gene may carry additional somatic missense *DICER1* mutations within the associated tumors, located in the metal ion-binding residues of the RNase IIIb domain. Our patient, a 10 yo girl was evaluated for thyroid enlargement. While her thyroid function tests were normal, thyroid sonography revealed three nodules in the right lobe (diameter: 1.9, 2.3, and 3 cm, respectively) that proved hyperplastic nonmalignant after thyroidectomy. She entered puberty at age 11 years and at age 12 years, both breast and pubic hair were Tanner stage IV. Ten months later, during follow-up, breast had regressed to Tanner stage I, hirsutism was present and her voice had notably deepened. Hormonal evaluation showed elevated levels of Testosterone, DHEAS, 17OH-progesterone, $\Delta 4$ androstenedione, while Cortisol was within normal range. Ovarian sonography revealed an enlarged right ovary (12cc) while the left ovary was normal (3cc). Right salpingo-oophorectomy was carried out and biopsy revealed an ovarian Sertoli-Leydig cell tumor. Androgens returned to normal levels postoperatively. Genetic analysis revealed a novel germline nonsense *DICER1* mutation (p.W1481*, c.4443G>A), inherited from her father (in whom sonography revealed multinodular goiter). Further studies in the ovarian and thyroid tissues showed two different somatic missense mutations of the same codon. The ovarian tissue carried the missense mutation p.E1813D (c.5439G>T) in heterozygosity whereas, the thyroid tissue carried the heterozygous missense variant p.E1813Q (c. 5437G>C). Immunohistochemistry revealed that within these tumors, the *DICER1* protein expression was significantly decreased compared to normal tissue. In conclusion we report a novel *DICER1* mutation (p.W1481*, c.4443G>A) and confirm that patients with germline *DICER1* mutations can develop somatic mutations within the RNase IIIb domain. Our findings show that impaired *DICER1* function affects the thyroid and ovarian tissue homeostasis and offer further proof that *DICER1* is a tumorigenic driver during childhood.

973T

***TNS1* gene mutations are present in sporadic pheochromocytomas and paragangliomas.** F.R. Faucz¹, N. Settas¹, A. Berthon¹, S.K. Flores², P.L. Dahia², C.A. Stratakis¹. 1) Section on Endocrinology and Genetics, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD; 2) Dept. Medicine and Mays Cancer Center, University of Texas Health Science Center at San Antonio, TX.

Tensin1 is an adaptor phosphoprotein linking integrin to the actin cytoskeleton; it is also involved in fibrillar adhesion formation. Tensin1 is encoded by the actin-binding protein Tensin1 (*TNS1*). *Tns1* knockout (KO) mice develop kidney abnormalities; female mice have also reduced fertility. A *Drosophila* Melanogaster *Tns1*KO model is known as blistery due to defective wing unfolding and consequent blister formation. *PTEN*(phosphatase and tensin homolog), the known tumor suppressor gene, is a paralog of *TNS1*. Genome-wide association studies have linked *TNS1* to asthma, lung disease and colorectal cancer, and *TNS1* expression is reduced or absent in several prostate and breast cancer cell lines. Pheochromocytomas (PH) and paragangliomas (PGL) are neural-crest-derived tumours that are genetically heterogeneous. More and more of these tumors are caused by inherited mutations. In the present study we screened 43 unrelated patients with PGL and/or PH through whole exome sequencing. Among the 27 patients where a known causative mutation was not found (including genes such as *SDHx*, *NF1*, *RET*, *HRAS*, *EPAS1*, *VHL*, and *FH*) we identified a total of 7 different *TNS1* coding sequence alterations, which were predicted to be damaging through in silico analysis. The frequency of patients with a predicted as damaging variation was 33.33% (9/27); a total of 7 variants were non-synonymous substitutions leading to a missense variation. One of these variants was novel (c.5174C>G, p.S1725P). The remaining 6 variants (p.R5W, p.V25I, p.I311M, p.G320R, p.R859W, and p.E869K) were rare or absent in the general population showing a statistical difference when compared with the frequencies found in our sample ($p < 0.001$ using "ALL" in the gnomAD database (<http://gnomad.broadinstitute.org>)). Immunostaining and additional genetic studies are ongoing to further clarify the role of these genetic changes in PGL and/or PH. We conclude that inactivating *TNS1* sequencing variants may be yet another cause of PH and PGL, adding to the long list of genes involved in these tumors. Our ongoing studies aim at further characterizing the role of *TNS1* in neuroendocrine tumor development.

974F

CRISPR/Cpf1 as a tool for locus specific delivery of suicide gene in chronic lymphocytic leukemia. F.C. Lorenzetti¹, F.T. Papa¹, S. Daga¹, S. Croci¹, F. Niccheri², F. Donati², D. Lopercolo¹, M.A. Mencarelli^{1,3}, A. Gozzetti⁴, M. Bocchia⁴, S. Conticello², A. Renieri^{1,3}, F. Mari^{1,3}. 1) Medical Genetic, University of Siena, Siena, Italy, Siena, Italy; 2) Core Research Laboratory - ISPRO, Florence, Italy; 3) Genetica Medica, Azienda Ospedaliera Universitaria Senese, Italy; 4) Department of Medicine and Immunological Sciences, Hematology Unit, Azienda Ospedaliera Universitaria Senese and University of Siena, Italy.

Patients with Chronic Lymphocytic Leukemia (CLL) harbouring *TP53* mutations, despite being subjected to burdensome cycles of therapy, do not have many possibilities of healing due to the acquired abilities of *TP53*-mutated clones to escape the control systems. We have recently provided evidence that fully supports the critical role of *TP53*-mutated clones in driving the neoplastic process and that these cells are present several months before disease manifestation and they are enhanced by chemotherapy (Pinto AM et al. Br J Haematol 2018). We present here a suitable example of transformative gene-based medicine using CRISPR/Cpf1 for locus specific delivery of suicide gene in lymphocytes successfully isolated from *TP53* mutated CLL patients. We are able to introduce the gene for the pro-drug converting enzyme herpes simplex virus type 1 thymidine kinase (HSV1-TK) into the genome of cancer cells carrying specific *TP53* mutations. Very promising results are obtained using our system in HEK293 cells line. Implementation of this toolkit and in vivo experiment on mouse are ongoing. The precision specificity and the integration rate of our approach for the genome therapy might make it possible to apply this protocol to clinical setting. Patenting process of this toolkit is ongoing at University of Siena.

975W

The Fanconi anemia proteins FANCM and FAAP24 limit BLM/Pol δ dependent telomere instability in the alternative lengthening of telomeres (ALT) pathway.

M.S. Meyn^{1,2,3,4,5,6}, J. Te³, M. Komosa³, F. Alazr^{3,5}. 1) Center for Human Genomics and Precision Medicine, University of Wisconsin, Madison, WI; 2) Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada; 3) Program in Genetics and Genome Biology, Hospital for Sick Children, Toronto, ON, Canada; 4) Division of Clinical and Metabolic Genetics, Hospital for Sick Children, Toronto, ON, Canada; 5) Institute of Medical Science, University of Toronto, Toronto, ON, Canada; 6) Department of Paediatrics, University of Toronto, Toronto, ON, Canada.

~15% of human tumours lack telomerase and use the Alternative Lengthening of Telomere (ALT) mechanism to maintain telomeres. ALT is characterized by variable telomere lengths, frequent telomere sister chromatid exchanges, and increases in extrachromosomal telomeric repeat (ECTR) DNA. We previously demonstrated that depletion of FANCD2, a member of the BRCA/Fanconi anemia DNA repair pathway, causes a hyper-ALT phenotype. Using ALT and telomerase-positive human cells, we extend our analysis to the DNA translocase FANCM and its partner FAAP24; proteins which stabilize stalled replication forks, target other Fanconi anemia proteins to damaged DNA, and promote DNA repair. We find siRNA depletion of FANCM or FAAP24 increases telomeric replicative stress in ALT cells as indicated by increased telomeric fragile sites. Depletion results in rapid increases in telomeric DNA content in ALT but not telomerase-positive cells as measured by Flow-FISH and Halo-FISH. Further, these increases in telomeric DNA are primarily due to 4-5 fold increases in ECTR DNA molecules, but not their length. Interestingly, FANCD2 protein localizes to telomeric foci in FANCM or FAAP24 depleted ALT cells, suggesting that FANCM and FAAP24 protect against telomeric replicative stress by directly interacting with replication forks, rather than by recruiting FANCD2 to stressed forks. siRNA depletion of either the BLM DNA helicase or the pol δ DNA polymerase in ALT cells resulted in telomere shortening and markedly fewer ECTR DNA molecules. Depletion of BLM or pol δ also suppressed the excessive ECTR production and increased telomere length seen in FANCM and FAAP24-depleted ALT cells, suggesting that these telomere abnormalities are an exaggeration of the ALT phenotype rather than due to a separate mechanism. Our results support a ALT model in which failed telomeric replication forks give rise to DNA breaks and gaps used by BLM/pol δ -dependent recombination-driven replication processes to drive ALT, which produces high levels of ECTR DNA. In this model, FANCM and FAAP24 restrain the ALT phenotype by promoting recovery of stalled telomeric replication forks. Loss of FANCM or FAAP24 increases the natural substrate for ALT, failed telomeric replication forks, hence the hyper-ALT phenotype observed in FANCM or FAAP24-depleted ALT cells. As the hyper-ALT phenotype is associated with increased cell death, disruption of the BRCA/Fanconi anemia pathway may be a viable treatment strategy for ALT tumours.